10-2014

# Self-referentiality in Constructive Semantics of Intuitionistic and Modal Logics

Junhua Yu
*Graduate Center, City University of New York*

# Self-referentiality in Constructive Semantics of Intuitionistic and Modal Logics

by

Junhua Yu

A dissertation submitted to the Graduate Faculty in Computer Science in partial fulfillment of the requirements for the degree of Doctor of Philosophy, The City University of New York.

2014

This manuscript has been read and accepted for the Graduate Faculty in Computer Science in satisfaction of the dissertation requirements for the degree of Doctor of Philosophy.

Sergei Artemov (*required signature*)

Date

Chair of Examining Committee

Robert Haralick (*required signature*)

Date

Executive Officer

Sergei Artemov

Yegor Bryukhov

Melvin Fitting

Rohit Parikh

Supervisory Committee

THE CITY UNIVERSITY OF NEW YORK

Abstract

**Self-referentiality in Constructive Semantics of Intuitionistic and Modal Logics**

by

Junhua Yu

Advisor: Sergei Artemov

This thesis explores self-referentiality in constructive semantics of intuitionistic and modal logics formalized via the framework of justification logic. In this framework initialed by Artemov, the language has formulas of the form $t:\phi$, which means "the term $t$ is a justification of the formula $\phi$." Moreover, terms can occur inside formulas and hence it is legal to have $t:\phi(t)$, which means "the term $t$ is a justification of the formula $\phi$ about $t$ itself." Formulas like this is not only interesting in the semantics of justification logic, but also, as we will see, necessary in applications of justification logic in formalizing constructive semantics implicitly carried by modal and intuitionistic logics.

Works initialed by Artemov and followed by Brezhnev, Fitting and others have successfully extracted constructive meaning packaged by modality in many modal logics. Roughly speaking, they offer methods of substituting modalities by terms in various justification logics, and then computing the exact structure of each term. After performing these methods, each (sub)formula prefixed by a modality becomes a (sub)formula prefixed by a term, which is then interpreted as the justification of the (sub)formula being

prefixed. In terminology of this framework, we say that modal logics are "realized" in justification logics.

Within the family of justification logics, the Logic of Proofs LP is perhaps the most important member. As Artemov showed, this logic is not only complete w.r.t. to arithmetical semantics about proofs, but also accommodates the modal logic S4 via realization. Combined with Gödel's modal embedding from intuitionistic propositional logic IPC to S4, the Logic of Proofs LP serves as an intermedium via which IPC receives its provability semantics, also known as Brouwer–Heyting–Kolmogorov semantics, or BHK semantics.

This thesis presents the candidate's works in two directions. (1) Following Kuznets' result that self-referentiality is necessary for the realization of several modal logics including S4, we show that it is also necessary for BHK semantics. (2) We find a necessary condition for a modal theorem to require self-referentiality in its realization, and using this condition to derive many interesting properties about self-referentiality.

# Acknowledgements

As one ages, he has more and more people to thank, but less and less time to do so. Hence it is not a good idea to start from my born. Instead, I will mention only those people whose helps have been necessary for either my current thread of research, or my good times as a student at the Graduate Center of the City University of New York.

The most important person on both entries is Sergei Artemov, my advisor and mentor. Besides his academic framework without which nearly all my research would go baseless, he has really influenced me to spend time on topics that are meaningful, rather than topics that are somehow well acknowledged, or merely hot. For times, his advises instructed me where to go next, and pointed out the meaning of proposed works. Guess I will realize more and more benefits of having studied with him in future years, or even decades.

It is my honor to have Yegor Bryukhov, Melvin Fitting, and Rohit Parikh in my dissertation supervisory committee. Their questions and comments have led me to think the work via different perspectives, and enlarge my field of view.

My current thread of research started in the year 2008 when I was a master student in Tsinghua University. From then on, comments and suggestions have been received from, other than those mentioned above, following persons, ordered chronologically: Fenrong Liu (also my mentor at Tsinghua University from 2007 to 2010 who has been always supporting me since then), Roman Kuznets (also the forerunner of the research of self-referentiality in the framework of justification logic), Johan van Benthem, Eric Pacuit, Dick de Jongh, anonymous referees of CSR 2010, Dafa Li, Lu Wang (also the person who led me to the discipline of logic), Feng Ye, Hidenori Kurokawa, Ren-June Wang, Evan Goris, Can Başkent, Konstantinos Pouliasis, Karen Kletter, anonymous referees of LFCS 2013, Rosalie Iemhoff, and Minghui Ma.

At the Graduate Center, I was lucky to have chances taking/auditing logical courses offered by the following instructors, ordered chronologically: Noson Yanofsky, Subash Shankar, Melvin Fitting, Russell Miller, Rohit Parikh, Saul Kripke, and Richard Mendelsohn.

Lina Garcia has been helping me with each step in the Ph.D. Program, for most of the times as the assistant program officer, and then voluntarily.

This thesis is prepared with the LaTeXclass file by George Leibman.

I have been financially supported by the Graduate Center and the Research Foundation of the City University of New York.

As their only child, I have been away from my parents for so long and so far. In memory of my grandfather who had been always supporting me unconditionally, I am guilty not was able to see him off. This thesis is not dedicated to any of them, since they each deserves a reunified family, which could be much better.

# Contents

# List of Tables

# List of Figures

# Chapter 1

# Introduction

Self-referentiality has be an interesting topic in many aspects of logic and related fields. Its trace can be seen in (versions of) the Liar Paradox that emerged during ancient times. Since the proposal of modern mathematical logic, there have been several milestone-like works where self-referentiality plays a key role. In Russell's paradox we see the presumed set of sets that each does not belong to it self; in Gödel's incompleteness theorem of arithmetics we see an arithmetical sentence that is equivalent to the assertion of its own unprovability; in Turing's unsolvable of the halting problem we see a presumed program that goto itself if the program coded by the input halts on the input itself; in Kleene's recursion theorem we see Rogers' construction that employs a function that, when given an input, attempts to compute the value of the function indexed by the input when the input itself is fed, and if succeeds, returns the index of a function that behaves the same as the function indexed by that value. Though more instances can be listed, those just mentioned are enough to indicate how important self-referentiality can be. Books like [59] and [67] have general surveys and detailed analysis of many instances.

In this thesis, we will explore self-referentiality in constructive semantics of intuitionistic and modal logics, formulated via the framework of justification logic. Self-referentiality is involved since a language of justification logic has the ability of saying something like "a justification justifies an assertion about the justification itself." As we will present in later chapters, this ability is not only interesting by its own, but also necessary in applications of justification logic in formalizing constructive contents implicitly carried by modal and intuitionistic logics.

Here is an outline of this thesis. The current Chapter 1 is an introduction in which we present backgrounds like the Brouwer–Heyting–Kolmogorov semantics and the framework of justification logic. Chapter 2 discusses self-referentiality in modal logics in terms of their constructive reading via justification logics. Some results presented in that chapter is gained in earlier works by Roman Kuznets. Chapter 3 takes a closer look at modal logics that are self-referential, and explores properties of their "non-self-referential fragments." Chapter 4 tracks one step further to the Brouwer–Heyting–Kolmogorov semantics of intuitionistic propositional logic via the Gödel–Artemov approach, the necessity of self-referentiality there, and properties of non-self-referential fragments. Chapter 5 is a concluding chapter where several directions for future works are suggested.

The current chapter is started by Section 1.1 presenting justification logic in general with formal definitions.

Followed is Section 1.2 carrying a brief historic introduction about the Brouwer–Heyting–Kolmogorov semantics, the formalization of which has served as the initial motivation of the first justification logic, Artemov's Logic of Proofs.

In Section 1.3, we formally present the phenomena of self-referentiality in the language of justification logic with their intuitive meanings in various kinds of semantics.

Before start, we present here some pre-settings effective in the whole thesis. Settings typically for specified chapters or sections will not be given until necessary. In this thesis, we denote formulas or subformulas by $\alpha$, $\beta$, ..., $\phi$, $\psi$, ..., and sets of formulas by $\Gamma$, $\Theta$, ..., with optional superscripts or subscripts. By a prime formula, we mean either a propositional atom like $p$, $q$, etc., or a falsehood $\perp$. In a propositional language, falsehood $\perp$ and implication $\rightarrow$ are always primitive connectives unless otherwise mentioned, and $\top$ is defined as an abbreviation of $\perp \rightarrow \perp$. In intuitionistic setting, conjunction $\wedge$ and disjunction $\vee$ are also primitive. Negation $\neg\phi$ is defined as the abbreviation of $\phi \rightarrow \perp$. In classical setting, $\alpha \vee \beta$ is defined as the abbreviation of $\neg\alpha \rightarrow \beta$, and $\alpha \wedge \beta$ as the abbreviation of $\neg(\alpha \rightarrow \neg\beta)$. Enabled by commutativity and associativity, generalized conjunction $\bigwedge$ and disjunction $\bigvee$ can be applied to sets of formulas, while $\bigwedge \varnothing := \top$ and $\bigvee \varnothing := \perp$ by definition. In a modal language, $\Box$ is the only primitive modality, and $\Diamond\phi$ is defined as the abbreviation of $\neg\Box\neg\phi$. For binding force, $\neg$ and $\Box$ are the strongest, followed by $\wedge$ and $\vee$, and then $\rightarrow$. We usually present all parentheses except for the outer-most, and in case that we do not, right associativity is assumed for connectives. In general, $\vdash$ means syntactical derivability, and $\Vdash$ means semantical satisfiability (semantical consequence is hardly used, if ever). By $\epsilon_0 \rhd \epsilon_1$, we mean that syntactical expression $\epsilon_0$ occurs in $\epsilon_1$ as a sub-expression. Accordingly, $\nvdash$, $\nVdash$, and $\ntriangleright$ respectively mean "not derivable," "not satisfied," and "does not occur as a sub-expression of." Polarities of subformulas (and their main connectives) are defined as usual. For example, $\phi \wedge \psi$ is a positive subformula of $\chi$ implies that $\phi$ is also a positive subformula of $\chi$. For another example, $\phi \rightarrow \psi$ is a negative subformula of $\chi$ implies that $\phi$ is a positive subformula of $\chi$, and $\psi$ is a negative subformula of $\chi$. For *classical propositional logic* $\mathsf{CPC}$, we do not specify a typical formulation of axioms.

## 1.1 The Framework of Justification Logic

The first justification logic is Artemov's Logic of Proofs $\mathsf{LP}$ [1, 2], which serves as an intermedium via which modal logic $\mathsf{S4}$ receives its provability semantics in Gödel sense, and so does the intuitionistic propositional logic $\mathsf{IPC}$. What just mentioned is called the Gödel–Artemov approach of formalizing the Brouwer–Heyting–Kolmogorov Semantics. An introduction for this approach can be given easier using notions defined in this section, as we will do in Section 1.2.

The key idea of $\mathsf{LP}$ is to employ inductively defined *(justification) terms* to play the role of modality $\Box$ in $\mathsf{S4}$. By substituting modalities in an $\mathsf{S4}$-theorem intelligently, we get an $\mathsf{LP}$-theorem, and hence "constructively explains" the meaning of modalities in the given $\mathsf{S4}$-theorem by structures of terms interpretable in semantics of $\mathsf{LP}$. Employing various classes of terms, this idea has been generalized by Brezhnev and many others to modal logics other than $\mathsf{S4}$, like $\mathsf{K}$, $\mathsf{D4}$, $\mathsf{S4.2}$, $\mathsf{S5}$, and many others. All other $\mathsf{LP}$-like logics considered in this thesis correspond to modal logics with only $\mathsf{k}$, $\mathsf{d}$, $\mathsf{t}$, $\mathsf{4}$ as axiom schemes, and are all included in [13]. All of them, together with those correspond to modal logics with $\mathsf{b}$ and $\mathsf{5}$ as axiom schemes can be found in [32]. There are also $\mathsf{LP}$-like logics correspond to less-familiar modal logics [26]. Now the whole group of $\mathsf{LP}$-like logics are called *justification logics*. Among them, five logics interesting in the current thesis are included in the following definition.

**Definition 1 (Justification Logics $\mathsf{J}, \mathsf{JD}, \mathsf{JT}, \mathsf{J4}, \mathsf{LP}$ [2, 13, 46])** Let *(justification) terms* be defined by

$$\begin{cases} \text{in } \mathsf{J}, \mathsf{JD}, \text{and } \mathsf{JT} & t ::= c \mid x \mid t \cdot t \mid t{+}t \\ \text{in } \mathsf{J4} \text{ and } \mathsf{LP} & t ::= c \mid x \mid t \cdot t \mid t{+}t \mid\, !t \end{cases}$$

Table 1.1: Axiom Schemes of Justification Logic

| Name | Form (scheme) | Adopted in Logic |
| --- | --- | --- |
| $A0.1$ | $\phi \to (\psi \to \phi)$ | All |
| $A0.2$ | $(\phi \to (\psi \to \chi)) \to ((\phi \to \psi) \to (\phi \to \chi))$ | All |
| $A0.3$ | $((\psi \to \bot) \to (\phi \to \bot)) \to (\phi \to \psi)$ | All |
| $A1.1$ | $t{:}\phi \to \phi$ | JT, LP |
| $A1.2$ | $t{:}\bot \to \bot$ | JD |
| $A2$ | $t_1 : (\phi \to \psi) \to (t_2 : \phi \to t_1 \cdot t_2 : \psi)$ | All |
| $A3$ | $t{:}\phi \to !t{:}t{:}\phi$ | J4, LP |
| $A4.1$ | $t_1 : \phi \to t_1 + t_2 : \phi$ | All |
| $A4.2$ | $t_2 : \phi \to t_1 + t_2 : \phi$ | All |

where $\cdot$, $+$, $!$ are called *operators*, and *(justification) formulas* be defined by

$$\phi ::= \bot \mid p \mid \phi \to \phi \mid t{:}\phi,$$

where $c$ is a *constant*, $x$ is a *variable*, and $p$ is a (propositional) atom. Constants and variables are called *atomic terms*. We sometimes generally say *justification language*, without specifying the logic that uses it.

*Axiom schemes* of justification logics are summarized in Table 1.1:

For *rules*: $MP$ with form

$$(MP) \quad \frac{\phi \to \psi \qquad \phi}{\psi}$$

is shared by all logics. J4 and LP have $AN.1$ with form

$$(AN.1) \quad \frac{}{c : A}$$

where $A$ is an axiom and $c$ is a constant. J, JD, and JT have $AN.2$ with form

$$(AN.2) \quad \frac{}{c_n : \cdots : c_1 : A}$$

where $A$ is an axiom, and $c_1 \cdots c_n$ is an initial segment of a series of constants. Rules $AN.1$ and $AN.2$ are called *axiom necessitation*, and the notation $AN$ in used in the general sense. ⊣

For Binding force of the three operators, "!" is in the strongest rank, "·" in the second, and "+" in the weakest. Left associativity is assumed for both "·" and "+," though we usually present enough parentheses to ease reading. For binding force of connectives, besides those mentioned in Chapter 1, "$t$:" shares the same rank with $\Box$, so for instance, $t:\phi \to \psi$ should be understood as $(t:\phi) \to \psi$.

Consider the definition of (justification) terms, roughly speaking, variables stand for justifications of hypothesis, constants stand for that of axioms (in *this* logic), operation "·" outputs the justification of the conclusion once that of the two premises of a Modus Ponens are input, operation "!" raises a term to another that justifies the given fact that a formula is justified by the given term, and operation "+" provides a joined justification that justifies anything justified by either of the two given terms. Given a justification proof, it is then possible to see how a conclusion gets justified, by analyzing the prefixed term and reducing it to variables and constants that justify hypothesis and axioms. Axiom necessitation is the rule that allows us to introduce axioms "justified" by constants, hence by taking care of applications of this rule, we can see the role of constants and axioms in a proof. This leads to the notion of *constant specification*, formally presented after the following notational convention.[1]

**Convention 2** Adopting the notation from [13] for $\mathsf{J}$, $\mathsf{JD}$, and $\mathsf{JT}$, we denote $c_n$ by $\underbrace{!\cdots!}_{n-1}c$, or in abbreviation, $!^{n-1}c$. Note that "!" dose not appear in the definition of the language of these three logics in Definition 1, and is totally notational. From now on we take

$$(AN.2) \quad \overline{!^n c : !^{n-1} c : \cdots : c : A}$$

[1]The definition of language of justification logic as presented in Definition 1 generally follows Kuznets' [46]. In Brezhnev's [13] where $\mathsf{J}$, $\mathsf{JD}$, and $\mathsf{JT}$ originally introduced, the notation for constant is different, and somehow misleading, since "!" is an operator in some logics while being purely notational in other logics. However, Brezhnev's notation is much more convenient when talking about self-referentiality, and hence we prefer to adopt it from now on.

as the form of *AN*.2. Introduction of "!" in these three logics does not indicate any operational meaning enjoyed by the same symbol in the other two logics. With this convention, the first element of a constant series becomes a sub-expression of any constant in that series.

**Definition 3 (Constant Specification)** A *constant specification* (notation $\mathcal{CS}$):

in $\mathsf{J4}$ and $\mathsf{LP}$, is a set of formulas of the form $c:A$ (where $c$ is a constant and $A$ is an axiom);

in $\mathsf{J}$, $\mathsf{JD}$, and $\mathsf{JT}$, the form is $!^n c:!^{n-1}c:\cdots:c:A$, and downward closure is required (for any natural number $z$, if $!^z c:!^{z-1}c:\cdots:c:A \in \mathcal{CS}$ then $!^{z-1}c:\cdots:c:A \in \mathcal{CS}$).

A constant specification $\mathcal{CS}$ is *injective* if for each constant $c$, there is at most one axiom $A$ s.t. $c:A \in \mathcal{CS}$.

For any constant specification $\mathcal{CS}$, if the *AN* rule of, say $\mathsf{LP}$, is restricted to introduce only formulas in $\mathcal{CS}$, then the resulting system is denoted by $\mathsf{LP}(\mathcal{CS})$. For instance, $\mathsf{LP}(\emptyset)$ is the system obtained by dropping *AN*.1 from $\mathsf{LP}$. We have similar notions and notations for the other justification logics in Definition 1.

Clearly, the collection of all *AN* rule-applications in a justification proof automatically forms a constant specification. We say the justification proof *calls for* that constant specification. $\dashv$

Justification logics enjoy the deduction theorem relativized to any constant specification, which can be proved by the standard method.

**Lemma 4 (Deduction Theorem)** *Let* $\mathsf{X}$ *ranges from* $\{\mathsf{J}, \mathsf{JD}, \mathsf{JT}, \mathsf{J4}, \mathsf{LP}\}$. *Then for any formulas* $\phi$, $\psi$, *and any set of formulas* $\Gamma$, *in* $\mathsf{X}(\mathcal{CS})$, *if* $\Gamma, \phi \vdash \psi$, *then* $\Gamma \vdash \phi \to \psi$.

Justification logics can be seen as the "explicit counterparts" of modal logics. If we replace each modality in a modal formula by a (justification) term, reading as the reason why the formula being prefixed holds, then we have a chance to explicitly explain structure of information implicitly carried by modalities.

Though there are several textbooks for modal logic (like [18, 11]), to make the thesis self-contained, we present some preliminaries about modal logics before going further. *(Propositional) modal language* is defined as the language got by adding the following closure rule to the definition of (classical) propositional language:

$$\text{If } \phi \text{ is a formula, then so is } \Box\phi,$$

where the modality $\Box$ is understood as "it is true in all accessible world that," "it is known that," "it is believed that," etc., depending on the environment in which this language is used [11]. In the framework of justification logic, we consider modal language as a language of (realizable) provability, and hence may understand it roughly as "we have an explicitly constructed justification of."

**Definition 5 (Modal Logics $\mathsf{K}, \mathsf{D}, \mathsf{T}, \mathsf{K4}, \mathsf{S4}$ [11])** In the modal language, $\mathsf{K}$ has, in addition to axiom schemes and rule of $\mathsf{CPC}$, the axiom scheme

$$\Box(\phi \to \psi) \to (\Box\phi \to \Box\psi)$$

and rule *Necessitation*

$$(Nec) \quad \frac{\vdash \phi}{\vdash \Box\phi},$$

and is complete w.r.t. the class of all Kripke frames.

$\mathsf{D}$ has, in addition to those in $\mathsf{K}$, the axiom scheme

$$\Box\bot \to \bot,$$

and is complete w.r.t. the class of serial Kripke frames (that is, each state has at least one accessible state).

T has, in addition to those in K, the axiom scheme

$$\Box\phi \to \phi,$$

and is complete w.r.t. the class of reflexive Kripke frames.

K4 has, in addition to those in K, the axiom scheme

$$\Box\phi \to \Box\Box\phi,$$

and is complete w.r.t. the class of transitive Kripke frames.

S4 has, in addition to those in T, the axiom scheme

$$\Box\phi \to \Box\Box\phi,$$

and is complete w.r.t. the class of reflexive and transitive Kripke frames. ⊣

**Definition 6 (Normal Modal Logics [11])** A collection of formulas in the modal language is a *normal modal logic* if it:

(i) contains all tautologies,

(ii) contains all instances of the scheme $\Box(\phi \to \psi) \to (\Box\phi \to \Box\psi)$,

(iii) is closed under ($Nec$), and

(iv) is closed under ($MP$). ⊣

It follows easily from the definition that the set of theorems of each of K, D, T, K4, and S4 is a normal modal logic.

We continue by presenting the notion of *realization*.

**Definition 7 (Realization [2])** A *realizer*, denoted by $r$, is a mapping from the set of all modal formulas to the set of all formulas in a justification language s.t. given any modal formula $\phi$, it replaces all $\Box$-occurrences in $\phi$ by (justification) terms.

The resulting justification formula, denoted by $\phi^r$, is called a *potential realization* of $\phi$. A potential realization $\phi^r$ of $\phi$ is a *realization* (of $\phi$), if it is a theorem in the justification logic.

A realizer is *normal*, if if assigns distinct variables to negative $\Box$-occurrences. A realization $\phi^r$ of $\phi$ is *normal*, if the realizer $r$ is normal, and there is a proof of $\phi^r$ in the justification logic that calls for a injective constant specification. $\dashv$

**Theorem 8 (Realization Theorem [2, 13])** *For any modal formula $\phi$, let* $\mathsf{X}$ *range from* $\{\mathsf{K}, \mathsf{D}, \mathsf{T}, \mathsf{K4}, \mathsf{S4}\}$*, and* $\mathsf{Y}$ *range from* $\{\mathsf{J}, \mathsf{JD}, \mathsf{JT}, \mathsf{J4}, \mathsf{LP}\}$*, resp., then what follows are equivalent:*

*(i)* $\mathsf{X} \vdash \phi$*;*

*(ii)* $\mathsf{Y} \vdash \phi^r$ *for some normal realization* $\phi^r$*.*

Theorem 8 was first proved by Artemov for $\mathsf{LP}$ in [1] by giving a mechanical procedure that, when given an $\mathsf{S4}$-theorem $\phi$, computes a normal realizer $r$ together with a proof of $\phi^r$ in $\mathsf{LP}$ showing the normality of the realization. A detailed instruction is available in [14].

The direction from (ii) to (i) is relatively trivial. Define *forgetful projection* as the mapping from the justification language to the modal language that replaces each term that is not a proper subterm-occurrence of another by a $\Box$. Observe that the forgetful projection of each axiom of $\mathsf{LP}$ is an axiom of $\mathsf{S4}$. (*MP*) is simply preserved, and (*AN*) corresponds to (*Nec*) applied on an axiom.

The other direction from (i) to (ii) is interesting. Generally speaking, each negative

$\Box$ or non-principal positive $\Box$ is replaced by a variable in the language of $\mathsf{LP}$, and each principal positive $\Box$ is replaced by the sum (terms connected by the "+" operator) of a finite list of "provisional variables." Based on a cut-free sequent calculus of $\mathsf{S4}$, each provisional variable corresponds to an $(R\Box)$ rule in the sequent proof tree, and by using Internalization Lemma (Lemma 40 in Section 2.2, also known as Lifting Lemma) on those rules while induction on the structure of the sequent proof tree, one gets actual $\mathsf{LP}$-terms to replace provisional variables while constructing an $\mathsf{LP}$-proof.

Artemov's procedure was then generalized by Brezhnev [13] to apply on several other modal logic - justification logic pairs including all concerned in the current thesis. In Section 2.2, we will present in very detail a refined proof that has additional properties when applied to theorems in the later defined "loop-free provable" fragment of modal logics.

Other proofs of this theorem include Fitting's methods in [23], [24], and [25], Brunnler et. al.'s algorithmatic method based on nested sequent calculus in [16] (also cf. Goetschi and Kuznets' [32]), Wang's method in [65], and others.

In this thesis, if Theorem 8 holds between modal logic $\mathsf{X}$ and justification logic $\mathsf{Y}$, then we say that $\mathsf{X}$ is the modal logic *corresponding* to $\mathsf{Y}$, and $\mathsf{Y}$ is the justification logic corresponding to $\mathsf{X}$.

## 1.2 The Brouwer–Heyting–Kolmogorov Semantics of Intuitionistic Logics

The first justification logic, Artemov's Logic of Proofs LP, was motivated by the desire of offering a provability semantics to intuitionistic logic (also known as Brouwer–Heyting–Kolmogorov semantics, or BHK semantics) following Gödel's modal embedding. This leads to the Gödel–Artemov approach of formalizing BHK, the only known successful formulation so far.

In this section, we briefly introduce the history of BHK while giving definitions when appropriate.

Intuitionistic logic is the logical basis of the school in philosophy of mathematics called intuitionism. The intuitionism in mathematics dates back into the 19th century, when it was an informal idea from works in mathematics or even science [60]. Brouwer was the first one to gave principles of intuitionism. According to Brouwer, only constructions achieve mathematical objects, and only proofs achieve mathematical truth [15], which is pretty natural given the thoughts of intuitionists that mathematics are purely constructive mental activities of humans. Brouwer's principles were then crystallized independently by Heyting and Kolmogorov, in different terminologies though, and the resulting stipulations are nowadays known after their names as Brouwer-Heyting-Kolmogorov (BHK) Semantics.

Since truth of a formula is interpreted as a proof of it, in order to say the truth of a composite formula in terms of truths of its components, we need to explain a proof of the composite formula in terms of proofs of its components. For the propositional language, the BHK semantics gives stipulations like follows:[2]

[2] Recall that $\neg\phi$ is defined as an abbreviation of $\phi \to \bot$.

- nothing can be a proof of $\bot$;
- a proof of $\phi \wedge \psi$ is the pair of a proof of $\phi$ and a proof of $\psi$;
- a proof of $\phi \vee \psi$ is a proof of $\phi$ or a proof of $\psi$;
- a proof of $\phi \to \psi$ is a construction that returns a proof of $\psi$ whenever a proof of $\phi$ is given.

These stipulations are still informal, with notions like "construction" and "proof" unexplained. Efforts has been devoted to clarify those notations, and give the BHK semantics a formalization.

In the 1930s, Gödel proposed in [30] the modal embedding that faithfully maps intuitionistic propositional logic into an axiomatization of provability (where $Bp$ means "$p$ is provable") that is essentially modal logic $\mathsf{S4}$. Then in [31], which remains unpublished until 1995 in the collection [20], Gödel proposed a way of using a decidable proof relation "$zBp, q$," which means "$z$ is a derivation of $q$ from $p$," instead of "$Bp$" from [30], together with functions over derivations to give a constructive explanation. However, without the provability semantics of $\mathsf{S4}$, and even without a general understanding of modal logic, this thread seemed to go nowhere.

In the 1940s, Kleene [39] proposed his realizability theory, which attempts to achieve intuitionistic truth via computable functions. In Kleene's approach, the predicate

natural number $n$ realizes intuituionistic arithmetical formula $\phi$

is not decidable, making it essentially different from the decidable proof predicate

natural number $n$ is the code of a proof of the formula $\phi$, and hence this approach does not address BHK semantics properly. This approach was accepted for long as a good attempt to formalize BHK semantics, and stipulations of BHK has been introduced by authors in an altered form to accommodate the connection. Since Kleene's realization predicate is not decidable in general, given a realizer of a disjunction, one cannot always mechanically determine which disjunct is realized. In order to compromise, an extra identifier is added to explicitly point out the realized disjunct. This identifier has also been seen in the stipulation of disjunction when introducing BHK semantics by authors (like in [60]).

In the 1960s, Kreisel [43] proposed an approach to formalize BHK vis a theory of construction. The theory was later shown to be inconsistent by Goodman [33] who established a "self-referential paradox," and attempts of fixing the problem do not preserve the relationship with BHK semantics. Despite of this, Kreisel should be credited as the first one who attempts to formalize $\phi \rightarrow \psi$ not only as a construction that returns a proof of $\psi$ when a proof of $\phi$ is given, but also a proof that the construction does work in addition to that.

Finally in 1995, Artemov [1] successfully finished the previously mentioned approach initialed by Gödel [30]. This was achieved by offering the Logic of Proofs $\mathsf{LP}$ (one of the justification logics in Definition 1) that serves as an intermedium via which $\mathsf{S4}$ reaches its provability semantics. The idea of using the decidable proof predicate $t{:}\phi$ (meaning "$t$ is a proof of $\phi$") can be seen as echoing Gödel's $zBp,q$ in [31], although Artemov's [1] was finished without knowing Gödel's [31] that appeared in the same year. This approach is now known as Gödel–Artemov approach, which will be introduced in more detail in the rest of this section.

The Gödel–Artemov approach can generally be displayed as:

$$\mathsf{IPC} \hookrightarrow \mathsf{S4} \hookrightarrow \mathsf{LP} \hookrightarrow \mathsf{PA}. \tag{1.1}$$

In this approach, a formula is a theorem of intuitionistic propositional logic IPC iff its embedding image is a theorem of S4, iff its embedding image has some realization that is a theorem of LP under some constant specification, iff this LP-theorem is interpreted as a Peano Arithmetic PA-theorem under any interpretation that admits this constant specification. Note that "realization" here (cf. Definition 7) has a meaning different from that it has in Kleene's realizability theory. Followed this approach, we can read each modality $\Box$ in S4 as "there is an explicit proof of." This reading is different from the reading of $\Box$ as "there is a proof of" in Gödel-Löb provability logic GL. In GL, $\Box\phi$ is understood as a $\Sigma_1$-sentence that there exists an $x$ that is a proof of (the interpretation of) $\phi$. In Gödel–Artemov approach, as we will see later, $\Box\phi$ will be interpreted as a decidable assertion where $\Box$ is assigned a term $t$ that explicitly displays the structure of the proof of (the interpretation of) $\phi$. This approach can be explored further by considering the first-order case, though provability semantics in first-order case are not recursively axiomatizable, as shown by Artemov and Yavorskaya (Sidon) in [8].

The first step in (1.1) is from IPC to S4.

The intuitionistic propositional logic IPC was gradually developed by Kolmogorov [41], Glivenko [29], Heyting (who is mostly acknowledged) [35], and others. Though each of them attempted to provide a formal proof system that adopts Brouwer's thoughts, non-equivalent axiomatizations were proposed because of differences in understanding Brouwer. Now decades later, by IPC, people usually mean a modern agreed axiomatization. In the current thesis, we adopt the following:

**Definition 9 (Intuitionistic Propositional Logic IPC [63])** The language is defined by

$$\phi ::= \bot \mid p \mid \phi \wedge \phi \mid \phi \vee \phi \mid \phi \to \phi.$$

$MP$ is the only inference rule, and axiom schemes are listed as follows, sorted in groups:

$$\bot \to \phi$$

$$\phi \to (\psi \to \phi) \qquad (\phi \to (\psi \to \chi)) \to ((\phi \to \psi) \to (\phi \to \chi))$$

$$\phi \wedge \psi \to \phi \qquad \phi \wedge \psi \to \psi \qquad \phi \to (\psi \to \phi \wedge \psi)$$

$$\phi \to \phi \vee \psi \qquad \psi \to \phi \vee \psi \qquad (\phi \to \chi) \to ((\psi \to \chi) \to (\phi \vee \psi \to \chi)).$$

$\dashv$

Extending IPC by allowing first-order quantifiers gives the intuitionistic first-order predicate logic IQC [40]. Adding appropriate clauses for quantifies into the BHK semantics of IPC gives that of IQC (cf. [9]).

Proposed by Gödel [30] and verified by McKinsey and Tarski [49], one can faithfully embed IPC into S4, by prefixing each subformula with a $\Box$, or more formally, by the embedding $(\cdot)^{\triangle}$ defined as:

$$\begin{cases} p^{\triangle} & := & \Box p; \\ \bot^{\triangle} & := & \Box\bot; \\ (\phi \odot \psi)^{\triangle} & := & \Box(\phi^{\triangle} \odot \psi^{\triangle}) \ \text{ for } \odot \in \{\wedge, \vee, \to\}. \end{cases}$$

**Theorem 10 (A Faithful Embedding of IPC into S4 [49])** *Let $\phi$ be a propositional formula. Then* IPC $\vdash \phi$ *iff* S4 $\vdash \phi^{\triangle}$.

This theorem was first proof by McKinsey and Tarski with algebraic method in [49]. Alternatively one can prove this via cut-free sequent calculus, as in [63].

This constitutes the first step from IPC to S4 in (1.1), i.e., with the help of Theorem 10, the provability semantics of S4 offers a natural provability semantics of IPC [2]. Note that there are many "similar" embeddings of IPC into S4 [63], and some of them will be considered in Section 4.1. We preserve $(\cdot)^{\triangle}$ for the "$\Box$ each subformula" embedding, and use $(\cdot)^{\times}$ for a general embedding. Other specific interesting embeddings will also be given notations in Section 4.1. Each of this embeddings actually gives a provability semantics of IPC, but only $(\cdot)^{\triangle}$ inductively meets the thought of "truth as provability."

The second step in (1.1) from S4 to LP has been presented in Section 1.1 (cf. Definition 7 and Theorem 8).

Combining Theorem 10 and Theorem 8, we have:

**Corollary 11** *For any propositional formula $\phi$:* IPC $\vdash \phi$ *iff* LP $\vdash \phi^{\triangle^r}$ *for some normal realization* $\phi^{\triangle^r}$.

Similar results hold, if we take other faithful embeddings from IPC to S4 instead of $(\cdot)^{\triangle}$.

The last step in (1.1) is from LP to formal proofs in PA.

Playing a rule here is the *provability semantics* of LP [2]. Let $Prf(\cdot,\cdot)$ be a $\Delta_1$-formula s.t.:

- for each arithmetic sentence $\phi$, PA $\vdash \phi$ iff $Prf(\overline{n}, \overline{\ulcorner\phi\urcorner})$ for some natural number $n$;
- for each natural number $k$, the set $T(k) = \{l \,|\, Prf(\overline{k}, \overline{l})\}$ is finite and its code is computable;
- for any natural numbers $k, l$, there is a natural number $n$ s.t. $T(k) \cup T(l) \subseteq T(n)$.

Then there are computable functions $\mathbf{m}(\cdot,\cdot)$, $\mathbf{a}(\cdot,\cdot)$, and $\mathbf{c}(\cdot)$, s.t. for any arithmetical formulas $\phi, \psi$, any natural numbers $k, n$, all of

- $Prf(\overline{k}, \overline{\ulcorner\phi\to\psi\urcorner}) \wedge Prf(\overline{n}, \overline{\ulcorner\phi\urcorner}) \to Prf(\overline{\mathbf{m}(k,n)}, \overline{\ulcorner\psi\urcorner})$,
- $(Prf(\overline{k}, \overline{\ulcorner\phi\urcorner}) \to Prf(\overline{\mathbf{a}(k,n)}, \overline{\ulcorner\phi\urcorner})) \wedge (Prf(\overline{n}, \overline{\ulcorner\phi\urcorner}) \to Prf(\overline{\mathbf{a}(k,n)}, \overline{\ulcorner\phi\urcorner}))$, and

- $Prf(\overline{k}, \overline{\ulcorner\phi\urcorner}) \to Prf(\overline{\mathbf{c}(k)}, \overline{\ulcorner Prf(\overline{k}, \overline{\ulcorner\phi\urcorner})\urcorner})$

are valid. An *arithmetical interpretation* (with the choice of $Prf$ and its appropriate functions $\mathbf{m}, \mathbf{a}, \mathbf{c}$ as parameters) assigns each propositional atomic formula an arithmetic sentence, assigns each atomic term a natural number, commutes with Boolean operations, and satisfies:

$$(t \cdot s)^* = \mathbf{m}(t^*, s^*), \qquad (t + s)^* = \mathbf{a}(t^*, s^*), \qquad (!t)^* = \mathbf{c}(t^*),$$

and

$$(t{:}\phi)^* = Prf(\overline{t^*}, \overline{\ulcorner\phi^*\urcorner}) \ .$$

For each constant specification $\mathcal{CS}$, an interpretation $*$ is a *$\mathcal{CS}$-interpretation*, if under $*$, all formulas in $\mathcal{CS}$ are $\mathsf{PA}$-theorems. An $\mathsf{LP}$-formula $\phi$ is *provably valid under constant specification* $\mathcal{CS}$, if $\mathsf{PA} \vdash \phi^*$ for any $\mathcal{CS}$-interpretation $*$.

**Theorem 12 (Arithmetical Completeness of $\mathsf{LP}$ [2])** *For any $\mathsf{LP}$-formula $\phi$:*

*(A) $\mathsf{LP}(\mathcal{CS}) \vdash \phi$ iff $\phi$ is provably valid under $\mathcal{CS}$;*

*(B) $\mathsf{LP} \vdash \phi$ iff $\phi$ is provably valid under some constant specification.*

This constitutes the last step in (1.1), and with the help of Theorem 8 and Theorem 10, we have the following two theorems.

**Theorem 13 (Arithmetical Completeness of $\mathsf{S4}$ [2])** *For any modal formula $\phi$: $\mathsf{S4} \vdash \phi$ iff $\phi^r$ is provably valid under some constant specification, for some normal realization $\phi^r$.*

**Theorem 14 (Arithmetical Completeness of $\mathsf{IPC}$ [2])** *For any faithful embedding $(\cdot)^\times$ and any propositional formula $\phi$: $\mathsf{IPC} \vdash \phi$ iff $\phi^{\times r}$ is provably valid under some constant specification, for some normal realization $\phi^{\times r}$.*

This finishes our description of the Gödel–Artemov approach of formalizing BHK semantics shown in (1.1).

## 1.3 Self-referentiality in Justification Logics

In the arithmetical semantics of $\mathsf{LP}$, a proof predicate $Prf(\cdot,\cdot)$ that satisfies certain requirements is employed as a parameter [2]. As we listed in Section 1.2, those requirements only includes: being a $\Delta_1$-formula, each proof proves a finite computable set of theorems, and the conjoinability of proofs. All those requirements are natural, but not all natural properties are required, which leaves spaces to explore. For instance, Artemov considered the symmetry of choice of proofs and introduced Symmetric Logic of Proofs in [3]. For an even earlier instance, Artemov and Stra$\beta$en tried to capture the Gödel proof predicate in [7].

Recall the language of justification logics (cf. Definition 1). The language admits formulas of the form $t\!:\!\phi$, where $\phi$ is a (sub)formula by itself and may contain occurrences of terms. Even the very term $t$ may occur in $\phi$, and hence we have a legal formula of the form $t\!:\!\phi(t)$. Formulas like this is said to be *self-referential*.

The same may happen when $t$ is a constant and $\phi$ is an axiom. Thus we have the following definition in all the five logics in Definition 1, where our notational Conversion 2 for "!" in $\mathsf{J}$, $\mathsf{JD}$, and $\mathsf{JT}$ applies.

**Definition 15 (Self-referentiality of Constant Specification [45, 46])** A constant specification $\mathcal{CS}$ is

(i) *self-referential*, if it has a subset of the form

$$\{\, c_1\!:\!A_1(c_2), \cdots, c_{n-1}\!:\!A_{n-1}(c_n), c_n\!:\!A_n(c_1)\,\}.$$

(ii) *directly self-referential*, if it has a formula of the form $c\!:\!A(c)$. $\dashv$

Given a self-referential formula $t\!:\!\phi(t)$, i.e., the formula $\phi$ which contains $t$ is prefixed by (another occurrence of) $t$. According to the arithmetical semantics of LP,

$$(t\!:\!\phi(t))^* = Prf(\overline{t^*}, \overline{\ulcorner \phi(t)^* \urcorner}) \ ,$$

that is, $t^*$ is a code of a proof of $(\phi(t))^*$. An observation indicates that this may not be consistent with the Gödel proof predicate. Actually, a natural coding of arithmetical formulas is defined inductively on parse tree of formulas. One plugs codes of subformulas into fixed computable functions associated to, e.g., connectives, and assigns the computed output to a formula as its code. Usually, this will lead to a property that the code of a formula is always larger than that of its proper subformulas. In our example, this often indicates

$$t^* < \ulcorner \phi(t)^* \urcorner. \tag{1.2}$$

On the other hand, the code of a proof is usually defined as the code of a list of formulas or that of a tree of formulas, depending on how the notion of "proof" is defined. This is to say, we have $x \nless y$ if $Prf(\overline{x}, \overline{y})$, which means, if $Prf(\overline{t^*}, \overline{\ulcorner \phi(t)^* \urcorner})$ holds, then so is

$$t^* \nless \ulcorner \phi(t)^* \urcorner \ ,$$

which contradicts (1.2). Therefore, an arithmetical interpretation with a natural proof predicate is likely to falsify self-referential formulas.

The arithmetical semantics of LP, as we mentioned in Section 1.2, captures a wide class of arithmetical interpretations by employing a proof predicate and three computable functions that satisfy the stipulations as parameters. This class contains combinations that are mostly intended (like what we just considered as "natural"), but also contains many others.

Artemov mentioned in [2] that the procedure given there to prove the realization theorem of S4 (in LP) may lead to (directly) self-referential constant specifications, and asked whether this is necessary. This is important, since according to the arithmetical completeness (Theorem 12), if a modal formula is realized into $\mathsf{LP}$ with a constant specification $\mathcal{CS}$, then the realized formula becomes a $\mathsf{PA}$-theorem under any interpretation that admits $\mathcal{CS}$. If a self-referential formula falsified by any natural interpretation is involved in $\mathcal{CS}$, then the realized formula is a $\mathsf{PA}$-theorem only under unnatural interpretations.

Self-referential formulas are syntactical objects, and hence express self-referential meanings in all various kinds of semantics. The arithmetical semantics we just mentioned is for $\mathsf{LP}$ only, and there are other semantics that work for some or all logics in Definition 1. Within them, Mkrtychev semantics [50] is the first that has built in syntactical component, called evidence function; Fitting semantics [23] is a multi-state version of Mkrtychev semantics, and works for all logics in Definition 1. Formal definitions of these semantics are given in Section 2.1. Artemov's modular semantics, though mathematically similar to Fitting semantics, suggests a better understanding of the ontology of justification in a logic setting [5]. In Mkrtychev or Fitting semantics, the self-referential formula $t\!:\!\phi(t)$ means $t$ justifies the assertion $\phi$ about $t$ itself (on a state, if in the Fitting semantics). In modular semantics that interprets a (justification) term on a state as the set of formulas justified by it on that state, $t\!:\!\phi(t)$ means $t$ is a set containing the formula $\phi(t)$ (on a state). Encountering $t\!:\!\phi(t)$, there is then no way to reduce the term $t$ to an expression free of $t$.

# Chapter 2

# Self-referentiality in Modal Logics

This chapter is about self-referentiality in modal logics in terms of their realization in justification logics.

It starts by Section 2.1 presenting Kuznets' works [14, 45, 46] on self-referentiality. Generally speaking, Kuznets' works show that: (1) each of $\mathsf{T}$, $\mathsf{K4}$, and $\mathsf{S4}$ has theorems whose realizations in $\mathsf{JT}$, $\mathsf{J4}$, and $\mathsf{LP}$ respectively necessarily call for directly self-referential constant specifications, and (2) $\mathsf{K}$ and $\mathsf{D}$ are both free of such theorems.

Then Section 2.2 carries the candidate's earlier (as a master student in Tsinghua University) work on refining Kuznets' result in modal logic $\mathsf{S4}$, presented uniformly with its later (since the candidate became a student in the Graduate Center, City University of New York) extension to both $\mathsf{T}$ and $\mathsf{K4}$. A decidable necessary condition for self-referentiality in realization is formulated in that section, and a refinement of Artemov's realization procedure (in order to prove Theorem 8 for $\mathsf{LP}$ in Artemov's [2]) is presented in detail.

## 2.1 Kuznets' Works on the Logic Level

In Section 1.3, we discussed the meaning of self-referential formulas, e.g., $t : \phi(t)$, in semantics of LP, and Artemov's question that whether such formulas are necessary in the realization of S4 in LP. One might think that the appearance of self-referential formulas is due to the realization procedure used in [2], which in principle does not cover all possibilities of realizing a specified formula, and other realization method like those in [24], [16], and [65], may help. This is not true, as Kuznets showed in [14] that self-referentiality is intrinsic in the realization of S4 in LP.

**Theorem 16 (Direct Self-referentiality of S4 [14])** *Any realization of the S4-theorem*

$$\neg\Box\neg(p \to \Box p) \tag{2.1}$$

*calls for a directly self-referential constant specification.*

**Remark:** The transitivity axiom $\Box\phi \to \Box\Box\phi$ in S4 can be read in an epistemic perspective as "if the agent knows $\phi$, then he also knows the fact that he knows $\phi$." In this way, together with the reading of $\Box(\phi \to \psi) \to (\Box\phi \to \Box\psi)$ as "the agent's knowledge of the succedent follows from that of the antecedent and that of the implication," and the reading of $\Box\phi \to \phi$ as "only true facts can be known by the agent," S4 becomes a (single agent) epistemic logic of knowledge with positive introspection. Taking an equivalent (under S4) form of (2.1),

$$\neg\Box(p \wedge \neg\Box p),$$

one can read this principle as "$M$ is not known," where $M$ stands for the Moore sentence $p \wedge \neg\Box p$, meaning "$p$ is true but not known." Hence the fact that (2.1) is an S4-theorem tells us that if positive introspection of knowledge is assumed, then Moore sentence is not knowable. The above observation has given an indication that self-referentiality may be related to interesting epistemic principles. However, this trend is diluted, as more self-referential examples (in Section 4.2) are found later.

Back to Kuznets' theorem, we offer an example here, displaying a realization of (2.1) via the realization procedure in [2] with a directly self-referential constant specification. Since the procedure involves many heavy computations, like converting a sequent proof to a Hilbert-style axiomatic proof, a step-by-step presentation is not only hard to read, but also fails to give a good intuition. Yet the realization is quite sensitive to the computation, and hence a sketch will look like a list of expressions with no obvious relationship between them. Therefore, we only display the final result, and omit all computations.

**Example 17** *The* S4*-theorem* $\neg\Box\neg(p \to \Box p)$ *has two* $\Box$*'s, one negative and one positive. The* LP*-theorem* $\neg x : \neg(p \to c_2 \cdot (c_1 \cdot !x) : p)$ *is a realization of it. An example proof of this* LP*-theorem calls for the constant specification*

$$\left\{ \begin{array}{l} c_1 : [x : \neg(p \to c_2 \cdot (c_1 \cdot !x) : p) \to \neg(p \to c_2 \cdot (c_1 \cdot !x) : p)], \\ c_2 : [\neg(p \to c_2 \cdot (c_1 \cdot !x) : p) \to p] \end{array} \right\}$$

*that is directly self-referential.*[1]

Note that the realization procedure is sensitive to propositional axioms of LP, to the way we treat negative $\Box$'s, to proof-search algorithm for S4, to algorithm used to convert a sequent proof into a Hilbert-style proof, etc., and hence the reader is likely to get a different result after computing by himself. A beauty of Theorem 16 is, its method is not sensitive to these choices, and even not sensitive to the general method of realizing S4-theorems. Theorem 16 turns down all possibilities of realizing (2.1) while avoiding direct self-referentiality.

[1]Note that both $c_1$ and $c_2$ occur in the axiom prefixed by the other, and hence we also have a reference cycle of length 2.

The proof of Theorem 16 employs a construction of a counter Mkrtychev model that admits the largest constant specification that is not directly self-referential, of any potential realization of (2.1). This method was generalized in Kuznets' later works [45, 46], where Fitting semantics is employed instead of Mkrtychev semantics, to get similar results for both $\mathsf{T}$ and $\mathsf{K4}$. Also proved in [45, 46] is the fact that both $\mathsf{K}$ and $\mathsf{D}$ do not enjoy such results.

It is helpful here to have the following definitions.

**Definition 18 (Self-referentiality of Modal Theorems)** A theorem of a modal logic is *self-referential* (*directly self-referential*), if each of its realizations in the corresponding justification logic necessarily calls for a constant specification that is self-referential (directly self-referential). $\dashv$

**Definition 19 (Self-referentiality of Modal Logics)** If a modal logic has a theorem that is self-referential (directly self-referential), then the logic itself is said to be *self-referential* (*directly self-referential*). $\dashv$

With the above definitions, we can summarize Kuznets' results in [45, 46] as

**Theorem 20 ([45, 46])** *(A) Both* $\mathsf{T}$ *and* $\mathsf{K4}$ *are directly self-referential.*

*(B) Both* $\mathsf{K}$ *and* $\mathsf{D}$ *are not self-referential.*

Self-referential theorems found by Kuznets for $\mathsf{T}$, $\mathsf{K4}$ are presented in Table 2.1 that summarizes Kuznets' results in [14, 45, 46].

For non-self-referential facts, Kuznets [45, 46] offers a refinement of Artemov's realization procedure, and hence gives an instruction following which all $\mathsf{K}$-theorem ($\mathsf{D}$-theorem) can be realized in $\mathsf{J}$ ($\mathsf{JD}$, resp.) non-self-referentially. For self-referential facts,

Table 2.1: A Summary of Kuznets' Results.

| | Modal mate | Self-referentiality | Instance |
|---|---|---|---|
| J | K | No | N.A. |
| JD | D | No | N.A. |
| JT | T | Directly | $\neg\Box\neg(p \to \Box p)$ |
| J4 | K4 | Directly | $\Box\neg(p \to \Box p) \to \Box\bot$ |
| S4 | K4 | Directly | $\neg\Box\neg(p \to \Box p)$ |

as we mentioned above, Fitting semantics is employed. As this semantics will be employed by our later sections, we present it here formally. Compared to other works like [23] and [46], our notations as presented below are slightly different. Employing such notations makes it easier to present our proofs in Section 3.3 and Section 4.2.

**Definition 21 (Evidence Function [50])** A unary function $*(\cdot)$ from the set of terms to the power set of formulas is called an *evidence function* if it satisfies the following closure conditions:

$$\begin{cases} \psi \in *(t) \text{ implies } (t\!:\!\psi) \in *(!t) & \text{for J4 and LP only} \\ \tau \to \phi \in *(t_1) \text{ and } \tau \in *(t_2) \text{ implies } \phi \in *(t_1 \cdot t_2) \\ *(t_1) \cup *(t_2) \subseteq *(t_1 + t_2). \end{cases}$$

An evidence function $*$ is *full*, if $\phi \in *(t)$ holds for any term $t$ and formula $\phi$.

For a set $X$ of formulas of the form $t\!:\!\phi$ and an evidence function $*$, if $\phi \in *(t)$ holds for any $t\!:\!\phi \in X$, then we say that $*$ *admits* $X$, also, that $*$ is an *$X$-evidence function*. This terminology is frequently used when $X$ is a constant specification.

Given a set $X$ of formulas of the form $t : \phi$, by $*_X$, we mean the closure of $X$ under these conditions, which is automatically an $X$-evidence function. $X$ is called the *initial set* of $*_X$. If $\phi \in *(t)$ is the case since $t\!:\!\phi \in X$, then $\phi \in *(t)$ is said to be achieved *via initial set*. If $\phi \in *(t)$ is the case but $t\!:\!\phi \notin X$, then $\phi \in *(t)$ is said to be achieved *via closure*. $\dashv$

**Lemma 22 ([50])** *$*_X$ is the smallest $X$-evidence function. That is, if $*'$ is an $X$-evidence* *function, then for any term $s$ and formula $\phi$, we have $\phi \in *_X(s)$ implies $\phi \in *'(s)$.*

**Definition 23 (Fitting Model [23])** A *Fitting model* is a tuple $\mathfrak{M} = (W, R, E, V)$, where $(W, R, V)$ is a Kripke model in the standard setting i.e., $W$ is a non-empty set of *states*, $R$ is a binary *accessible relation* on $W$, and $V$ is a propositional valuation. $R$ must be serial in case of $\mathsf{JD}$, be reflexive in cases of $\mathsf{JT}$ and $\mathsf{LP}$, and be transitive in cases of $\mathsf{J4}$ and $\mathsf{LP}$. $E$ is an *evidence function assignment* that assigns to each state $w \in W$ an evidence function (usually denoted by $*^w$). In cases of $\mathsf{J4}$ and $\mathsf{LP}$, the property of *monotonicity* is further required. That is, if $uRv$, then $*^u(t) \subseteq *^v(t)$ for any term $t$.

An evidence function assignment $E$ is a $\mathcal{CS}$*-evidence function assignment*, if it assigns to each state a $\mathcal{CS}$-evidence function.

If $E$ is a $\mathcal{CS}$-evidence function assignment, then $\mathfrak{M} = (W, R, E, V)$ is called a $\mathcal{CS}$-*model*.

For $\mathfrak{M} = (W, R, E, V)$, by $\mathfrak{M}, w \vDash \phi$, we mean $\mathfrak{M}$ *satisfies* $\phi$ on $w$. This is defined by:

(i) $\mathfrak{M}, w \nvDash \bot$,

(ii) $\mathfrak{M}, w \vDash p$ iff $w \in V(p)$,

(iii) $\mathfrak{M}, w \vDash \phi \rightarrow \psi$ iff $\mathfrak{M}, w \nvDash \phi$ or $\mathfrak{M}, w \vDash \psi$,

(iv) $\mathfrak{M}, w \vDash t{:}\phi$, iff $\mathfrak{M}, v \vDash \phi$ for each $v$ s.t. $wRv$, and $\phi \in E(w, t)$. $\dashv$

**Theorem 24 (Fitting Completeness [23])** *Let* $\mathsf{X}$ *range from* $\{\mathsf{J}, \mathsf{JD}, \mathsf{JT}, \mathsf{J4}, \mathsf{LP}\}$*, then for any justification formula $\phi$:*

$\mathsf{X}(\mathcal{CS}) \vdash \phi$ *iff* $\mathfrak{M}, w \vDash \phi$ *for any* $\mathcal{CS}$*-model* $\mathfrak{M}$ *and state* $w$.

For $\mathsf{LP}$, we have the completeness w.r.t. Mkrtychev semantics [50], which is discovered earlier and mathematically simpler than Fitting semantics. With Definition 23 of Fitting models, we can formulate Mkrtychev semantics as follows.

**Definition 25 (Mkrtychev Model [50])** A Mkrtychev model for $\mathsf{LP}$ is a Fitting model for $\mathsf{LP}$ with only one reflexive state.

More formally, a Mkrtychev model is a pair $\mathfrak{M} = (*, V)$ where $*$ is an evidence function (as in Definition 21), and $V$ is a propositional valuation.

If $*$ is a $\mathcal{CS}$-evidence function, then $\mathfrak{M} = (*, V)$ is called a $\mathcal{CS}$*-model*.

By $\mathfrak{M} \models \phi$, we mean that $\mathfrak{M}$ *satisfies* $\phi$. This is defined by:

(i) $\mathfrak{M} \not\models \bot$,

(ii) $\mathfrak{M} \models p$ iff $V(p) = 1$,

(iii) $\mathfrak{M} \models \phi \rightarrow \psi$ iff $\mathfrak{M} \not\models \phi$ or $\mathfrak{M} \models \psi$,

(iv) $\mathfrak{M} \models t{:}\phi$, iff $\mathfrak{M} \models \phi$ and $\phi \in *(t)$. $\dashv$

**Theorem 26 (Mkrtychev Completeness [50])** *For any* $\mathsf{LP}$*-formula* $\phi$*:*

$\mathsf{LP}(\mathcal{CS}) \vdash \phi$ *iff* $\mathfrak{M} \models \phi$ *for any* $\mathcal{CS}$*-model* $\mathfrak{M}$*.*

Note that completeness here are relativized to specific constant specifications. In order to show that a justification formulas is not provable in a logic armed by a constant specification, it is sufficient to construct a counter model that admits that constant specification, of the formula.

Fitting (Mkrtychev, modular) models, as semantic objects, involve syntactical contents in the evidence function. This gives those models a somehow direct way to control justifications in a logic. It is well-agreed that this kind of semantics is much more wieldy than, e.g., arithmetical semantics.

## 2.2 A Necessary Condition for Self-referentiality of Theorems

This Section contains materials reused from the candidate's published paper [72], with the kind permission from Springer, the owner of the copyright of [72].

A notational remark: in this section only, $R^+$ ($R^*$) stands for the transitive (reflexive and transitive, respectively) closure of the binary relation $R$.

In Section 2.1, we introduced Kuznets results about self-referentiality of modal logics T, K4, and S4. Those results can be seen as at a logic-level, i.e., which modal logic is self-referential and which is not.

The candidate's earlier (before becoming a student in the Graduate Center, City University of New York) work [68, 69] considered this topic at the theorem-level, i.e., which S4-theorems need self-referentiality while being realized and which do not. The notion of *prehistoric loop* was defined in a Gentzen-style sequent calculus of S4, and it was shown that an S4-theorem is self-referential only if it does not have any proof that is prehistoric-loop-free in that calculus.

Being an extension of [68], the candidate's recent [72] performs the same idea in both T and K4, the two (known) minimal logics that are self-referential. Both T and K4, together with S4, is treated uniformly.

### 2.2.1 G3-style Sequent Calculi

Sequent calculi were invented by Gentzen [27], and later used in the proof theory of modal logic by authors like Ohnishi and Matsumoto [53, 54]. Gentzen's original idea have also been improved and refined by authors like Ketonen, Kleene, Dragalin, Troelstra, Negri, and von Plato, in a direction that leads to the G3-style sequent calculi, which generally "absorb" structural rules like Weakening and Contraction into other rules. Calculus with G3-style flavor for modal logic may be traced back to Kanger [37]. Current formulations are available in [63] and [52].

In [63], a sequent calculus of S4, called G3s, was presented, and then shown to be depth-preserved weakening admissible, depth-preserved contraction admissible, and cut admissible. Those properties make it a nice calculus based on which realization procedure works, like in [14]. We include similar calculi for K4 and T in the coming definition.

**Definition 27 (G3-style Calculi)** Consider modal multi-set sequent calculus with rules:[2]

$$\frac{}{p,\Gamma\Rightarrow\Delta,p}\ (p \text{ is an atom})\quad (Ax) \qquad \frac{}{\bot,\Gamma\Rightarrow\Delta}\quad (L\bot)$$

$$\frac{\Gamma\Rightarrow\Delta,\alpha\quad \beta,\Gamma\Rightarrow\Delta}{\alpha\to\beta,\Gamma\Rightarrow\Delta}\ (L\to) \qquad \frac{\alpha,\Gamma\Rightarrow\Delta,\beta}{\Gamma\Rightarrow\Delta,\alpha\to\beta}\ (R\to)$$

$$\frac{\theta,\Box\theta,\Gamma\Rightarrow\Delta}{\Box\theta,\Gamma\Rightarrow\Delta}\ (L\Box) \qquad \frac{\Box\Theta\Rightarrow\eta}{\Box\Theta,\Gamma\Rightarrow\Delta,\Box\eta}\ (R\Box)$$

$$\frac{\Theta,\Box\Theta\Rightarrow\eta}{\Box\Theta,\Gamma\Rightarrow\Delta,\Box\eta}\ (4\Box) \qquad \frac{\Theta\Rightarrow\eta}{\Box\Theta,\Gamma\Rightarrow\Delta,\Box\eta}\ (K\Box)$$

In $(Ax)$, $(L\bot)$: $p,\bot$ are called principal formulas, and (formulas in) $\Gamma,\Delta$ are called weakening[3] formulas. In the conclusion of $(L\to)$, $(R\to)$, $(L\Box)$: $\alpha\to\beta,\Box\theta$ are principal, and $\Gamma,\Delta$ are side. In the conclusion of $(K\Box)$ and $(4\Box)$: $\Box\eta,\Box\Theta$ are principal, and $\Gamma,\Delta$ are weakening. In the conclusion of $(R\Box)$: $\Box\Theta$ is side, $\Box\eta$ is principal, and $\Gamma,\Delta$ are weakening. The distinction of principal, side, and weakening formulas is summarized in Table 2.2.

By *correspondence*, we mean the relation between, for instance, the displayed $\alpha$ occurrence in the left premise of $(L\to)$ and the displayed $\alpha$ occurrence (as a subformula of $\alpha\to\beta$) in the conclusion of $(L\to)$. Each occurrence in the premise of a rule has

[2]In what follows, $\Gamma,\Delta$, ..., are multi-sets of formulas, and $\Box\Gamma$ is an abbreviation of $\{\Box\gamma\,|\,\gamma\in\Gamma\}$. For notational convenience, we may omit some "," in sequents. Thus $\Gamma\Delta$ stands for $\Gamma\cup\Delta$, and $\Gamma\alpha$ stans for $\Gamma\cup\{\alpha\}$, etc..

[3]In standard textbooks like [63], $\Gamma,\Delta$ are called side formulas. Since we will distinguish side formulas and weakening formulas in modal rules, and $\Gamma,\Delta$ here behave closely to weakening formulas in modal rules, we call them "weakening" formulas.

Table 2.2: Principal, Side, and Weakening Formulas in Rules.

| | Principal | Side | Weakening | | Principal | Side | Weakening |
|---|---|---|---|---|---|---|---|
| $(Ax)$ | $p \Rightarrow p$ | $\times$ | $\Gamma \Rightarrow \Delta$ | $(L\Box)$ | $\Box\theta \Rightarrow$ | $\Gamma \Rightarrow \Delta$ | $\times$ |
| $(L\bot)$ | $\bot \Rightarrow$ | $\times$ | $\Gamma \Rightarrow \Delta$ | $(R\Box)$ | $\Rightarrow \Box\eta$ | $\Box\Theta \Rightarrow$ | $\Gamma \Rightarrow \Delta$ |
| $(L\rightarrow)$ | $\alpha\rightarrow\beta \Rightarrow$ | $\Gamma \Rightarrow \Delta$ | $\times$ | $(K\Box)$ | $\Box\Theta \Rightarrow \Box\eta$ | $\times$ | $\Gamma \Rightarrow \Delta$ |
| $(R\rightarrow)$ | $\Rightarrow \alpha\rightarrow\beta$ | $\Gamma \Rightarrow \Delta$ | $\times$ | $(4\Box)$ | $\Box\Theta \Rightarrow \Box\eta$ | $\times$ | $\Gamma \Rightarrow \Delta$ |

exactly one corresponding occurrence in the conclusion of that rule. An occurrence in the conclusion may have two corresponding occurrences in premise(s). For example, the displayed $\theta$ occurrence (as a subformula of $\Box\theta$) in the conclusion of $(L\Box)$ has the two displayed $\theta$ occurrences (one by its own, the other as a subformula of $\Box\theta$) in the premise as its corresponding occurrences. For another example, a formula $\delta$ in $\Delta$ in the conclusion of $(L\rightarrow)$ has two corresponding occurrences, one in the displayed $\Delta$ in the left premise, and the other in the displayed $\Delta$ in the right premise. The phenomenon that two occurrences in premise(s) share a same corresponding occurrence in the conclusion is called a *unification*. These terminologies are assumed to be generalized to structures defined on occurrences in later subsections.

As in [63], we have in modal language that:

$$\begin{aligned}\mathsf{G3cp} &= \{(Ax), (L\bot), (L\rightarrow), (R\rightarrow)\},\\ \mathsf{G3s} &= \mathsf{G3cp} \cup \{(L\Box), (R\Box)\}.\end{aligned}$$

Precisely, e.g., $\mathsf{G3cp}$ is the set of sequents closed under $\{(Ax), (L\bot), (L\rightarrow), (R\rightarrow)\}$. Also, we have:

$$\begin{aligned}\mathsf{G3t} &= \mathsf{G3cp} \cup \{(L\Box), (K\Box)\},\\ \mathsf{G34} &= \mathsf{G3cp} \cup \{(4\Box)\}.\end{aligned}$$

⊣

For each rule $(R)$, by $\overrightarrow{(R)}$, we mean several (possibly 0) continuous applications of $(R)$. Following [63], by $\mathsf{G3[st4]}$, we mean a calculus $\mathsf{X} \in \{\mathsf{G3s}, \mathsf{G3t}, \mathsf{G34}\}$. Similarly, if more, or less calculi are involved.

By G3[st4] $\vdash_n \Gamma \Rightarrow \Delta$ (or G3[st4] $\vdash_{<n} \Gamma \Rightarrow \Delta$, resp.), we mean that sequent $\Gamma \Rightarrow \Delta$ is derivable in G3[st4] by a proof tree of height $\leq n$ (or $< n$, resp.). G3[st4] $\vdash \Gamma \Rightarrow \Delta$ means G3[st4] $\vdash_n \Gamma \Rightarrow \Delta$ for some natural number $n$.

In the rest of this subsection, we will establish some standard results for those calculi defined above. Those results are depth-preserved-admissibility ("DPA" for short) of Weakening, DPA of Contraction, and Cut-elimination. In [63, Section 3.5, 4.1], those results are proved for G3c, a calculus extending G3cp in the first order language instead of propositional modal language. It is easy to see that the proof there offers all we need for G3cp. Then in [63, Section 9.1], a similar method is designed for G3s. Similar methods work for G3t and G34.

**Theorem 28 (DPA of Weakening)** *For any sets of formulas $\Gamma$, $\Gamma'$, $\Delta$, and $\Delta'$:*

*If* G3[st4] $\vdash_n \Gamma \Rightarrow \Delta$, *then* G3[st4] $\vdash_n \Gamma', \Gamma \Rightarrow \Delta, \Delta'$.

**Proof.** An easy induction on $n$. ⊣

**Theorem 29 (Depth-preserved Inversion Lemma)** *For any $\alpha$, $\beta$, $\theta$, $\Gamma$, and $\Delta$:*

*(A)* G3[st4] $\vdash_n \alpha \to \beta, \Gamma \Rightarrow \Delta$ *implies* G3[st4] $\vdash_n \Gamma \Rightarrow \Delta, \alpha$ *and* G3[st4] $\vdash_n \beta, \Gamma \Rightarrow \Delta$.

*(B) If* G3[st4] $\vdash_n \Gamma \Rightarrow \Delta, \alpha \to \beta$ *then* G3[st4] $\vdash_n \alpha, \Gamma \Rightarrow \Delta, \beta$.

*(C) If* G3[st] $\vdash_n \Box\theta, \Gamma \Rightarrow \Delta$ *then* G3[st] $\vdash_n \theta, \Box\theta, \Gamma \Rightarrow \Delta$.

**Proof.** (A) We generate the method of [63] for G3cp to accommodate modal rules.

When $n = 0$: $\alpha \to \beta, \Gamma \Rightarrow \Delta$ is an instance of $(Ax)$ or $(L\bot)$. This case is trivial.

Induction step: Consider $(R)$, the last rule of $\vdash_n \alpha \to \beta, \Gamma \Rightarrow \Delta$.

Case $(i)$: $(R)$ is $(L\to)$ and the displayed $\alpha \to \beta$ is principal. Take premises of $(R)$.

Case (*ii*): $(R) \in \{(L\to), (R\to), (L\Box)\}$, and the displayed $\alpha\to\beta$ is side:

$$\frac{\vdash_{n-1} \alpha\to\beta, \Gamma_0 \Rightarrow \Delta_0 \quad \vdash_{n-1} \alpha\to\beta, \Gamma_1 \Rightarrow \Delta_1}{\vdash_n \alpha\to\beta, \Gamma \Rightarrow \Delta}(R).$$

By IH, we have $\vdash_{n-1} \Gamma_0 \Rightarrow \Delta_0\alpha$, $\vdash_{n-1} \beta\Gamma_0 \Rightarrow \Delta_0$, $\vdash_{n-1} \Gamma_1 \Rightarrow \Delta_1\alpha$, and $\vdash_{n-1} \beta\Gamma_1 \Rightarrow \Delta_1$. Use the same rule to get:

$$\frac{\vdash_{n-1} \Gamma_0 \Rightarrow \Delta_0, \alpha \quad \vdash_{n-1} \Gamma_1 \Rightarrow \Delta_1, \alpha}{\vdash_n \Gamma \Rightarrow \Delta, \alpha}(R), \quad \frac{\vdash_{n-1} \beta, \Gamma_0 \Rightarrow \Delta_0 \quad \vdash_{n-1} \beta, \Gamma_1 \Rightarrow \Delta_1}{\vdash_n \beta, \Gamma \Rightarrow \Delta}(R).$$

Case (*iii*): $(R) \in \{(R\Box), (K\Box), (4\Box)\}$. By Table 2.2, principal and side formulas of those rules are all $\Box$-ed, hence the displayed $\alpha\to\beta$ is a weakening formula. We can apply the same rule, but with $\alpha$ or $\beta$ as the weakening formula instead.

(B) Similarly.

(C) Follows directly from Theorem 28. $\dashv$

**Theorem 30 (DPA of Contraction)** *For any $\alpha$, $\beta$, $\Gamma$, and $\Delta$:*

*(A) If* G3[st4] $\vdash_n \beta, \beta, \Gamma \Rightarrow \Delta$ *then* G3[st4] $\vdash_n \beta, \Gamma \Rightarrow \Delta$.

*(B) If* G3[st4] $\vdash_n \Gamma \Rightarrow \Delta, \alpha, \alpha$ *then* G3[st4] $\vdash_n \Gamma \Rightarrow \Delta, \alpha$.

**Proof.** Follow the method used in [63].

(A) When $n = 0$: $\beta, \beta, \Gamma \Rightarrow \Delta$ is an instance of $(Ax)$ or $(L\bot)$. Trivial.

Induction step: Consider $(R)$, the last rule of $\vdash_n \beta, \beta, \Gamma \Rightarrow \Delta$.

Case (1): $(R)$ is $(L\to)$. Subcase (1.1): Both displayed (in the conclusion of $(R)$) $\beta$'s are side. Use IH. Subcase (1.2): Only one[4] displayed $\beta$'s is principal. Use Theorem 29 and IH.

Case (2): $(R)$ is $(R\to)$. Similar to Case (1).

[4]Note that in $(L\to), (R\to), (L\Box)$, there can be only one principal formula in the conclusion.

Case (3): $(R)$ is $(L\Box)$. Subcase (3.1): Both displayed $\beta$'s are side. Use IH. Subcase (3.2): only one of displayed $\beta$'s is principal. Then $\beta \equiv \Box\beta_0$, and

$$\frac{\vdash_{n-1} \beta_0, \Box\beta_0, \beta, \Gamma \Rightarrow \Delta}{\vdash_n \Box\beta_0, \beta, \Gamma \Rightarrow \Delta}(L\Box).$$

Applying IH, we have $\vdash_{n-1} \beta_0, \Box\beta_0, \Gamma \Rightarrow \Delta$, and then $\vdash_n \beta, \Gamma \Rightarrow \Delta$ follows from $(L\Box)$.

Case (4): $(R)$ is $(4\Box)$. Subcase (4.1): (at least) one of displayed $\beta$'s is weakening. Apply $(4\Box)$ on the same premise, but without introducing that occurrence of $\beta$. Subcase (4.2): both displayed $\beta$'s are principal. Then $\beta \equiv \Box\beta_0$, and

$$\frac{\vdash_{n-1} \beta_0, \beta_0, \Box\beta_0, \Box\beta_0, \Theta, \Box\Theta \Rightarrow \eta}{\vdash_n \Box\beta_0, \Box\beta_0, \Box\Theta, \Gamma \Rightarrow \Delta, \Box\eta}(4\Box).$$

By IH, $\vdash_{n-1} \beta_0, \Box\beta_0, \Theta, \Box\Theta \Rightarrow \eta$, and then $\vdash_n \Box\beta_0, \Box\Theta, \Gamma \Rightarrow \Delta, \Box\eta$ follows from $(4\Box)$.

Case (5): $(R) \in \{(K\Box), (R\Box)\}$. Similar to Case (4).

(B) Similar to the proof for (A). $\dashv$

Let $(Cut)$ rule be:

$$\frac{\vdash_a \Gamma \Rightarrow \Delta, \phi \qquad \vdash_b \phi, \Gamma \Rightarrow \Delta}{\vdash_{max\{a,b\}+1} \Gamma \Rightarrow \Delta} \quad (Cut).$$

The *rank* of this $(Cut)$ is the degree of *cut-formula* $\phi$, notation $dg(\phi)$, and the *level* of this $(Cut)$ is $a + b$, the sum of depths of premises. A $(Cut)$ with rank $dg(\phi) = r$ and level $l$ (resp. $< l$) is denoted by $(Cut^r_l)$, or $(Cut^\phi_l)$ (resp. $(Cut^r_{<l})$, or $(Cut^\phi_{<l})$). We may use only superscript or subscript, if we only care about the rank or level of a $(Cut)$.

**Remark:** There is another version of cut rule in [63] (where our $(Cut)$ above is called context-sharing cut $(Cut_{cs})$):

$$\frac{\Gamma_0 \Rightarrow \Delta_0, \phi \qquad \phi, \Gamma_1 \Rightarrow \Delta_1}{\Gamma_0, \Gamma_1 \Rightarrow \Delta_0, \Delta_1}(Cut').$$

With the help of Theorem 28 and 30, it is easy to show that $\mathsf{X} \cup \{(Cut)\}$ and $\mathsf{X} \cup \{(Cut')\}$ derive exactly the same set of sequents. It is thus reasonable to work with $(Cut)$ only.

**Theorem 31** **((*Cut*)-elimination)** *For any* $\Gamma$ *and* $\Delta$*:*

$\mathsf{G3[st4]} \vdash \Gamma \Rightarrow \Delta$ *iff* $\mathsf{G3[st4]} \cup \{(Cut)\} \vdash \Gamma \Rightarrow \Delta$.

**Proof.** Follow the method in [63], recursively eliminate the minimal level maximal rank $(Cut)$:

$$\frac{\vdash_a \Gamma \Rightarrow \Delta, \phi \qquad \vdash_b \phi, \Gamma \Rightarrow \Delta}{\Gamma \Rightarrow \Delta}(Cut^{\phi}_{a+b}).$$

(1): When $dg(\phi) = 0$: $\phi$ is an atom $p$, or $\bot$.

(1.1): One of cut-formulas is weakening. Trivial.

(1.2): One of cut-formulas is side. W.l.o.g., consider:

$$\frac{\dfrac{\vdash_{<a} \Gamma_0 \Rightarrow \Delta_0, \phi \quad \vdash_{<a} \Gamma_1 \Rightarrow \Delta_1, \phi}{\vdash_a \Gamma \Rightarrow \Delta, \phi}(R) \qquad \dfrac{\cdots}{\vdash_b \phi, \Gamma \Rightarrow \Delta}}{\Gamma \Rightarrow \Delta}(Cut^{\phi}_{a+b}).$$

By Table 2.2, $(R) \in \{(L\to), (R\to), (L\Box)\}$, since only these rules may have side formulas of degree 0. Replace the subproof with that $(Cut)$ as its last rule by:

$$\dfrac{\dfrac{\dfrac{\vdash_{<a} \Gamma_0 \Rightarrow \Delta_0 \phi}{\vdash_{<a} \Gamma_0 \Gamma \Rightarrow \Delta_0 \Delta \phi} \quad \dfrac{\vdash_b \phi \Gamma \Rightarrow \Delta}{\vdash_b \phi \Gamma_0 \Gamma \Rightarrow \Delta_0 \Delta}}{\Gamma_0 \Gamma \Rightarrow \Delta_0 \Delta}(Cut^{\phi}_{<a+b}) \quad \dfrac{\dfrac{\vdash_{<a} \Gamma_1 \Rightarrow \Delta_1 \phi}{\vdash_{<a} \Gamma_1 \Gamma \Rightarrow \Delta_1 \Delta \phi} \quad \dfrac{\vdash_b \phi \Gamma \Rightarrow \Delta}{\vdash_b \phi \Gamma_1 \Gamma \Rightarrow \Delta_1 \Delta}}{\Gamma_1 \Gamma \Rightarrow \Delta_1 \Delta}(Cut^{\phi}_{<a+b})}{\dfrac{\Gamma\Gamma \Rightarrow \Delta\Delta}{\Gamma \Rightarrow \Delta}}(R),$$

where by double line, we mean an application of weakening or contraction. By Theorem 28 and 30, these structural rules are depth-preserved eliminable.

(1.3): Both cut-formulas are principal.

(1.3.1): $\phi$ is $\bot$. $(L\bot)$ has principal formula only in antecedent. This subcase is impossible.

(1.3.2): $\phi$ is an atom $p$. Both premises are $(Ax)$'s:

$$\frac{\dfrac{}{\vdash_0 \Gamma \Rightarrow \Delta, p} \qquad \dfrac{}{\vdash_0 p, \Gamma \Rightarrow \Delta}}{\Gamma \Rightarrow \Delta}(Cut^{p}_{0}).$$

By the form of $(Ax)$, $p \in \Gamma \cap \Delta$, so $\Gamma \Rightarrow \Delta$ is also an instance of $(Ax)$.

(2): For induction step. Suppose that $(Cut_l^r)$ is eliminable for any $r < r_0$, or $r = r_0$ but $l < l_0$. Now consider $(Cut_{l_0}^{r_0})$.

(2.1): One cut-formula is weakening. Similar to (1.1), where $dg(\phi)$ is not essential. This is also the base case of sub-induction on level, since any cut-formula of degree $> 0$ can only be weakening in premises of a $(Cut_0)$.

(2.2): One cut-formula is side in $(R) \in \{(L\rightarrow), (R\rightarrow), (L\Box)\}$. Similar to (1.2), where the condition $dg(\phi) = 0$, which holds there but not here, is only used to get the condition that $(R) \in \{(L\rightarrow), (R\rightarrow), (L\Box)\}$, a given condition here.

(2.3): Both cut-formulas are principal. Note that $dg(\phi) > 0$.

(2.3.1): $\phi$ is $\alpha\rightarrow\beta$. This case is proved in the standard way (cf. [63]).

(2.3.2): $\phi$ is $\Box\psi$.

(2.3.2.1): In G3t, only $(K\Box)$ has principal formula in succedent, and both $(L\Box)$ and $(K\Box)$ have principal formulas in antecedent. In $(K\Box)$-$(L\Box)$ subcase:

$$\dfrac{\dfrac{\vdash_{a-1}\ \Theta \Rightarrow \psi}{\vdash_a\ \Box\Theta, \Gamma \Rightarrow \Delta, \Box\psi}(K\Box) \quad \dfrac{\vdash_{b-1}\ \psi, \Box\psi, \Box\Theta, \Gamma \Rightarrow \Delta}{\vdash_b\ \Box\psi, \Box\Theta, \Gamma \Rightarrow \Delta}(L\Box)}{\Box\Theta, \Gamma \Rightarrow \Delta}(Cut_{a+b}^{\Box\psi}),$$

can be replaced by:

$$\dfrac{\dfrac{\dfrac{\vdash_{a-1}\ \Theta\Rightarrow\psi}{\overline{\Theta\Box\Theta\Gamma\Rightarrow\Delta\psi}}}{\Box\Theta\Gamma\Rightarrow\Delta\psi}(\overrightarrow{L\Box}) \quad \dfrac{\dfrac{\vdash_{a-1}\ \Theta \Rightarrow \psi}{\vdash_a\ \psi\Box\Theta\Gamma\Rightarrow\Delta\Box\psi}(K\Box) \quad \dfrac{\cdots}{\vdash_{b-1}\ \psi\Box\psi\Box\Theta\Gamma\Rightarrow\Delta}}{\psi\Box\Theta\Gamma \Rightarrow \Delta}(Cut_{<a+b}^{\Box\psi})}{\Box\Theta\Gamma \Rightarrow \Delta}(Cut^{\psi}).$$

In $(K\Box)$-$(K\Box)$ subcase, we have:

$$\dfrac{\dfrac{\vdash_{a-1}\ \Theta_0 \Rightarrow \psi}{\vdash_a\ \Box\Theta_0, \Gamma_0 \Rightarrow \Delta, \Box\eta, \Box\psi}(K\Box) \quad \dfrac{\vdash_{b-1}\ \psi, \Theta_1 \Rightarrow \eta}{\vdash_b\ \Box\psi, \Box\Theta_1, \Gamma_1 \Rightarrow \Delta, \Box\eta}(K\Box)}{\Gamma \Rightarrow \Delta, \Box\eta}(Cut_{a+b}^{\Box\psi}),$$

where $\Gamma_0 \cup \Box\Theta_0 = \Gamma_1 \cup \Box\Theta_1 = \Gamma$. We can replace it by:

$$\dfrac{\dfrac{\dfrac{\vdash_{a-1}\ \Theta_0 \Rightarrow \psi}{\overline{\Theta_0, \Theta_1 \Rightarrow \psi, \eta}} \quad \dfrac{\vdash_{b-1}\ \psi, \Theta_1 \Rightarrow \eta}{\overline{\psi, \Theta_0, \Theta_1 \Rightarrow \eta}}}{\Theta_0, \Theta_1 \Rightarrow \eta}(Cut^{\psi}).}{\Gamma \Rightarrow \Delta, \Box\eta}(K\Box)$$

(2.3.2.2): In G3s and G34. Similar to (2.3.2.1).

(2.4): One of cut-formulas is side in ($R\Box$), while the other is principal (in ($R\Box$)).[5] Thus we have:

$$\dfrac{\dfrac{\vdash_{a-1} \Box\Theta_0 \Rightarrow \psi}{\vdash_a \Box\Theta_0, \Gamma_0 \Rightarrow \Delta, \Box\psi, \Box\eta}(R\Box) \quad \dfrac{\vdash_{b-1} \Box\psi, \Box\Theta_1 \Rightarrow \eta}{\vdash_b \Box\psi, \Box\Theta_1, \Gamma_1 \Rightarrow \Delta, \Box\eta}(R\Box)}{\Gamma \Rightarrow \Delta, \Box\eta}(Cut^{\Box\psi}_{a+b}),$$

where $\Box\Theta_0 \cup \Gamma_0 = \Box\Theta_1 \cup \Gamma_1 = \Gamma$. Replace this by

$$\dfrac{\dfrac{\dfrac{\dfrac{\vdash_{a-1} \Box\Theta_0 \Rightarrow \psi}{\vdash_a \Box\Theta_0, \Box\Theta_1 \Rightarrow \Box\psi, \eta}(R\Box) \quad \dfrac{\vdash_{b-1} \Box\psi, \Box\Theta_1 \Rightarrow \eta}{\vdash_{b-1} \Box\psi, \Box\Theta_0, \Box\Theta_1 \Rightarrow \eta}}{\Box\Theta_0, \Box\Theta_1 \Rightarrow \eta}(Cut^{\Box\psi}_{<a+b})}{\Box\Theta_0, \Box\Theta_1, \Gamma \Rightarrow \Delta, \Box\eta}(R\Box)}{\Gamma \Rightarrow \Delta, \Box\eta}.$$

⊣

**Corollary 32** G3[st4] *has subformula property.*

We benefit from ($Cut$)-elimination in two points. The first one is, our G3-style calculi are ($Cut$)-free formulations of corresponding modal logics, as presented in the following theorem.

**Theorem 33** G3s, G3t, G34 *are formulations of* S4, T, K4 *respectively.*

**Proof.** These calculi are equivalent to well-known Hilbert-systems. Associate to sequent $\Gamma \Rightarrow \Delta$ the formula $\bigwedge\Gamma \to \bigvee\Delta$. For one direction, the set of theorems in the Hilbert-system are closed under (axioms and) rules of our G3-style calculus. For the other direction, first note that all Hilbert-style axioms have their sequent forms derivable in corresponding G3-style calculus. Closure under $MP$ follows from ($Cut$)-elimination, and closure under Necessitation holds since from $\Rightarrow \eta$, sequent $\Rightarrow \Box\eta$ is derivable by any of ($R\Box$), ($K\Box$), and ($4\Box$). ⊣

[5]Precisely speaking, the remaining case is that one cut-formula is side in ($R\Box$), and the other is side in ($R\Box$) or principal. By Table 2.2, side formulas in ($R\Box$) are all negative, which implies that the positive cut-formula is principal. Since the negative cut-formula is from ($R\Box$), which determines the calculus (G3s), the positive cut-formula is principal in ($R\Box$).

The second point that we benefit from (*Cut*)-elimination is, polarities of formulas are preserved, which enables realization procedures in [2] and [14] to be applied.

### 2.2.2 Prehistoric Graph

Many notations will be introduced in this subsection as we go, and will be used in later sections like Section 3.1 and Section 4.5. A Gentzen-style sequent proof (tree) is denoted by $\mathcal{T}$, with each of its node (a sequent) denoted by $s_1, s_2, \cdots$, and its conclusion-premise relation denoted by $R$. Thus the root of a proof tree is the conclusion sequent of that proof. For any proof $\mathcal{T}$ and any sequent $s$ in $\mathcal{T}$, by $\mathcal{T} \upharpoonright s$, we mean the subproof of $s$ in $\mathcal{T}$.

In all G3-style calculi presented in Subsection 2.2.1, all rules preserve polarities of formula occurrences. Specifically, polarities are preserved for $\Box$ occurrences. Based on this property, families of $\Box$'s are defined in [2] and [13] for sequent calculi there.

We employ the same idea. A *family* is a set of related $\Box$ occurrences. Corresponding $\Box$ occurrences in premise and conclusion of a rule belong to a same family. Extending this to an equivalent relation, we can see that $\Box$ occurrences in a proof belong to pairwise disjoint families. We denote families by $1, 2, \cdots$, and an occurrence of $\Box$ in family $i$ by $\Box_i$. All $\Box$ occurrences in a family have a same polarity, which is defined as the *polarity* of the family. Hence, it is notationally safe to denote a positive (or negative) $\Box$ in a proof by $\boxplus$ (or $\boxminus$). A positive family is *principal*, if there is a positive $\Box$ in that family introduced principally in a modal rule (i.e., the displayed $\Box$ in formula $\Box\eta$ in $(R\Box), (K\Box), (4\Box)$ in Definition 27). In what follows, we say family $i$ *occurs* in a formula $\phi$, if there is a $\Box_i$ occurs in $\phi$, or $\phi(\Box_i)$ in notation. we say $i$ occurs in a set $\Gamma$ of formulas, if there is some formula $\phi(\Box_i)$ in $\Gamma$.

In [68], prehistoric relation is defined in G3s, on principal positive families. Based on that idea, we extend that notion, make it more generally.

**Definition 34 (Prehistoric Graph)** In a G3s, G3t, or G34 proof $\mathcal{T}$, the *prehistoric graph* $\mathcal{P}(\mathcal{T})$ is defined as the directed graph

$$\mathcal{P}(\mathcal{T}) := (F, \prec_L, \prec_R, \prec),$$

where $F$ is the set of principal positive families in $\mathcal{T}$, all of $\prec_L, \prec_R, \prec$ are binary relations on $F$, $\prec := \prec_L \cup \prec_R$, and:

1) in G3s:

$$\prec_L := \{(i,j) | \frac{\boxminus\Theta(\boxplus_i) \Rightarrow \eta}{\boxminus\Theta(\boxplus_i), \Gamma \Rightarrow \Delta, \boxplus_j\eta}(R\Box) \text{ is in } \mathcal{T}\},$$

$$\prec_R := \{(i,j) | \frac{\boxminus\Theta \Rightarrow \eta(\boxplus_i)}{\boxminus\Theta, \Gamma \Rightarrow \Delta, \boxplus_j\eta(\boxplus_i)}(R\Box) \text{ is in } \mathcal{T}\};$$

2) in G3t:

$$\prec_L := \{(i,j) | \frac{\Theta(\boxplus_i) \Rightarrow \eta}{\boxminus\Theta(\boxplus_i), \Gamma \Rightarrow \Delta, \boxplus_j\eta}(K\Box) \text{ is in } \mathcal{T}\},$$

$$\prec_R := \{(i,j) | \frac{\Theta \Rightarrow \eta(\boxplus_i)}{\boxminus\Theta, \Gamma \Rightarrow \Delta, \boxplus_j\eta(\boxplus_i)}(K\Box) \text{ is in } \mathcal{T}\};$$

3) in G34:

$$\prec_L := \{(i,j) | \frac{\Theta(\boxplus_i), \boxminus\Theta(\boxplus_i) \Rightarrow \eta}{\boxminus\Theta(\boxplus_i), \Gamma \Rightarrow \Delta, \boxplus_j\eta}(4\Box) \text{ is in } \mathcal{T}\},$$

$$\prec_R := \{(i,j) | \frac{\Theta, \boxminus\Theta \Rightarrow \eta(\boxplus_i)}{\boxminus\Theta, \Gamma \Rightarrow \Delta, \boxplus_j\eta(\boxplus_i)}(4\Box) \text{ is in } \mathcal{T}\}.$$

$\dashv$

Some results about prehistoric graphs were proved for G3s in [68]. We generalize those proofs to make them work for G3t and G34.

**Lemma 35** *In a proof $\mathcal{T}$, each family has a unique occurrence in the root.*

**Proof.** In any rule in those calculi, each occurrence in a premise has exactly one corresponding occurrence in the conclusion. $\dashv$

**Lemma 36** *In $\mathcal{P}(\mathcal{T})$ of a proof $\mathcal{T}$, for any $i \in F$, $(i, i) \notin \prec_R$.*

**Proof.** Assume $(i, i) \in \prec_R$. No matter which calculus $\mathcal{T}$ is from, there is a rule application in $\mathcal{T}$ that puts $(i, i)$ into $\prec_R$. Moreover, in the conclusion of that rule, there is an occurrence of $\boxplus_i \eta(\boxplus_i)$. This occurrence has exactly one corresponding occurrence in the conclusion of each step, and eventually in the root. Now there are two occurrences of $\boxplus_i$ in the root, which contradicts with Lemma 35. ⊣

**Lemma 37** *In $\mathcal{P}(\mathcal{T})$ of a proof $\mathcal{T}$, for any $i, j, k \in F$:*

*(A) if $k \prec_R j \prec_L i$ then $k \prec_L i$; and*

*(B) if $k \prec_R j \prec_R i$ then $k \prec_R i$.*

**Proof.** No matter which calculus $\mathcal{T}$ is from, $k \prec_R j$ implies that $\boxplus_j \psi(\boxplus_k)$ occurs in a sequent in $\mathcal{T}$ for some formula $\psi$. This occurrence has a unique related occurrence in the root. Consider any occurrence of $\boxplus_j$ in a sequent $s$ in $\mathcal{T}$. By Lemma 35, This $\boxplus_j$ is related to the occurrence of $\boxplus_j$ in $\boxplus_j \psi(\boxplus_k)$ in the root. Thus, this $\boxplus_j$ in $s$ is also in an occurrence of formula $\boxplus_j \psi(\boxplus_k)$.

(A) In G34, $j \prec_L i$ implies that

$$\frac{\Theta(\boxplus_j), \boxminus\Theta(\boxplus_j) \Rightarrow \eta}{\boxminus\Theta(\boxplus_j), \Gamma \Rightarrow \Delta, \boxplus_i \eta}(4\Box) \text{ is in } \mathcal{T}.$$

By the observation above, each $\boxplus_j$ in premise is in an occurrence of formula $\boxplus_j \psi(\boxplus_k)$. Hence

$$\frac{\Theta(\boxplus_j \psi(\boxplus_k)), \boxminus\Theta(\boxplus_j \psi(\boxplus_k)) \Rightarrow \eta}{\boxminus\Theta(\boxplus_j \psi(\boxplus_k)), \Gamma \Rightarrow \Delta, \boxplus_i \eta}(4\Box)$$

is in $\mathcal{T}$. So we have $k \prec_L i$. In G3s and G3t, similar proofs work.

(B) Proved in a similar way. ⊣

**Theorem 38** *In $\mathcal{P}(\mathcal{T})$ of a proof $\mathcal{T}$:*

*(A)* $(F, \prec_R)$ *has no loop; and*

*(B)* $(F, \prec)$ *has a loop iff* $(F, \prec_L)$ *has a loop.*

**Proof.** (A) Assume that $(F, \prec_R)$ has a loop $i_1 \prec_R i_2 \prec_R \cdots \prec_R i_n \prec_R i_1$. Using Lemma 37(B) $n-1$ times, we have $i_1 \prec_R i_1$, which contradicts Lemma 36.

(B) The ($\Leftarrow$) direction is trivial. For the ($\Rightarrow$) direction. Suppose that $(F, \prec)$ has a loop. By (A), there is a $\prec_L$ step in this loop. If there are no $\prec_R$ steps in that loop, we already have a loop in $(F, \prec_L)$. If $\prec_R$ steps also exist, then the loop has a form of

$$\cdots \prec i_x \prec_R i_y \prec_L i_z \prec \cdots .$$

By Lemma 37(A), we have

$$\cdots \prec i_x \prec_L i_z \prec \cdots$$

as a loop in $(F, \prec)$ with fewer $\prec_R$ steps. Doing this recursively until no $\prec_R$ remains, we get a loop in $(F, \prec_L)$. $\dashv$

By saying $\mathcal{P}(\mathcal{T}) = (F, \prec_L, \prec_R, \prec)$ *has a loop*, we mean $(F, \prec)$ has a loop, or equivalently, $(F, \prec_L)$ has a loop. A proof $\mathcal{T}$ is *loop-free*, if $\mathcal{P}(\mathcal{T})$ has no loop. A modal theorem is *loop-free provable*, if it has a loop-free proof in the corresponding calculus.

**Theorem 39 (Decidability of Loop-free Provability)** *There is a mechanical procedure to determine whether a modal formula is loop-free provable.*

**Proof.** As a sequent calculus with sub-formula property, $\mathsf{G3[st4]}$ is decidable, and it actually offers a decision procedure for the corresponding modal logic (cf. [22]). To proof $\phi$ in $\mathsf{G3[st4]}$, one can take $\Rightarrow \phi$ as the main goal, and get subgoals via rules (in the reverse direction from conclusion to premises), or close branches via axioms. Our procedure to determine loop-free provability is based on this idea.

In our setting, for all rules expect $(R\Box)$, $(K\Box)$, and $(4\Box)$, premise(s) are uniquely determined after specifying the principal formula from the conclusion. For an $(R\Box)$, $(K\Box)$, or $(4\Box)$ rule, after specifying the principal formula, we have options on carrying formulas from the antecedent to the premise - formulas prefixed by $\boxminus$ can either be seen as side or be seen as weakening. If such a formula is seen as side, it will occur in the premise and principal $\boxplus$'s in it will contribute to the $\prec_L$ in prehistoric graph. To make sure that we are constructing a loop-free proof, we note down each of those edges involved, and a later trial via $(R\Box)$, $(K\Box)$, or $(4\Box)$ is allowed, only if new edges it brings in do not form a loop with existing edges. $\dashv$

### 2.2.3 Realization of Loop-free Provable Theorems

In this subsection, we consider realizations of loop-free provable S4, T, or K4 theorems, into corresponding justification logics LP, JT, J4, resp., with the property that self-referential constant specification is not employed. Our method is based on the realization procedure in [2]. Another version that has more similarities with ours came from [14]. We present some preliminaries from [2] and [14], with necessary extensions to fit our goal.

To deal with $(R\Box)$, *Internalization Lemma* is employed by realization procedures in [2] and [14], through in different versions. For $(K\Box)$, a corresponding treatment can be found in [46] with name *Internalization Property*. We present here a version that will work in our setting. We denote series of terms or formulas $x_1, \cdots, x_n$ by $X$, $\phi_1, \cdots, \phi_n$ by $\Phi$, and $x_1 : \phi_1, \cdots, x_n : \phi_n$ by $X : \Phi$, etc..

**Lemma 40 (Internalization Lemma [2][46])** *(A) In* LP*, from a derivation* $X : \Theta \vdash \eta$*, we can construct a derivation* $X : \Theta \vdash t(X) : \eta$ *for a term* $t$*.*

*(B) In* JT*, similarly from* $\Theta \vdash \eta$ *to* $X{:}\Theta \vdash t(X){:}\eta$ *for a term* $t$ *and variables* $X$*.*

*(C) In* J4*, similarly from* $\Theta, X{:}\Theta \vdash \eta$ *to* $X{:}\Theta \vdash t(X){:}\eta$ *for a term* $t$*.*

*In all three cases, we can construct in such a way that no positive variables are introduced, and if the original derivation has non-self-referential (non-direct-self-referential, resp.) constant specification, then so is the constructed derivation.*

**Proof.** (A) Inductively substitute formulas in original derivation by series of formulas as follows.

Substitute axiom $A$ by $c{:}A$, where $c$ is a new constant.

Substitute hypotheses $x{:}\theta$ by

$$\begin{array}{l} x{:}\theta \\ x{:}\theta \to !x{:}x{:}\theta \\ !x{:}x{:}\theta \end{array}$$

Substitute $AN$ $c{:}A$ by

$$\begin{array}{l} c{:}A \\ c{:}A \to !c{:}c{:}A \\ !c{:}c{:}A \end{array}$$

For an $MP$ on $\phi \to \psi$ and $\phi$ in the original derivation, by IH, we have got subderivations of $s_1 : (\phi \to \psi)$ and $s_2 : \phi$ for some terms $s_1, s_2$. Combine these two subderivations, and extend by

$$\begin{array}{l} s_1{:}(\phi \to \psi) \to (s_2{:}\phi \to s_1 \cdot s_2{:}\psi) \\ s_2{:}\phi \to s_1 \cdot s_2{:}\psi \\ s_1 \cdot s_2{:}\psi \end{array}$$

Note that no variables are introduced, and we can always take new constant while dealing with $AN$.

(B) Treatments for axiom and $MP$ are similar to (A).

For each hypotheses $\theta \in \Theta$, substitute $\theta$ by $x : \theta$, where $x$, being negative, is the only variable introduced.

Substitute $AN$[6] $!^{n-1}c{:}\cdots{:}c:A$ by $!^{n}c{:}\cdots:c:A$.

[6]Note our special notation for constants in JT.

(C) Treatments for axiom, $AN$, and $MP$ are similar to (A). Here we have two kinds of hypotheses. For each hypotheses $x{:}\Theta$ where $x \in X$ and $\theta \in \Theta$, substitute $x{:}\theta$ by

$$\begin{array}{l} x{:}\theta \\ x{:}\theta \rightarrow !x{:}x{:}\theta \\ !x{:}x{:}\theta. \end{array}$$

For each hypotheses $\theta$ where $\theta \in \Theta$, substitute $\theta$ by $x:\theta$, where $x$, being negative, is the variable corresponds to $\theta$ in $X{:}\Theta$. ⊣

A corollary of Lemma 40 will be recalled later in Section 3.2.

**Corollary 41** *If* S4 *(*T*,* K4*, resp.) theorem $\phi$ is non-self-referential (non-directly-self-referential), then so is $\Box\phi$.*

**Proof.** Suppose $\phi$ is non-self-referential, i.e., there is a realization $\psi$ of $\phi$ s.t. $\vdash \psi$ holds with a non-self-referential constant specification. Apply Lemma 40 on this derivation, we got a derivation of $t:\psi$ for a term $t$ with a non-self-referential constant specification. Observe that $t{:}\psi$ is a realization of $\Box\phi$.

This verification works in all three pairs of logics, and the non-direct-self-referential case is similar. ⊣

Now we are ready to present a realization procedure that can realize a modal theorem with a loop-free proof non-self-referentially. The idea was originated in [68], but for S4 only there. We now amend it to accommodate both T and K4.

**Theorem 42 (Non-self-referentiality of Loop-free Theorems)** *Assume the following:*

*(i)* X *ranges over* {T, K4, S4}*,*

*(ii)* Y *respectively ranges over* {G3t, G34, G3s}*, and*

*(iii)* Z *respectively ranges over* {JT, J4, LP}.

*Then for any* $\mathsf{X}$*-theorem* $\phi$*, if* $\Rightarrow\ \phi$ *has a loop-free proof in* $\mathsf{Y}$*, then there is a normal realization* $\phi^r$ *of* $\phi$ *in* $\mathsf{Z}$*, s.t.* $\mathsf{Z}(\mathcal{CS}) \vdash \phi^r$ *for some non-self-referential constant specification* $\mathcal{CS}$*.*

**Proof.** We prove this by presenting a realization procedure that works for all three modal-justification pairs. The main idea is to incorporate an order, inherited from loop-free proofs, in the realization procedure from [2].

We start with a $\mathsf{G3[st4]}$ loop-free proof of $\Rightarrow\ \phi$, denoted by $\mathcal{T}$. The "conclusion-to-premise" relation of $\mathcal{T}$ is denoted by $R$, and notes (sequents) in $\mathcal{T}$ are denoted by $s_0, s_1, \cdots$, and specifically, the root sequent by $s_r$. $R^+$ and $R^*$ are transitive closure and reflexive-transitive closure of $R$, resp.. Denote principal positive families by $1, \cdots, m$, and recall our notation of a $\boxplus$ of family $i$ as $\boxplus_i$.

Substitute $\boxminus$'s and non-principal $\boxplus$'s in $\mathcal{T}$ by variables $x_1, x_2, \cdots$. In what follows, there will be no further changes on those variables. So they will not contribute in forming self-referentiality in constant specification.

There could be many $(R\Box)$, $(K\Box)$, or $(4\Box)$ rules that introduce principal $\boxplus$'s in a family. Each rule introduces only one principal $\boxplus$, and hence each rule is associated to a unique family. In $\mathsf{G3s}$, we denote $(R\Box)$'s associated to family $i$ by $(R\Box)_{i.1}, \cdots, (R\Box)_{i.m_i}$, where $m_i$ is the number of those $(R\Box)$'s. The only premise and conclusion of $(R\Box)_{i.j}$ is denoted by $I_{i.j}$ and $O_{i.j}$, respectively. Similar notations for $\mathsf{G3t}$ and $\mathsf{G34}$. Associate a *provisional* variable $u_{i.j}$ to $(R\Box)_{i.j}$, $(K\Box)_{i.j}$, or $(4\Box)_{i.j}$. Substitute each $\boxplus_i$ by sum of provisional variables of family $i$, i.e., $u_{i.1} + u_{i.2} + \cdots + u_{i.m_i}$. Later, $u_{i.j}$ will be substituted by a term, denoted by $t_{i.j}$, that comes from further treatment on $(R\Box)_{i.j}$, $(K\Box)_{i.j}$, or $(4\Box)_{i.j}$.

A formula, set of formula, sequent, or tree is *provisional-variable-free*, if there is no provisional variable that occurs in it.

What we have done so far is the first step of realization, called *variable substitution*. Results of variable substitution on tree, sequent, rule $(R)$, set of formulas, and formula, are denoted by $\mathcal{T}'$, $s'$, $(R)'$, $\Gamma'$, and $\phi'$, resp..

The second step of realization, called *inductively substitution*, offers inductively a Hilber-style justification derivation of the modal theorem associated to each sequent in $\mathcal{T}$. It calls for Internalization Lemma and a substitution of provisional variables by provisional-variable-free terms to go through $(R\Box)'$, $(K\Box)'$, and $(4\Box)'$ rules.

In realization procedures in [2] and [14], inductively substitution is performed from leaf-side nodes of $\mathcal{T}'$ to root-side nodes, and constants are only introduced while employing Internalization Lemma. We follow this line, but with the $\epsilon$-function introduced. The idea is, in $\mathsf{G3s}$ for instance, $(R\Box)'_{i.j}$ is treated as the $\epsilon(i.j)$-th one among all $(R\Box)'$ instances. Taking care of the "leaf-side to root-side" requirement, we must guarantee that:

$$O_{i_1.j_1} R^+ O_{i_2.j_2} \text{ implies } \epsilon(i_2.j_2) < \epsilon(i_1.j_1).$$

This is always possible on a tree, but we can have more restrictions, given the fact that $\mathcal{T}$ is loop-free.

**Claim 43** *There is an $\epsilon$ satisfies all conditions above together with the condition that for any $j_1 \in \{1, \cdots, m_{h_1}\}$, $j_2 \in \{1, \cdots, m_{h_2}\}$:*

$$h_2 \prec h_1 \textit{ implies } \epsilon(h_2.j_2) < \epsilon(h_1.j_1).$$

**Proof.** We claim that there is a principal positive family $i_1$ s.t. $h \not\prec i_1$ for any principal positive family $h$. Otherwise, since there are only finitely many principal positive families in $\mathcal{T}$, there would be a prehistoric loop, which contradicts the assumption.

Similarly, we can show that there are families $i_2, i_3, \cdots$ s.t.: if $h \prec i_2$, then $h \in \{i_1\}$; if $h \prec i_3$, then $h \in \{i_1, i_2\}$; $\cdots$; if $h \prec i_z$, then $h \in \{i_1, \cdots, i_{z-1}\}$; $\cdots$.

Since there are only finitely many principal positive families in $\mathcal{T}$, we will enumerate all of them as $i_1, i_2, \cdots, i_m$ in the way above. Thus, the desired $\epsilon$ is obtained by setting $\epsilon(i_z.j) = j + \sum_{w=1}^{z-1} m_{i_w}$ for each family $i_z$ and $j \in \{1, \cdots, m_{i_z}\}$. ⊣

Note that by the $\epsilon$ from the proof above, once we treat an instance of family $i$, we will not treat any instance from other families, until all instances from $i$ are treated. Thus an order of families is also respected by $\epsilon$, and we can take $\epsilon(i)$ to denote the position of $i$ in that order.

**Claim 44** *There is a way of assigning names to families and an $\epsilon$ s.t.*

$$h_2 \prec h_1 \textit{ implies } h_2 < h_1.$$

**Proof.** Take $\epsilon$ as in Claim 43, and assume we happen to assigned names to families in a way that $\epsilon(i) = i$.[7] Then Claim 43 implies what desired. ⊣

We have a series of new constants $c_{i.j.1}, \cdots, c_{i.j.m_{i.j}}$ associated with, say $(R\Box)'_{i.j}$, where $m_{i.j}$ is the number of those constants. We always have enough new constants in each series, so $m_{i.j}$ is determined by how much constants we need in treating $(R\Box)'_{i.j}$. In our procedure, any constant $c_{i.j.k}$ is introduced at most once. Therefore, the axiom being prefixed by $c_{i.j.k}$ is uniquely determined. We denote this axiom by $A_{i.j.k}$.

An application of $AN.1$ (or $AN.2$),

$$\frac{}{c_{i.j.k} : A_{i.j.k}} \text{ (or } \frac{}{!^n c_{i.j.k} : !^{n-1} c_{i.j.k} : \cdots : c_{i.j.k} : A_{i.j.k}}\text{)},$$

is *$\epsilon$-allowed*, if $A_{i.j.k}$ is provisional-variable-free, and does not contain any constant $c_{i'.j'.k'}$ with $\epsilon(i') \geq \epsilon(i)$ (or $i' \geq i$ in our lucky way of assigning names to families). A constant

[7] This is just a matter of notation. There are no requirements on how we assign names to families (i.e., which family has number $i$ as its name). Observations above show that $\epsilon$ can be taken to respect the order of families, thus we may assume we happened to assign it in the way that cooperates with $\epsilon$.

specification is *$\epsilon$-allowed*, if all formulas in it, as applications of *AN*.1 or *AN*.2, are $\epsilon$-allowed.

We perform inductively substitution ordered by this $\epsilon$ on $\mathcal{T}'$, round by round. Let $\mathcal{T}^0$ be $\mathcal{T}'$ at the beginning. At round $i$, all instances of family $i$ are treated. IH gives the following items:

- A tree $\mathcal{T}^{i-1}$, which is obtained by substituting $t_{i-1.j_0}$ for provisional variable $u_{i-1.j_0}$ in $\mathcal{T}^{i-2}$ for each $j_0 \in \{1, \cdots, m_{i-1}\}$. (We have round-superscripted notations for subtree, sequent, rule, derivation, set of formulas (e.g., constant specification), formula, and term, whenever necessary.)

- A Hilbert-style justification derivation $d^{i-1}_{i_0.j_0}$ of the modal formula associated to $O^{i-1}_{i_0,j_0}$, with all provisional variables occur in it also occur in $O^{i-1}_{i_0,j_0}$, and with an $\epsilon$-allowed injective constant specification $\mathcal{CS}^{i-1}_{i_0.j_0}$ of $d^{i-1}_{i_0.j_0}$, for each $j_0 \in \{1, \cdots, m_{i_0}\}$, $i_0 \in \{1, \cdots, i-1\}$.

The goal of round $i$ is to:

- Generate a provisional-variable-free term $t_{i.j}$ for each $j \in \{1, \cdots, m_i\}$, and get $\mathcal{T}^i$ by substituting $t_{i.j}$ for $u_{i.j}$ in $\mathcal{T}^{i-1}$ for each $j \in \{1, \cdots, m_i\}$.

- Generate a Hilbert-style justification derivation $d^i_{i.j}$ of the modal formula associated to $O^i_{i.j}$, with all provisional variables occur in it also occur in $O^i_{i.j}$, and with an $\epsilon$-allowed injective constant specification $\mathcal{CS}^i_{i.j}$ of $d^i_{i.j}$, for each $j \in \{1, \cdots, m_i\}$.

Followed is a detailed instruction of inductively substitution.

From $\mathcal{T}^{i-1}$, take $(\mathcal{T} \upharpoonright O_{i.j})^{i-1}$ for each $j \in \{1, \cdots, m_i\}$, to get a bunch of subtrees. Note that those subtrees are pair-wisely disjoint, since otherwise, we would have $(\mathcal{T} \upharpoonright O_{i.j'})^{i-1}$

as a proper subtree of $(\mathcal{T} \restriction O_{i.j})^{i-1}$, which would then implies that $O_{i.j}R^{+}O_{i.j'}$ in $\mathcal{T}$, and then $\boxplus_i$ occurs in $I_{i.j}$, and hence $i \prec i$, contradicts with the loop-free assumption.

The following claim holds in our specified way of assigning names to families and choice of $\epsilon$.

**Claim 45** $(\mathcal{T} \restriction I_{i.j})^{i-1}$ *is provisional-variable-free.*

**Proof.** For any provisional variable $u_{i_1.j_1}$ that occurs in $(\mathcal{T} \restriction I_{i.j})^{0}$, we have $i_1 \prec i$, and hence by Claim 44, $i_1 < i$, which implies that $u_{i_1.j_1}$ has been substituted by the provisional-variable-free $t_{i_1.j_1}$ and does not occur in $(\mathcal{T} \restriction I_{i.j})^{i-1}$. ⊣

Then, from each $(\mathcal{T} \restriction O_{i.j})^{i-1}$, eliminate all subtrees $(\mathcal{T} \restriction I_{i_0.j_0})^{i-1}$ for $i_0 \in \{1, \cdots, i-1\}$ and $j_0 \in \{1, \cdots, m_{i_0}\}$. What remains are a bunch of pair-wisely disjoint tree-like structures of sequents, where leaf nodes are axioms or $O$-sequents of earlier families, root nodes are $O^{i-1}_{i.j}$ for some $j$, and despite of the last rule from $I^{i-1}_{i.j}$ to $O^{i-1}_{i.j}$, conclusion-premise relations are rules from $(Ax)^{i-1}$, $(L\bot)^{i-1}$, $(L\rightarrow)^{i-1}$, $(R\rightarrow)^{i-1}$, and $(L\Box)^{i-1}$.

We make a sub-induction on each of those tree-like structures, from leaves to $I^{i-1}_{i.j}$ instead of $O^{i-1}_{i.j}$, generate a provisional-variable-free Hilbert-style derivation inductively on it. In this sub-induction, unless specifically mentioned, all objects are read as being superscripted by $(\cdot)^{i-1}$.

Sub-leaf case.

Sequent $s$ introduced by $(Ax)$ has the form of

$$\overline{p, \Gamma \Rightarrow \Delta, p}.$$

We can take

$$\begin{array}{l} p \wedge (\bigwedge \Gamma) \\ p \wedge (\bigwedge \Gamma) \rightarrow (\bigvee \Delta) \vee p \\ (\bigvee \Delta) \vee p \end{array}$$

as the Helbert-style derivation desired. Note that all formulas in this derivation have occurrences in $s$, which is provisional-variable-free by Claim 45. Thus this derivation is provisional-variable-free. The constant specification of this derivation is empty, and hence is $\epsilon$-allowed.

Similarly, if $s$ is introduced by $(L\bot)$.

If $s$ is $O^{i-1}_{i_0.j_0}$, since $O_{i.j}R^{+}O_{i_0.j_0}$, we have $i_0 \prec i$, and hence $i_0 < i$ by Claim 44. Then the main IH gives a derivation $d^{i-1}_{i_0.j_0}$ of $O^{i-1}_{i_0.j_0}$, with all provisional variables in it occur in $O^{i-1}_{i_0.j_0}$, and with an $\epsilon$-allowed $\mathcal{CS}^{i-1}_{i_0.j_0}$ of $d^{i-1}_{i_0.j_0}$. Since $O^{i-1}_{i_0.j_0}$ is a sequent of $(\mathcal{T} \upharpoonright I_{i.j})^{i-1}$, by Claim 45, $d^{i-1}_{i_0.j_0}$ is provisional-variable-free.

Sub-induction step.

For sequent $s$ introduced by $(L\rightarrow)$:

$$\frac{\Gamma \Rightarrow \Delta, \phi \qquad \psi, \Gamma \Rightarrow \Delta}{\phi \rightarrow \psi, \Gamma \Rightarrow \Delta}.$$

By sub-IH, we have provisional-variable-free derivations $d_L$ and $d_R$ of the two premises, each with an $\epsilon$-allowed constant specification. We apply the deduction theorem of Hilbert-style formulation (Lemma 4) to $d_R$, and denote the resulting derivation $(\bigwedge \Gamma \vdash \psi \rightarrow \bigvee \Delta)$ by $d_{R'}$. Specifically, if we employ the standard method to calculate, the resulting $d_{R'}$ is also provisional-variable-free, and the constant specification is not effected. Now what follows is a derivation of $s$.

$$
\begin{array}{l}
\bigwedge \Gamma \\
(d_L) \\
(\bigvee \Delta) \vee \phi \\
\phi \rightarrow \psi \\
((\bigvee \Delta) \vee \phi) \rightarrow (\phi \rightarrow \psi) \rightarrow ((\bigvee \Delta) \vee \psi) \\
(\phi \rightarrow \psi) \rightarrow ((\bigvee \Delta) \vee \psi) \\
(\bigvee \Delta) \vee \psi \\
\bigwedge \Gamma \\
(d_{R'}) \\
\psi \rightarrow \bigvee \Delta \\
((\bigvee \Delta) \vee \psi) \rightarrow (\psi \rightarrow \bigvee \Delta) \rightarrow \bigvee \Delta \\
(\psi \rightarrow \bigvee \Delta) \rightarrow \bigvee \Delta \\
\bigvee \Delta
\end{array}
$$

Despite of provisional-variable-free $d_L$ and $d_{R'}$, all other formulas are consisted of $\Gamma, \Delta, \phi, \psi$. All of them occur in the conclusion of $(L\rightarrow)$, which is a sequent in $(\mathcal{T} \upharpoonright I_{i.j})^{i-1}$. By Claim 45, all of them are provisional-variable-free, hence so is the derivation above. The constant specification of the derivation is the union of that from both side, which implies that each $AN.1$ or $AN.2$ in the resulting constant specification is inherited from an $\epsilon$-allowed constant specification. Thus, the constant specification of the derivation above is also $\epsilon$-allowed.

For a sequent introduced by $(R \rightarrow)$, also employ the deduction theorem, and the case is similar but easier than the $(L\rightarrow)$ case.

In G3[st], after the variable substitution, $(L\Box)$ will have a form of

$$\frac{\phi, t{:}\phi, \Gamma \Rightarrow \Delta}{t{:}\phi, \Gamma \Rightarrow \Delta}.$$

For the case that $s$ is introduce by $(L\Box)$, by sub-IH, we have a provisional-variable-free derivation $d$ of the premise. The desired provisional-variable-free derivation of $s$ is then gained by adding

$$
\begin{array}{l}
t{:}\phi \\
t{:}\phi \rightarrow \phi
\end{array}
$$

to the beginning of $d$. The constant specification is not effected.

This is the end of the sub-induction.

By this sub-induction, we have got a provisional-variable-free derivation of $I^{i-1}_{i.j}$, with an $\epsilon$-allowed constant specification. The remaining step from $I^{i-1}_{i.j}$ to $O^{i-1}_{i.j}$ can be treated as follows.

(i) Apply the Internalization Lemma (Lemma 40) on the provisional-variable-free derivation of $I^{i-1}_{i.j}$, and get a derivation of

$$X : \Theta^{i-1} \Rightarrow t_{i.j} : \eta^{i-1}.$$

The term $t$ from Internalization Lemma is taken as our term $t_{i.j}$, which will be used later to replace $u_{i.j}$ throughout $\mathcal{T}^{i-1}$. Lemma 40 says, no new positive variables are introduced. Since provisional variables are all positive, no provisional variables are introduced, and hence the resulting derivation is provisional-variable-free. The constant specification of this new derivation is got by adding all $AN.1$ or $AN.2$ used in Internalization Lemma to the original $\epsilon$-allowed constant specification. To show that the new constant specification is also $\epsilon$-allowed, it is sufficient to show that all $AN.1$ or $AN.2$ used in Internalization Lemma are $\epsilon$-allowed, i.e., all axioms being prefixed are provisional-variable-free, and do not have constants from $\geq i$ families. The original derivation of $I^{i-1}_{i.j}$ is provisional-variable-free, so are all of its axioms. Those provisional-variable-free axioms are exactly those being prefixed during applying the Internalization Lemma. Constants from $\geq i$ families do not occur in those axioms, since these constants have not been introduced by our procedure when generating the original derivation of $I^{i-1}_{i.j}$. Hence, the new constant specification is also $\epsilon$-allowed.

(ii) Use axioms $A4.1$, $A4.2$ and rule $MP$ to get the sum of $t_{i.j}$ with provisional variables associated with other $(R\Box)$, $(K\Box)$, or $(4\Box)$ rules of family $i$. The result is a derivation of

$$X:\Theta^{i-1} \Rightarrow u_{i.1}+\cdots+u_{i.j-1}+t_{i.j}+u_{i.j+1}+\cdots+u_{i.m_i}:\eta^{i-1}.$$

In this derivation, all provisional variables are in the family $i$, and constant specification is not effected, and hence is $\epsilon$-allowed.

(iii) Use Boolean axioms and rule $MP$ to get weakening formulas of $(R\Box)^{i-1}_{i.j}$, $(K\Box)^{i-1}_{i.j}$, or $(4\Box)^{i-1}_{i.j}$. We may introduce some provisional variables here, but all of them occur in $O^{i-1}_{i.j}$. The constant specification is not effected, and hence is $\epsilon$-allowed.

(iv) Doing (i)-(iii) for each of tree-like structures of family $i$, we get the desired series of provisional-variable-free terms $t_{i.1},\cdots,t_{i.m_i}$. Substituting $u_{i.j_0}$ by $t_{i.j_0}$ for each $j_0\in\{1,\cdots,m_i\}$ in $\mathcal{T}^{i-1}$, we get $\mathcal{T}^i$. Substituting $u_{i.j_0}$ by $t_{i.j_0}$ for each $j_0\in\{1,\cdots,m_i\}$ in the derivation generated above while treating $O^{i-1}_{i.j}$, we get the desired derivation $d^i_{i.j}$ of $O^i_{i.j}$ for each $j\in\{1,\cdots,m_i\}$. Note that all provisional variables introduced in (ii) are in family $i$, and hence are eliminated by this substitution. So, all provisional variables remain are introduced in (iii), and hence occur in $O^{i-1}_{i.j}$. Within these provisional variables, those from family $i$ are eliminated by the substitution; the others, from later families, are not eliminated. This substitution also brings $O^{i-1}_{i.j}$ to $O^i_{i.j}$, thus all provisional variables occur in $d^i_{i.j}$ also occur in $O^i_{i.j}$. After (iii), the constant specification is $\epsilon$-allowed, and hence is provisional-variable-free. The substitution above will only introduce new terms to positions where provisional variables of family $i$ located. Therefore, constant specification is not effected in the substitution above, and is still $\epsilon$-allowed.

To finish our proof for round $i$, we need to emphasize that, even in a later round $i_2$ and more provisional variables are substituted by corresponding terms, $d^{i_2}_{i.j}$ is still a derivation of $O^{i_2}_{i.j}$, with all its provisional variables occur in $O^{i_2}_{i.j}$, and with an $\epsilon$-allowed constant specification. Since substitutions of provisional variables by provisional-variable-free terms are performed on both the proof tree (where $O$-sequent resides) and the derivation $d_{i.j}$, the requirement for derivation and its provisional variables are satisfied. Later substitutions of provisional variables by provisional-variable-free terms will not effect our constant specification, which is $\epsilon$-allowed, and hence provisional-variable-free.

Now we have finished the instruction for round $i$ in the inductively substitution, and also have proved some results. The base step where $i = 1$ of the inductively substitution is similar but easier, with no earlier families, and hence no $O$-sequents to the leaf-side of tree-like structures.

This is the *end of inductively substitution*.

After the inductively substitution, we have treated all $(R\Box)'$, $(K\Box)'$, $(4\Box)'$ rules and all rules to the leaf-side of them in $\mathcal{T}'$. We have also eliminated all provisional variables, and got a Hilbert-style derivation of each root-most $O$-sequent, with the constant specification of this derivation being $\epsilon$-allowed. Since we have eliminated all provisional variables, all those derivation are fully in $\mathsf{LP}$, $\mathsf{JT}$, or $\mathsf{J4}$. From those $O$-sequents to the root $s^m_r$ in $\mathcal{T}^m$, all rules are $(Ax)^m$, $(L\bot)^m$, $(L\to)^m$, $(R\to)^m$, and $(L\Box)^m$. We can go through those rules just like what we did in inductively substitution. The result will be a derivation of $s^m_r$ fully in $\mathsf{LP}$, $\mathsf{JT}$, or $\mathsf{J4}$. In out treatment for rules in inductively substitution, $(Ax)$, $(L\bot)$, $(R\to)$,

$(L\Box)$ have no effect on constant specification; the case for $(L\rightarrow)$ will union two constant specification's from premises, but will get an $\epsilon$-allowed constant specification, if both constant specification's from premises are $\epsilon$-allowed. Thus, the constant specification of the derivation of the root sequent is also $\epsilon$-allowed. We denote this constant specification by $\mathcal{CS}''$.

This is the *end of realization procedure*. Since we assign distinct variables to negative modalities, and $\mathcal{CS}''$ is injective, the resulting realization is normal.

In the very beginning, we are given a loop-free proof $\mathcal{T}$ of $\Rightarrow \phi$. That is, $s_r$ is $\Rightarrow \phi$, hence $s_r^m$ has the form of $\Rightarrow \psi$. Having eliminated all provisional variables, $\psi$ is an LP, JT, or J4 formula. From $s_r$, through $s_r'$, then $s_r^1, \cdots$, till $s_r^m$, only $\Box$'s in $\phi$ are substituted by variables, or sums of provisional variables, and then provisional variables are substituted by terms. Thus $\psi$ is a realization of $\phi$. The generated Hilbert-style derivation of $s_r^m$ is then a derivation $\vdash \psi$, with an $\epsilon$-allowed $\mathcal{CS}''$.

To conclude our proof, it is sufficient to show that $\mathcal{CS}''$ is not self-referential. Suppose that it were self-referential, then in S4 and J4 cases,

$$\left\{\begin{array}{l} c_{i_1.j_1.k_1} : A_{i_1.j_1.k_1}(c_{i_2.j_2.k_2}), \\ \qquad \cdots\cdots \\ c_{i_{z-1}.j_{z-1}.k_{z-1}} : A_{i_{z-1}.j_{z-1}.k_{z-1}}(c_{i_z.j_z.k_z}), \\ c_{i_z.j_z.k_z} : A_{i_z.j_z.k_z}(c_{i_1.j_1.k_1}) \end{array}\right\} \subseteq \mathcal{CS}'';$$

and in JT case,

$$\left\{\begin{array}{l} !^{n_1}c_{i_1.j_1.k_1} : \cdots : c_{i_1.j_1.k_1} : A_{i_1.j_1.k_1}(c_{i_2.j_2.k_2}), \\ \qquad \cdots\cdots \\ !^{n_{z-1}}c_{i_{z-1}.j_{z-1}.k_{z-1}} : \cdots : c_{i_{z-1}.j_{z-1}.k_{z-1}} : A_{i_{z-1}.j_{z-1}.k_{z-1}}(c_{i_z.j_z.k_z}), \\ !^{n_z}c_{i_z.j_z.k_z} : \cdots : c_{i_z.j_z.k_z} : A_{i_z.j_z.k_z}(c_{i_1.j_1.k_1}) \end{array}\right\} \subseteq \mathcal{CS}''.$$

In all cases, since $\mathcal{CS}''$ is $\epsilon$-allowed, we have $i_z < i_{z-1} \cdots < i_2 < i_1 < i_z$, which is impossible. Therefore, $\mathcal{CS}''$ is not self-referential. $\dashv$

The above theorem says, prehistoric-loop-free provable modal theorems are non-self-referential, and prehistoric loop is a decidable necessary condition of self-referentiality.

The procedure presented in our proof of Theorem 42 is essentially Artemov's procedure in [2] with some careful preservations of orders. The question that whether or not the other direction of Theorem 42 holds has been open for years. Note that if it does hold, then self-referentiality of model theorems would be decidable. This thesis, however, is not able to answer this question. Instead, we will utilize what we have so far to explore self-referentiality as much as possible, as we do in Chapter 3.

# Chapter 3

# Non-self-referential Fragments

In Section 2.2, we have shown that for a modal theorem, prehistoric-loop-free provability implies non-self-referentiality. The other direction, though has be desired for years, is still open. So far, prehistoric-loop-free provability is the best known approximation of non-self-referentiality, and similar ideas have been applied to other sequent calculi to prove non-self-referentiality of theorems, like in [56] and [44]. While authors of these works try to fit non-self-referentiality into the framework of substructural logics, we insist on the the problem itself, understanding the collection of non-self-referential theorems in a modal logic as a fragment of it, and explore properties of that fragment.

In Section 3.1, we study the fragment of a modal logic whose theorems are prehistoric-loop-free provable as in Section 2.2.

Results in Section 3.1 will be employed in Section 3.2 to deduce properties of the non-self-referential fragment. In these two sections, we will treat all of $\mathsf{T}$, $\mathsf{K4}$, and $\mathsf{S4}$ uniformly.

Section 3.3 compares same kinds of fragments of $\mathsf{T}$, $\mathsf{K4}$, and $\mathsf{S4}$, and shows that the non-self-referentiality of a modal theorem can be sensitive to the choice of logics.

Many properties gained in this chapter about the non-self-referential fragment, even about the prehistoric-loop-free provable fragment as its approximation, are bad in a traditional point of view. This indicates that the notion of non-self-referentiality is not addressed so well by traditional logical tools.

## 3.1 Prehistoric-loop-free Provable Fragments

We start with the definition of *prehistoric-loop-free provable fragments*.

**Definition 46 (Prehistoric-loop-free Provable Fragments)** We write $\mathsf{G3[st4]}^{\otimes} \vdash \Gamma \Rightarrow \Delta$, if sequent $\Gamma \Rightarrow \Delta$ has a loop-free proof in $\mathsf{G3[st4]}$.

The *prehistoric-loop-free provable fragment* of a modal logic $\mathsf{X}$, notation $\mathsf{X}^{\otimes}$, is defined for $\mathsf{T}$, $\mathsf{K4}$, and $\mathsf{S4}$, by

$$\begin{cases} \mathsf{T}^{\otimes} := \{\phi \,|\, \mathsf{G3t}^{\otimes} \vdash\, \Rightarrow \phi\}; \\ \mathsf{K4}^{\otimes} := \{\phi \,|\, \mathsf{G34}^{\otimes} \vdash\, \Rightarrow \phi\}; \\ \mathsf{S4}^{\otimes} := \{\phi \,|\, \mathsf{G3s}^{\otimes} \vdash\, \Rightarrow \phi\}. \end{cases}$$

⊣

As a corollary of Theorem 42, any modal theorem that is self-referential is not loop-free provable. For instance, we have the following corollary for Kuznets' self-referential instances presented in Table 2.1.

**Corollary 47** *(A)* $\neg\Box\neg(p \to \Box p) \notin \mathsf{T}^{\otimes}$*;*

*(B)* $\Box\neg(p \to \Box p) \to \Box\bot \notin \mathsf{K4}^{\otimes}$*;*

*(C)* $\neg\Box\neg(p \to \Box p) \notin \mathsf{S4}^{\otimes}$.

**Proof.** Suppose that, say, $\neg\Box\neg(p \to \Box p) \in \mathsf{S4}^{\otimes}$. By Definition 46, $\mathsf{G3s}^{\otimes} \vdash\, \Rightarrow \phi$, which implies that $\neg\Box\neg(p \to \Box p)$ has a realization in $\mathsf{LP}$ that calls for a constant specification that is not self-referential. This contradicts with Kuznets' result Theorem 16. ⊣

We will concentrate on these loop-free fragments and show their proof-theoretical properties in the rest of this section. To get a start, we observe how prehistoric graphs are changed by rules on sequents. By observing forms of rules and the definition of prehistoric graph, we have what follows.

**Observation 48 (Effects of Rules of Prehistoric Graphs)** $(Ax)$ and $(L\bot)$ each creates a graph with only vertices but no edges.

Applying $(R\to)$ on the root of a proof $\mathcal{T}_0$ gives a larger proof $\mathcal{T}$ with $\mathcal{P}(\mathcal{T}_0) = \mathcal{P}(\mathcal{T})$.

Applying $(L\to)$ on roots of proofs $\mathcal{T}_0$ and $\mathcal{T}_1$, like

$$\frac{\Gamma \Rightarrow \Delta, \alpha \qquad \beta, \Gamma \Rightarrow \Delta}{\alpha \to \beta, \Gamma \Rightarrow \Delta}(L\to),$$

gives a larger proof $\mathcal{T}$ s.t. $\mathcal{P}(\mathcal{T})$ is the unification of $\mathcal{P}(\mathcal{T}_0)$ and $\mathcal{P}(\mathcal{T}_1)$ in the sense that: vertices from side formulas (those in $\Gamma$ and $\Delta$) are unified as formulas are unified; vertices from active formulas $\alpha, \beta$ are preserved; edges are changed along with vertices they go from and to.

In G3s and G3t, $(L\Box)$ is the only rule that can unify two occurrences in a same sequent. Suppose we use $(L\Box)$ on the root of a proof $\mathcal{T}_0$ to get $\mathcal{T}$, like

$$\frac{\theta, \boxminus\theta, \Gamma \Rightarrow \Delta}{\boxminus\theta, \Gamma \Rightarrow \Delta}(L\Box).$$

From $\mathcal{P}(\mathcal{T}_0)$ to $\mathcal{P}(\mathcal{T})$, only vertices in active $\theta$ and $\boxminus\theta$ in the premise are unified; edges are changed along with vertices.

In G3s, $(R\Box)$ is the only rule that can introduce new edges. Suppose we use $(R\Box)$ on the root of a proof $\mathcal{T}_0$ with $\mathcal{P}(\mathcal{T}_0) = (F^0, \prec_L^0, \prec_R^0, \prec^0)$ to get $\mathcal{T}$, like

$$\frac{\boxminus\Theta \Rightarrow \eta}{\boxminus\Theta, \Gamma \Rightarrow \Delta, \boxplus_j\eta}(R\Box).$$

Then $\mathcal{P}(\mathcal{T}) = (F, \prec_L, \prec_R, \prec)$, and in which $F = F^0 \cup \{j\} \cup \{x \mid \boxplus_x \text{ occurs in } \Gamma \text{ or in } \Delta\}$, $\prec_L = \prec_L^0 \cup \{(x, j) \mid \boxplus_x \text{ occurs in } \boxminus\Theta\}$, and $\prec_R = \prec_R^0 \cup \{(x, j) \mid \boxplus_x \text{ occurs in } \eta\}$.

$(K\Box)$ is the only rule in G3t that can introduce new edges. Effects of $(K\Box)$ in G3t is quite same as that of $(R\Box)$ in G3s.

In G34, $(4\Box)$ is the only modal rule. It unifies vertices like what $(L\Box)$ does in G3s, while introducing new edges like what $(R\Box)$ does in G3s. In more detail, applying $(4\Box)$

on $\mathcal{T}_0$ to get $\mathcal{T}$, like

$$\frac{\Theta, \boxminus\Theta \Rightarrow \eta}{\boxminus\Theta, \Gamma \Rightarrow \Delta, \boxplus_j \eta}(4\Box),$$

we get $\mathcal{P}(\mathcal{T})$ by following steps: (i) unify vertices from active $\Theta$ and $\boxminus\Theta$ in the premise while changing edges along with vertices; (ii) add the vertex $j$ and vertices from $\Gamma$ and $\Delta$; (iii) add in $\prec_L$ pairs $(x, j)$ where $\boxplus_x$ occurs in $\boxminus\Theta$; and (iv) add in $\prec_R$ pairs $(x, j)$ where $\boxplus_x$ occurs in $\eta$. ⊣

**Corollary 49** *If $\mathcal{T}$ is loop-free, then any subproof of $\mathcal{T}$ is also loop-free.*

**Proof.** Observation 48, together with an easy observation shows that once a loop is formed in a subproof, there is no rule that can eliminate it (although we may unify vertices to shorten the length of that loop). ⊣

By *unified* formula occurrences, we mean displayed $\Gamma, \Delta$ in the conclusion of $(L \rightarrow)$, displayed $\Box\theta$ in the conclusion of $(L\Box)$, and displayed $\Box\Theta$ in the conclusion of $(4\Box)$. By the observation above, rules have different effects on prehistoric graphs. Those effects include: introduce new vertices, introduce new edges, and unify vertices. It is helpful to see what rules can make a loop, and how.

**Lemma 50** *Given loop-free subproofs $\mathcal{T}_0$ (and $\mathcal{T}_1$):*

*(A) Only* $(L\rightarrow), (L\Box), (4\Box)$ *can make a loop in $\mathcal{P}(\mathcal{T})$.*

*(B) Only* $(L\Box), (4\Box)$ *can make a loop of length* $1$ *in $\mathcal{P}(\mathcal{T})$.*

*(C) A loop is made, only if a pair of vertices from $\mathcal{P}(\mathcal{T}_0)$ (and $\mathcal{P}(\mathcal{T}_1)$), each with edges, are unified.*

**Proof.** By Observation 48, neither $(Ax)$, $(L\bot)$, nor $(R\rightarrow)$ can make a loop.

In $(L \rightarrow)$, vertices from loop-free graphs $\mathcal{P}(\mathcal{T}_0)$ and $\mathcal{P}(\mathcal{T}_1)$ are unified, but without introducing any new edge. Each occurrence from a side formula of one premise is unified with exactly one peer occurrence from the other premise. To make a loop, some chains from $\mathcal{P}(\mathcal{T}_0)$ and $\mathcal{P}(\mathcal{T}_1)$ are connected by unifying the minimal of a chain to the maximal of another chain from the other premise. The simplest case is, we have two chains, each from one premise. That is,

$$1^0 \prec 2^0 \prec \cdots \prec i^0 \text{ in } \mathcal{P}(\mathcal{T}_0) \text{ and } 1^1 \prec 2^1 \prec \cdots \prec j^1 \text{ in } \mathcal{P}(\mathcal{T}_1),$$

with $i^0$ and $1^1$ being unified, $j^1$ and $1^0$ being unified. The loop formed is:

$$1^0 \prec 2^0 \prec \cdots \prec i^0 (\equiv 1^1) \prec 2^1 \prec \cdots \prec j^1 (\equiv 1^0).$$

Since $\mathcal{P}(\mathcal{T}_0)$ is loop free, $1^0, 2^0, \cdots, i^0$ are pairwisely distinct, so are $1^1, 2^1, \cdots, j^1$. In another word, each chain will have two distinct vertices contributed to the loop. Thus, the loop formed has length > 1.

In $(L\Box)$ vertices from $\theta, \Box\theta$ in loop-free $\mathcal{P}(\mathcal{T}_0)$ are unified, but without introducing any new edge. To make a loop, chain(s) from $\mathcal{P}(\mathcal{T}_0)$ are connected by unifying minimal and maximal vertices. The simplest case is, we have only one chain $1 \prec 2 \prec \cdots \prec i$ in $\mathcal{P}(\mathcal{T}_0)$, and then 1 is unified with $i$. The loop formed is

$$1 \prec 2 \prec \cdots \prec i (\equiv 1).$$

If $i = 2$ in $\mathcal{P}(\mathcal{T}_0)$, then the formed loop has length 1.

By Observation 48, $(R\Box)$ and $(K\Box)$ do not unify vertices, and all edges they may add go to the vertex being principally introduced by them (vertex $j$ in Observation 48). This vertex, however, does not exist in $\mathcal{P}(\mathcal{T}_0)$, and hence has out-degree 0 in $\mathcal{P}(\mathcal{T})$. Other vertices and edges from $\mathcal{P}(\mathcal{T}_0)$ are not effected, hence $\mathcal{P}(\mathcal{T})$ is also loop-free.

By Observation 48, the effect of $(4\Box)$ can be seen as a "multi-$(L\Box)$", followed by an $(R\Box)$. Based on previous cases, we know that $(4\Box)$ can make loops by "multi-$(L\Box)$" effect, but the $(R\Box)$-effect cannot make a loop.

In all cases above, *only unification of vertices with edges can make a loop.* ⊣

Loop-free fragments have some nice closure properties. Here is the first one. Recall that in a modal logic, the necessitation rule has the form of

$$(Nec)\frac{\vdash \phi}{\vdash \Box\phi},$$

which means, if $\phi$ is provable as a theorem, then so is $\Box\phi$.

**Theorem 51 (Closure under Necessitation and De-necessitation)** *For any $\phi$:*

*(A)* $\mathsf{G3[st4]}^{\otimes} \vdash \Rightarrow \phi$ *iff* $\mathsf{G3[st4]}^{\otimes} \vdash \Rightarrow \Box\phi$.

*(B) For* $\mathsf{X} \in \{\mathsf{S4}, \mathsf{T}, \mathsf{K4}\}$, *we have* $\phi \in \mathsf{X}^{\otimes}$ *iff* $\Box\phi \in \mathsf{X}^{\otimes}$.

**Proof.** (A) For one direction, suppose that we have a loop-free proof of $\Rightarrow \phi$. Applying an $(R\Box)$, $(K\Box)$, or $(4\Box)$ (depending on the calculus) on the root sequent gives a proof of $\Rightarrow \Box\phi$. By Lemma 50, neither $(R\Box)$ nor $(K\Box)$ can make a loop. In the case that the rule applied is $(4\Box)$, the antecedent (where unification happens) is empty. Thus there is no unification to make a loop.

For the other direction, suppose that we have a loop-free proof of $\Rightarrow \Box\phi$. By an observation of all our rules, we know that the last rule of that proof is $(R\Box)$, $(K\Box)$, or $(4\Box)$ (depending on the calculus), and the premise of that rule is $\Rightarrow \phi$. By Corollary 49, the subproof with $\Rightarrow \phi$ as root is also loop-free.

(B) A direct consequence of (A). ⊣

Other closure properties will be presented once we are ready to prove them. We now define *isolated* occurrence, which is a formula occurrence with all vertices from it having no edge, and hence is “safe” for unification.

**Definition 52 (Isolated Occurrence)** In proof $\mathcal{T}$, a (formula or subformula) occurrence $\phi$ in the root is *isolated*, if all vertices from $\phi$ have no edges. More precisely, if in $\mathcal{P}(\mathcal{T}) = (F, \prec_L, \prec_R, \prec)$, we have $\prec \subseteq (F \backslash F^{\phi}) \times (F \backslash F^{\phi})$, where $F^{\phi} := \{i \in F \mid \boxplus_i \text{ occurs in } \phi\}$. ⊣

**Lemma 53** *Applying a rule* $(R)$ *on loop-free subproof(s) to get* $\mathcal{T}$. *Assume that each unified formula of* $(R)$ *has a corresponding occurrence being isolated in the subproof of the premise it resides, then* $\mathcal{T}$ *is also loop-free.*

**Proof.** Each pair of vertices being unified occur in a pair of formula-occurrences being unified. By assumption, in each such pair of occurrences, there is one occurrence being isolated in the premise it comes from. Vertices from this occurrence has no edge in the prehistoric graph of that premise. By Lemma 50, to make a loop, we need at least one pair of vertices, each with edge in the graph of premise, to be unified. Now we see this is impossible under our assumption. ⊣

**Lemma 54** *Suppose that we apply* $(L\rightarrow)$, $(R\rightarrow)$, *or* $(L\Box)$, *on* $\mathcal{T}_0$ *(and* $\mathcal{T}_1$*) to get* $\mathcal{T}$. *If a formula occurrence* $\phi$ *in the root of* $\mathcal{T}$ *has all of its corresponding occurrences isolated in* $\mathcal{T}_0$ *(and* $\mathcal{T}_1$*), then* $\phi$ *is isolated in* $\mathcal{T}$.

**Proof.** By Observation 48, none of $(L\rightarrow)$, $(R\rightarrow)$, and $(L\Box)$ can introduce new edges. Thus each edge in $\mathcal{P}(\mathcal{T})$ has corresponding edge(s) from at least one premise. If $\phi$ is not isolated in $\mathcal{T}$, then there is an edge going from (or to) a vertex from $\phi$. This edge has a corresponding edge, w.l.o.g., in $\mathcal{P}(\mathcal{T}_0)$, which means that $\phi$ is not isolated in $\mathcal{T}_0$, a contradiction. ⊣

To highlight isolated formula occurrences, we employ a notation. For proof $\mathcal{T}$ and sequent $s$ in $\mathcal{T}$, if $\phi$ is isolated in $\mathcal{T} \upharpoonright s$, then we highlight this fact by putting this $\phi$ in a pair of $\langle\rangle$, like $\langle\phi\rangle$, in $s$. Recall that $\mathcal{T} \upharpoonright s$ is the subproof of sequent $s$ in $\mathcal{T}$.

**Theorem 55 (Loop-free Admissible of Weakening)** *(A) If sequent $\Gamma \Rightarrow \Delta$ has a proof $\mathcal{T}$ with prehistoric graph $\mathcal{P}(\mathcal{T})$, then sequent $\Gamma, \Gamma' \Rightarrow \Delta', \Delta$ has a proof $\mathcal{T}'$ of the same height, and with prehistoric graph $\mathcal{P}(\mathcal{T}')$ that is a disjoint union of an isomorphic copy of $\mathcal{P}(\mathcal{T})$ and a graph with vertices from $\Gamma'$ or from $\Delta'$ and no edges. In another word, edges in $\mathcal{P}(\mathcal{T}')$ are exactly edges in $\mathcal{P}(\mathcal{T})$, and formulas in $\Gamma', \Delta'$ are isolated.*

*(B)* $\mathsf{G3[st4]}^{\otimes} \vdash_n \Gamma \Rightarrow \Delta$ *implies* $\mathsf{G3[st4]}^{\otimes} \vdash_n \Gamma, \Gamma' \Rightarrow \Delta', \Delta$.

**Proof.** (A) Consider the following algorithm.

Start by adding $\Gamma'$ and $\Delta'$ at the root of $\mathcal{T}$. Go from conclusion to premises of each rule in $\mathcal{T}$, step by step, put $\Gamma'$ and $\Delta'$ as side formulas in more and more sequents in $\mathcal{T}$, until there is an $(Ax)$, $(L\bot)$, $(R\Box)$, $(K\Box)$, or $(4\Box)$, where $\Gamma'$ and $\Delta'$ can been seen as weakening formulas introduced.

Note that in the resulting proof $\mathcal{T}'$, all occurrences of $\Gamma'$ and $\Delta'$ are to the root-side of any $(R\Box)$, $(K\Box)$, or $(4\Box)$. Occurrences $\Gamma'$ and $\Delta'$ are introduced as weakening formulas, and hence are isolated there. Then they are either inherited, or unified with only other occurrences of $\Gamma'$ and $\Delta'$, which are also isolated, by only $(L\to)$, $(R\to)$, and $(L\Box)$. By Lemma 54, $\Gamma'$ and $\Delta'$ are isolated everywhere they appear.

Therefore, in $\mathcal{P}(\mathcal{T}')$, vertices from $\Gamma'$ and $\Delta'$ have no edges, which means all edges in $\mathcal{P}(\mathcal{T}')$ are inherited from $\mathcal{P}(\mathcal{T})$.

(B) An easy consequence of (A). ⊣

We have shown in Theorem 28 that our $\mathsf{G3}$-calculi are closed under weakening. Now Theorem 55 says that applying the derivable weakening will not introduce loop in prehistoric graph. From now on, we use $(W)$ to denote an application of weakening:

$$\frac{\Gamma \Rightarrow \Delta}{\Gamma, \langle\Gamma'\rangle \Rightarrow \langle\Delta'\rangle, \Delta}(W).$$

This rule is allowed anywhere in a loop-free proof.

In G3[st], the $(L\Box)$ rule has the form of

$$\frac{\theta, \Box\theta, \Gamma \Rightarrow \Delta}{\Box\theta, \Gamma \Rightarrow \Delta} \quad (L\Box).$$

By $(L\Box W)$, we mean an admissible rule of the form

$$\frac{\theta, \Gamma \Rightarrow \Delta}{\Box\theta, \Gamma \Rightarrow \Delta} \quad (L\Box W),$$

which can be seen as an abbreviation of

$$\cfrac{\cfrac{\theta, \Gamma \Rightarrow \Delta}{\theta, \langle\Box\theta\rangle, \Gamma \Rightarrow \Delta}(W)}{\Box\theta, \Gamma \Rightarrow \Delta}(L\Box).$$

According to Observation 48, the only effect of $(L\Box)$ is the unification of the two active $\theta$'s in the premise (one by its own, one as a subformula of $\Box\theta$). In the application of $(L\Box)$ in $(L\Box W)$, one of the unified $\theta$-occurrence is from the isolated $\Box\theta$ just introduced by the $(W)$, and hence the application of $(L\Box W)$ on the root of a proof $\mathcal{T}_0$ gives a larger proof $\mathcal{T}$ with $\mathcal{P}(\mathcal{T}_0) = \mathcal{P}(\mathcal{T})$. Specifically, if $\mathcal{P}(\mathcal{T}_0)$ is loop-free, then so is $\mathcal{P}(\mathcal{T})$.

For a similar reason, in In G34, we have the admissible rule

$$\frac{\Theta \Rightarrow \eta}{\Box\Theta, \Gamma \Rightarrow \Delta, \Box\eta} \quad (4\Box W),$$

which can be seen as an abbreviation of

$$\cfrac{\cfrac{\Theta \Rightarrow \eta}{\Theta, \langle\Box\Theta\rangle \Rightarrow \eta}(W)}{\Box\Theta, \Gamma \Rightarrow \Delta, \Box\eta}(4\Box),$$

and does not make a loop.

In this thesis, $\neg\eta$ is defined as an abbreviation of $\eta \to \bot$. This saves us rules about $\neg$ in our calculi. If one likes to explicitly give those rules, they will have the following form:

$$\frac{\Gamma \Rightarrow \Delta, \eta}{\neg\eta, \Gamma \Rightarrow \Delta}(L\neg) \qquad \frac{\eta, \Gamma \Rightarrow \Delta}{\Gamma \Rightarrow \Delta, \neg\eta}(R\neg).$$

Within them, $(L\neg)$ can be seen as an abbreviation of

$$\dfrac{\overline{\bot, \langle\Gamma\rangle \Rightarrow \langle\Delta\rangle} \qquad \Gamma \Rightarrow \Delta, \eta}{\eta \to \bot, \Gamma \Rightarrow \Delta}(L\to),$$

where one copy of side formulas unified by $(L\to)$ is isolated. Thus, an application of $(L\neg)$ cannot make a loop. For $(R\neg)$, we understand it as an abbreviation of

$$\dfrac{\dfrac{\eta, \Gamma \Rightarrow \Delta}{\eta, \Gamma \Rightarrow \Delta, \bot}(W)}{\Gamma \Rightarrow \Delta, \eta \to \bot}(R\to),$$

where $(W)$ even does not introduce a vertex and $(R\to)$ simply preserves the graph. Thus, an application of $(R\neg)$ cannot make a loop.

In summary, the we have the following fact.

**Fact 56 (Loop-free Admissible Rules)** *All of* $(Ax)$, $(L\bot)$, $(L\neg)$, $(R\neg)$, $(R\to)$, $(R\Box)$, $(K\Box)$, $(W)$, $(L\Box W)$, *and* $(4\Box W)$ *(whenever applicable) are admissible in* $\mathsf{G3[st4]}^{\otimes}$.

**Lemma 57** $\mathsf{G3[st4]}^{\otimes} \vdash \phi \Rightarrow \phi$ *for any* $\phi$.

**Proof.** Induction on $\phi$.

(i) Trivial for prime formulas.

(ii) For $\phi \equiv \alpha\to\beta$, consider the following proof:

$$\dfrac{\dfrac{\dfrac{\alpha \Rightarrow \alpha}{\alpha \Rightarrow \alpha, \langle\beta\rangle}(W) \qquad \dfrac{\beta \Rightarrow \beta}{\langle\alpha\rangle, \beta \Rightarrow \beta}(W)}{\alpha\to\beta, \alpha \Rightarrow \beta}(L\to),}{\alpha\to\beta \Rightarrow \alpha\to\beta}(R\to)$$

where $\alpha \Rightarrow \alpha$ and $\beta \Rightarrow \beta$ are given by IH. By Theorem 55, $(W)$ is allowed to introduce isolated formulas. In the displayed $(L\to)$, a pair of side $\alpha$-occurrences are unified, with the occurrence from right premise being isolated. Similarly for a pair of side $\beta$'s. By Lemma 53, we can apply this $(L\to)$ without introducing a loop in prehistoric graph.

(iii) For $\phi \equiv \Box\psi$, consider the following three proofs in G3s, G3t, and G34 respectively:

$$\dfrac{\dfrac{\dfrac{\psi \Rightarrow \psi}{\langle \boxminus\psi \rangle, \psi \Rightarrow \psi}(W)}{\boxminus\psi \Rightarrow \psi}(L\Box)}{\boxminus\psi \Rightarrow \boxplus\psi}(R\Box), \quad \dfrac{\psi \Rightarrow \psi}{\boxminus\psi \Rightarrow \boxplus\psi}(K\Box), \quad \dfrac{\dfrac{\psi \Rightarrow \psi}{\langle \boxminus\psi \rangle, \psi \Rightarrow \psi}(W)}{\boxminus\psi \Rightarrow \boxplus\psi}(4\Box),$$

where $\psi \Rightarrow \psi$ follows from IH, $(W)$ is allowed by Theorem 55, $(L\Box)$ and $(4\Box)$ is allowed by Lemma 53, $(R\Box)$ and $(K\Box)$ cannot make a loop (by Lemma 50). $\dashv$

We are ready to show the last closure property in this section, the closure under uniform substitution.

**Theorem 58 (Closure under Substitution)** *For any $\phi$, $\psi$, $\Gamma$, $\Delta$, and atom $p$:*

*(A)* $\mathsf{G3[st4]}^{\otimes} \vdash \Gamma \Rightarrow \Delta$ *implies* $\mathsf{G3[st4]}^{\otimes} \vdash \Gamma[p/\psi] \Rightarrow \Delta[p/\psi]$.

*(B)* $\phi \in \mathsf{X}^{\otimes}$ *implies* $\phi[p/\psi] \in \mathsf{X}^{\otimes}$ *for* $\mathsf{X} \in \{\mathsf{S4}, \mathsf{T}, \mathsf{K4}\}$.

**Proof.** (A) We present an effective procedure to generate a proof of $\Gamma[p/\psi] \Rightarrow \Delta[p/\psi]$, and then show that it is actually loop-free.

By assumption, we have a $\mathsf{G3[st4]}^{\otimes}$ proof of $\Gamma \Rightarrow \Delta$. Denote this proof by $\mathcal{T}_0$. Substitute each occurrence of $p$ in $\mathcal{T}_0$ by $\psi$, and denote the resulting tree by $\mathcal{T}_1$. Clearly, all applications of rules except $(Ax)$'s with $p$ as principal formula become an application of the same rule in $\mathcal{T}_1$, and each application of $(Ax)$ with form

$$\overline{p, \Gamma_0 \Rightarrow \Delta_0, p}$$

becomes $\psi, \Gamma_0[p/\psi] \Rightarrow \Delta_0[p/\psi], \psi$, which is temporally called a *principal-$\psi$ sequent*. That is, except principal-$\psi$ sequents, $\mathcal{T}_1$ is a G3[st4]-proof of $\Gamma[p/\psi] \Rightarrow \Delta[p/\psi]$.

By Lemma 57, there is a loop-free proof of $\psi \Rightarrow \psi$. Denote it by $\mathcal{T}_\psi$. Substitute each principal-$\psi$ sequent with form $\psi, \Gamma_0[p/\psi] \Rightarrow \Delta_0[p/\psi], \psi$ in $\mathcal{T}_1$ by

$$\dfrac{\begin{matrix}\mathcal{T}_\psi \\ \psi \Rightarrow \psi\end{matrix}}{\psi, \langle \Gamma_0[p/\psi] \rangle \Rightarrow \langle \Delta_0[p/\psi] \rangle, \psi}(W),$$

we get $\mathcal{T}_2$. Since $\mathcal{T}_\psi$ is loop-free, and $(W)$ is eliminable without introducing any loop (Theorem 55), we know that this gives principal-$\psi$ sequents $\mathsf{G3[st4]}^{\otimes}$-proofs in $\mathcal{T}_2$, and hence $\mathcal{T}_2$ is a $\mathsf{G3[st4]}$-proof of $\Gamma[p/\psi] \Rightarrow \Delta[p/\psi]$. What remains is to show that $\mathcal{T}_2$ is loop-free.

For each sequent $s$ in $\mathcal{T}_2$, let $\mathcal{P}(\mathcal{T}_2 \upharpoonright s) = (F^s, \prec^s_L, \prec^s_R, \prec^s)$ be the graph of the subproof of $s$ in $\mathcal{T}_2$. Sequent $s$ in $\mathcal{T}_2$ that is not above any principal-$\psi$ sequent has corresponding sequent in $\mathcal{T}_0$, which is denoted by $s'$. For sequent $s$ not above any principal-$\psi$ sequent, let $F^s_\psi$ be the set of vertices in $\mathcal{P}(\mathcal{T}_2 \upharpoonright s)$ that are from occurrences of $\psi$'s substituted in,[1] and let

$$\mathcal{P}_\psi(\mathcal{T}_2 \upharpoonright s) := (F^s_\psi, \prec^s_{L\psi}, \prec^s_{R\psi}, \prec^s_\psi)$$

where $\rhd^s_\psi := \rhd^s \upharpoonright (F^s_\psi \times F^s_\psi)$ for $\rhd \in \{\prec_L, \prec_R, \prec\}$, be the subgraph of $\mathcal{P}(\mathcal{T}_2 \upharpoonright s)$ that contains exactly those vertices in $F^s_\psi$ and edges in $(F^s_\psi \times F^s_\psi)$. Also let

$$\mathcal{P}_{\overline{\psi}}(\mathcal{T}_2 \upharpoonright s) := (F^s_{\overline{\psi}}, \prec^s_{L\overline{\psi}}, \prec^s_{R\overline{\psi}}, \prec^s_{\overline{\psi}})$$

where $F^s_{\overline{\psi}} := F^s \setminus F^s_\psi$, and $\prec^s_{L\overline{\psi}} := \prec^s_L \upharpoonright (F^s_{\overline{\psi}} \times F^s_{\overline{\psi}})$, etc.. Note that $F^s$ is a disjoint union of $F^s_\psi$ and $F^s_{\overline{\psi}}$, but it is more complex for edges. For instance, $\prec^s_L$ may contains edges in neither $\prec^s_{L\psi}$ nor $\prec^s_{L\overline{\psi}}$, since there may be edges go from a vertex in $F^s_\psi$ to a vertex in $F^s_{\overline{\psi}}$. Temporally call these edges *cross edges*.

Now we are ready to show that $\mathcal{T}_2$ is loop-free. Any sequent $s$ above a principal-$\psi$ sequent is in the loop-free subproof $\mathcal{T}_\psi$. By Corollary 49, $\mathcal{P}(\mathcal{T}_2 \upharpoonright s)$ is loop-free. For any sequent $s$ not above any principal-$\psi$ sequent in $\mathcal{T}_2$, we prove this by three claims.

**Claim 59** *There is no loop inside $\mathcal{P}_{\overline{\psi}}(\mathcal{T}_2 \upharpoonright s)$.*

[1] Thus vertices from an occurrence of $\psi$ that has already occurred in $s'$ are not included. This distinction is clear since a position occupied by a substituted-in $\psi$ in $s$ is occupied by an occurrence of $p$ in $s'$, and this occurrence of $p$ is unified with only other occurrences of $p$ in $\mathcal{T}_0$, which implies the substituted-in $\psi$ is unified with only other substituted-in $\psi$'s in $\mathcal{T}_2$.

**Proof.** All vertices and edges involved by the substitution is not included in, and hence $\mathcal{P}_{\overline{\psi}}(\mathcal{T}_2 \upharpoonright s)$ is isomorphic to $\mathcal{P}(\mathcal{T}_0 \upharpoonright s')$, which is loop-free. $\dashv$

**Claim 60** *There is no cross edge that goes from a vertex in $F^s_{\overline{\psi}}$ to a vertex in $F^s_\psi$.*

**Proof.** Assume there is such an edge $(i, j)$ where $i \in F^s_{\overline{\psi}}$ and $j \in F^s_\psi$. By Definition 34, this means $\boxplus_i$ occurs in the premise of a rule $(R)$ of $\mathsf{G3[st4]}$ that principally introduces $\boxplus_j$. Now $j \in F^s_\psi$ implies $\boxplus_j$ occurs in a substituted-in $\psi$, and hence the application of rule $(R)$ is in $\mathcal{T}_\psi$. Thus $\boxplus_i$ occurs $\mathcal{T}_\psi$, and also its root sequent $\psi \Rightarrow \psi$. This means that $\boxplus_i$ also occurs in a substituted-in $\psi$, which contradicts with $i \in F^s_{\overline{\psi}}$. $\dashv$

**Claim 61** *There is no loop inside $\mathcal{P}_\psi(\mathcal{T}_2 \upharpoonright s)$.*

**Proof.** To prove this claim, it is sufficient to care about only vertices from substituted-in $\psi$'s. Since we are talking about sequent $s$ with corresponding $s'$ in $\mathcal{T}_0$, we can make an induction on $\mathcal{T}_0$. By this induction, we will show that (abbreviated by (*) later in this proof):

> In $\mathcal{P}_\psi(\mathcal{T}_2 \upharpoonright s) = (F^s_\psi, \prec^s_{L\psi}, \prec^s_{R\psi}, \prec^s_\psi)$, for any edge $e \in \prec^s_\psi$, there is a subgraph[2] $\mathcal{G} = (F_{\psi^-} \cup F_{\psi^+}, \prec^{\mathcal{G}}_L, \prec^{\mathcal{G}}_R, \prec^{\mathcal{G}})$ of $\mathcal{P}_\psi(\mathcal{T}_2 \upharpoonright s)$, where $F_{\psi^-}$ (resp. $F_{\psi^+}$) is the collection of vertices from $\psi^-$ (resp. $\psi^+$), a negative (resp. positive) occurrence of substituted-in $\psi$ in $s$, s.t. $\mathcal{G} \cong \mathcal{P}(\mathcal{T}_\psi)$ and $e \in \prec^{\mathcal{G}}$.

By (*), each edge in $\mathcal{P}_\psi(\mathcal{T}_2 \upharpoonright s)$ is in an isomorphic copy of $\mathcal{P}(\mathcal{T}_\psi)$, and hence has its isomorphic image in $\mathcal{P}(\mathcal{T}_\psi)$. If $\mathcal{P}_\psi(\mathcal{T}_2 \upharpoonright s)$ has a loop, then isomorphic images of edges that constitute this loop would form a loop in $\mathcal{P}(\mathcal{T}_\psi)$, which contradicts with the fact that $\mathcal{P}(\mathcal{T}_\psi)$ is loop-free.

---

[2] $\mathcal{G}_1 = (F^{\mathcal{G}_1}, \prec^{\mathcal{G}_1}_L, \prec^{\mathcal{G}_1}_R, \prec^{\mathcal{G}_1})$ is a subgraph of $\mathcal{G}_2 = (F^{\mathcal{G}_2}, \prec^{\mathcal{G}_2}_L, \prec^{\mathcal{G}_2}_R, \prec^{\mathcal{G}_2})$, if $F^{\mathcal{G}_1} \subseteq F^{\mathcal{G}_2}$, $\prec^{\mathcal{G}_1}_L \subseteq \prec^{\mathcal{G}_2}_L$, $\prec^{\mathcal{G}_1}_R \subseteq \prec^{\mathcal{G}_2}_R$, and $\prec^{\mathcal{G}_1} \subseteq \prec^{\mathcal{G}_2}$.

We finish this proof by presenting the induction that verifies (*):

For $s'$ got by an application of $(Ax)$ with principal formula $p$, $s$ is a principal-$\psi$ sequent. If $s'$ is $\Gamma, p \Rightarrow p, \Delta$, then $s$ is $\Gamma[p/\psi], \psi \Rightarrow \psi, \Delta[p/\psi]$. Recall that we use copies of a same proof $\mathcal{T}_\psi$ to get $\psi \Rightarrow \psi$ throughout the whole procedure, and the proof of $s$ is got by further applying a $(W)$ that introduces $\Gamma[p/\psi]$ and $\Delta[p/\psi]$. By Theorem 55, this application of $(W)$ can introduce only vertices but no edges. Hence $\mathcal{P}_\psi(\mathcal{T}_2 \restriction s)$ is a disjoint union of a graph $\mathcal{G} \cong \mathcal{P}(\mathcal{T}_\psi)$ and a graph with vertices from substituted-in $\psi$ occurrences in $\Gamma[p/\psi]$ and $\Delta[p/\psi]$ but no edges. Thus all edges are in $\mathcal{G}$. Also note that vertices of $\mathcal{G}$ are from the pair of displayed $\psi$ occurrences in $\Gamma[p/\psi], \psi \Rightarrow \psi, \Delta[p/\psi]$.

For $s'$ got by an application of $(L\bot)$, or of $(Ax)$ with principal formula other than $p$. Since $p$ is not principal in $s'$, we get $s = s'[p/\psi]$ by an application of the same rule. By Observation 48, $\mathcal{P}(\mathcal{T}_2 \restriction s)$ has only vertices but no edges, and hence (*) is vacuously true.

For $s'$ got by an application of rule $(R)$ with premise(s) $s'_0$ (and $s'_1$). For any $e \in \prec^s_\psi$, $e = (i, j)$ is an edge between two vertices $i$ and $j$ from substituted-in $\psi$('s). Assume that $e$ has no corresponding edges in $\prec^{s_0}_\psi \cup \prec^{s_1}_\psi$. Then $e$ is introduced by $(R)$, and hence Observation 48 tells us that $(R) \in \{(R\Box), (K\Box), (4\Box)\}$, and the positive principal formula has the form of $\boxplus_j \eta$. Now the fact that $j$ is from a substituted-in $\psi$ implies that this substituted-in $\psi$ is to be formed by its subformluas at or under sequent $s$ in $\mathcal{T}_2$. This means that an atom $p$ is to be formed by its subformulas at or under sequent $s'$ in $\mathcal{T}_0$, which is impossible. Therefore, the assumption is not true, and $e$ has a corresponding edge $e'$ in

$\prec_{\psi}^{s_0} \cup \prec_{\psi}^{s_1}$. In what follows, we assume $e' \in \prec_{\psi}^{s_0}$, w.l.o.g..

By IH, there is a subgraph $\mathcal{G}_0 = (F_{\psi_0^-} \cup F_{\psi_0^+}, \prec_L^{\mathcal{G}_0}, \prec_R^{\mathcal{G}_0}, \prec^{\mathcal{G}_0})$ of $\mathcal{P}_{\psi}(\mathcal{T}_2 \upharpoonright s_0)$, where $F_{\psi_0^-}$ and $F_{\psi_0^+}$ are collections of vertices from a pair of substituted-in $\psi$ occurrences (denoted by $\psi_0^-$ and $\psi_0^+$) in $s_0$, s.t. $\mathcal{G}_0 \cong \mathcal{P}(\mathcal{T}_{\psi})$ and $e' \in \prec^{\mathcal{G}_0}$.

$\psi_0^-$ and $\psi_0^+$ have corresponding occurrences in $s$. Denote them by $\psi^-$ and $\psi^+$, and collections of vertices of $\mathcal{P}_{\psi}(\mathcal{T}_2 \upharpoonright s)$ from $\psi^-$ and $\psi^+$ by $F_{\psi^-}$ and $F_{\psi^+}$. All rules respect polarity, and hence no pair of vertices from $F_{\psi_0^-} \cup F_{\psi_0^+}$ are unified. Thus, vertices of $\mathcal{G}_0$ are bijectively mapped by the correspondence to $F_{\psi^-} \cup F_{\psi^+}$. Edges are changed along with vertices they go from and to. Hence edges of $\mathcal{G}_0$ in $\prec_L^{\mathcal{G}_0}$, $\prec_R^{\mathcal{G}_0}$, and $\prec^{\mathcal{G}_0}$ are injectively mapped by the correspondence to $\mathcal{P}_{\psi}(\mathcal{T}_2 \upharpoonright s)$. Denote images of $\prec_L^{\mathcal{G}_0}$, $\prec_R^{\mathcal{G}_0}$, $\prec^{\mathcal{G}_0}$ by $\prec_L^{\mathcal{G}}$, $\prec_R^{\mathcal{G}}$, $\prec^{\mathcal{G}}$, respectively. Now $\mathcal{G} = (F_{\psi^-} \cup F_{\psi^+}, \prec_L^{\mathcal{G}}, \prec_R^{\mathcal{G}}, \prec^{\mathcal{G}})$ is an isomorphic image of $\mathcal{G}_0$, and hence is an isomorphic image of $\mathcal{P}(\mathcal{T}_{\psi})$. Note that $e'$ is the corresponding edge of $e$ in $\prec_{\psi}^{s_0}$, located in its subset $\prec^{\mathcal{G}_0}$, which implies $e \in \prec^{\mathcal{G}}$.

Therefore, for an arbitrary edge $e \in \prec_{\psi}^{s}$, we have found a subgraph $\mathcal{G}$ of $\mathcal{P}_{\psi}(\mathcal{T}_2 \upharpoonright s)$, with all its vertices (in $F_{\psi^-} \cup F_{\psi^+}$) from a pair of negative and positive occurrences of substituted-in $\psi$ in $s$, s.t. $\mathcal{G} \cong \mathcal{P}(\mathcal{T}_{\psi})$ and $e \in \prec^{\mathcal{G}}$.

This completes our induction.

$\dashv$

By the above three claims, we know that $\mathcal{P}(\mathcal{T}_2 \upharpoonright s)$ is loop-free. Suppose a loop exists, and that loop has a vertex in $F_{\overline{\psi}}^{s}$. Then Claim 60 prevent this loop to visit any vertex in $F_{\psi}^{s}$. This requires the loop to be in $\mathcal{P}_{\overline{\psi}}(\mathcal{T}_2 \upharpoonright s)$, which contradicts Claim 59. Thus any potential loop cannot has any vertex in $F_{\overline{\psi}}^{s}$. This requires the loop to be in $\mathcal{P}_{\psi}(\mathcal{T}_2 \upharpoonright s)$, which contradicts Claim 61. In summary, no loop exists in $\mathcal{P}(\mathcal{T}_2 \upharpoonright s)$. This prove is finished, by letting $s$ to be the end-sequent of $\mathcal{T}_2$.

(B) An easy consequence of (A). $\dashv$

For ordered sets $P = (p_1, p_2, \cdots, p_k)$ and $\Psi = (\psi_1, \psi_2, \cdots, \psi_k)$, let $\phi[P/\Psi]$ be an abbreviation of $\phi[p_1/\psi_1][p_2/\psi_2]\cdots[p_k/\psi_k]$. Then we have the following corollary.

**Corollary 62** *For any propositional tautology $\phi$ (in the propositional language), atoms $P$, and modal formulas $\Psi$, we have* $\mathsf{G3[st4]}^{\otimes} \vdash \Rightarrow \phi[P/\Psi]$*, and hence* $\phi[P/\Psi] \in \mathsf{X}^{\otimes}$ *for* $\mathsf{X} \in \{\mathsf{S4}, \mathsf{T}, \mathsf{K4}\}$.

**Proof.** Since $\phi$ is a tautology in the propositional language, $\mathsf{G3cp} \vdash \Rightarrow \phi$. Note that each $\mathsf{G3cp}$-proof is also a $\mathsf{G3[st4]}$-proof, and viewed as a $\mathsf{G3[st4]}$-proof, the absence of modal rules makes the proof loop-free. By Theorem 58(A), $\mathsf{G3[st4]}^{\otimes} \vdash \Rightarrow \phi[P/\Psi]$, which further implies $\phi[P/\Psi] \in \mathsf{X}^{\otimes}$. $\dashv$

Consequences similar to Corollary 62 hold for modal cases.

**Lemma 63** *For any modal formula $\phi$ and $\psi$:*

*(A)* $\Box\phi \to \phi \in \mathsf{T}^{\otimes}$*;*

*(B)* $\Box\phi \to \Box\Box\phi \in \mathsf{K4}^{\otimes}$*;*

*(C)* $\Box(\phi \to \psi) \to (\Box\phi \to \Box\psi) \in \mathsf{K4}^{\otimes} \cap \mathsf{T}^{\otimes}$.

**Proof.** (A) In $\mathsf{G3t}$, the proof

$$\dfrac{\dfrac{\dfrac{p \Rightarrow p}{\langle \boxminus p \rangle, p \Rightarrow p}(W)}{\boxminus p \Rightarrow p}(L\Box)}{\Rightarrow \boxminus p \to p}(R\to)$$

is loop-free. By Theorem 58(B), $\Box\phi \to \phi \in \mathsf{T}^{\otimes}$.

(B) In $\mathsf{G34}$ the proof

$$\dfrac{\dfrac{\dfrac{\boxminus p \Rightarrow \boxplus p}{p, \boxminus p \Rightarrow \boxplus p}(W)}{\boxminus p \Rightarrow \boxplus\boxplus p}(4\Box)}{\Rightarrow \boxminus p \to \boxplus\boxplus p}(R\to)$$

is loop-free. By Theorem 58(B), $\Box\phi\to\Box\Box\phi\in\mathsf{K4}^{\otimes}$.

(C) The proof

$$\dfrac{\dfrac{\dfrac{\dfrac{p\Rightarrow p}{p\Rightarrow\langle q\rangle,p}(W)\qquad\dfrac{q\Rightarrow q}{q,\langle p\rangle\Rightarrow q}(W)}{p\to q,p\Rightarrow q}(L\to)}{\boxminus(p\to q),\boxminus p\Rightarrow\boxplus q}(K\Box)}{\Rightarrow\boxminus(p\to q)\to(\boxminus p\to\boxplus q)}(\overrightarrow{R\to})$$

in $\mathsf{G3t}$ and the proof

$$\dfrac{\dfrac{\dfrac{\dfrac{p\Rightarrow p}{\langle\boxminus(p\to q)\rangle,\langle\boxminus p\rangle,p\Rightarrow\langle q\rangle,p}(W)\qquad\dfrac{q\Rightarrow q}{\langle\boxminus(p\to q)\rangle,\langle\boxminus p\rangle,q,\langle p\rangle\Rightarrow q}(W)}{\langle\boxminus(p\to q)\rangle,\langle\boxminus p\rangle,p\to q,p\Rightarrow q}(L\to)}{\boxminus(p\to q),\boxminus p\Rightarrow\boxplus q}(4\Box)}{\Rightarrow\boxminus(p\to q)\to(\boxminus p\to\boxplus q)}(\overrightarrow{R\to})$$

in $\mathsf{G34}$ are both loop-free. By Theorem 58(B), $\Box(\phi\to\psi)\to(\Box\phi\to\Box\psi)\in\mathsf{K4}^{\otimes}\cap\mathsf{T}^{\otimes}$. ⊣

In fact, similar arguments show, e.g., $\Box\phi\to\phi\in\mathsf{S4}^{\otimes}$. But we prefer to put it as a consequence of the following fact that the inclusion of sub-logics is preserved also for the loop-free setting, or in another word, $\mathsf{G3s}$ has at least the same ability of avoiding loops as $\mathsf{G3t}$ and $\mathsf{G34}$.

**Theorem 64** *(A)* $\mathsf{G3t}^{\otimes}\vdash\Gamma\Rightarrow\Delta$ *implies* $\mathsf{G3s}^{\otimes}\vdash\Gamma\Rightarrow\Delta$ *for any* $\Gamma$ *and* $\Delta$*;*

*(B)* $\mathsf{G34}^{\otimes}\vdash\Gamma\Rightarrow\Delta$ *implies* $\mathsf{G3s}^{\otimes}\vdash\Gamma\Rightarrow\Delta$ *for any* $\Gamma$ *and* $\Delta$*;*

*(C)* $\mathsf{T}^{\otimes}\subsetneq\mathsf{S4}^{\otimes}$*;*

*(D)* $\mathsf{K4}^{\otimes}\subsetneq\mathsf{S4}^{\otimes}$.

**Proof.** (A) Given a loop-free proof in $\mathsf{G3t}$, by a substitution for each $(K\Box)$ like:

$$\dfrac{\Theta\Rightarrow\eta}{\boxminus\Theta,\Gamma\Rightarrow\Delta,\boxplus\eta}(K\Box)\quad\rightsquigarrow\quad\dfrac{\dfrac{\dfrac{\Theta\Rightarrow\eta}{\langle\boxminus\Theta\rangle,\Theta\Rightarrow\eta}(W)}{\boxminus\Theta\Rightarrow\eta}(\overrightarrow{L\Box})}{\boxminus\Theta,\Gamma\Rightarrow\Delta,\boxplus\eta}(R\Box),$$

we get a loop-free proof in $\mathsf{G3s}$ of the same conclusion. To see this, by IH, we have a loop-free proof of $\Theta\Rightarrow\eta$ in $\mathsf{G3s}$, then Theorem 55, Lemma 53, and Lemma 50 allow us to derive like displayed without introducing a loop.

(B) Given a loop-free proof in G34, apply a substitution for each (4□) like:

$$\dfrac{\Theta, \boxminus\Theta \Rightarrow \eta}{\boxminus\Theta, \Gamma \Rightarrow \Delta, \boxplus\eta}(4\Box) \quad \rightsquigarrow \quad \dfrac{\dfrac{\Theta, \boxminus\Theta \Rightarrow \eta}{\boxminus\Theta \Rightarrow \eta}(\overrightarrow{L\Box})}{\boxminus\Theta, \Gamma \Rightarrow \Delta, \boxplus\eta}(R\Box).$$

By Observation 48, both sides have exactly the same effect on prehistoric graphs, so we get a loop-free proof in G3s of the same conclusion.

(C) $\mathsf{T}^{\otimes} \subseteq \mathsf{S4}^{\otimes}$ is a direct consequence of (A), and it is sufficient to find a formula in $\mathsf{S4}^{\otimes}$ but not in $\mathsf{T}^{\otimes}$. The desired formula is $\Box p \to \Box\Box p$, obviously not in T, and hence not in $\mathsf{T}^{\otimes}$. Lemma 63(B) shows $\Box p \to \Box\Box p \in \mathsf{K4}^{\otimes}$, and hence $\Box p \to \Box\Box p \in \mathsf{S4}^{\otimes}$ by (B).

(D) Similar to (C). ⊣

We have just shown that the loop-free provable fragment contains all tautologies and modal principles (in that logic). Followed that, we have the following theorem.

**Theorem 65** *For* $\mathsf{X} \in \{\mathsf{S4}, \mathsf{T}, \mathsf{K4}\}$*, any axiom of* X *is also in* $\mathsf{X}^{\otimes}$.

**Proof.** Any propositional axiom $\chi$ is a tautology, and can be written as $\phi[P/\Psi]$ for some tautology (in propositional language) $\phi$, atoms $P$, and modal formulas $\Psi$. Then $\chi \in \mathsf{X}^{\otimes}$ by Corollary 62.

For modal axioms in T, $\Box(\phi \to \psi) \to (\Box\phi \to \Box\psi), \Box\phi \to \phi \in \mathsf{T}^{\otimes}$ follows from Lemma 63(C).

For modal axioms in K4, $\Box(\phi \to \psi) \to (\Box\phi \to \Box\psi), \Box\phi \to \Box\Box\phi \in \mathsf{K4}^{\otimes}$ follows from Lemma 63(C).

For modal axioms in S4, $\Box(\phi \to \psi) \to (\Box\phi \to \Box\psi), \Box\phi \to \phi, \Box\phi \to \Box\Box\phi \in \mathsf{T}^{\otimes} \cup \mathsf{K4}^{\otimes}$ follows from Lemma 63. Then by Theorem 64, these axioms are also in $\mathsf{S4}^{\otimes}$. ⊣

Recalled that in Corollary 47, there are modal theorems in each of T, K4, and S4 that are not included in corresponding loop-free provable fragments. The following theorem is now a consequence of what we know.

**Theorem 66 (Non-admissibility of $MP$)** *If* $\mathsf{X} \in \{\mathsf{S4}, \mathsf{T}, \mathsf{K4}\}$*, then* $\mathsf{X}^{\otimes}$ *is not closed under* $MP$*.*

**Proof.** By Theorem 65, $\mathsf{X}^{\otimes}$ contains all axioms in $\mathsf{X}$. By Theorem 51, $\mathsf{X}^{\otimes}$ is closed under Necessitation, hence Necessitation is admissible in $\mathsf{X}^{\otimes}$.

Suppose with the seek of a contradiction that $\mathsf{X}^{\otimes}$ is closed under $MP$, then $MP$ is also admissible in $\mathsf{X}^{\otimes}$, which then implies $\mathsf{X} \subseteq \mathsf{X}^{\otimes}$, which contradicts with Corollary 47. $\dashv$

**Corollary 67 (Non-normalities)** *If* $\mathsf{X} \in \{\mathsf{S4}, \mathsf{T}, \mathsf{K4}\}$*, then* $\mathsf{X}^{\otimes}$ *is not a normal modal logic.*

**Proof.** Directly by Theorem 66 and Definition 6. $\dashv$

In this thesis, prehistoric graph generally serves as a technical tool in the research of self-referentiality. However, as instanced by properties shown in this section, prehistoric graph is interesting by its own, and should be considered as a notion extending $\mathsf{G3}$-style calculi of modal logics. The framework of substructural logic (cf. [55]) studies sequent calculi with restricted structural rules. With the notion of prehistoric graph, the introduction, weakening, and contraction (we call it unification on the perspective of formulas, rather than sequents) of positive modalities are graphically presented. It will be an interesting direction to explore the relationship between prehistoric graph and substructural stuffs.

## 3.2 Non-self-referential Fragments

We start with the definition of the largest constant specification that is non-self-referential.

**Definition 68 (The Largest Non-self-referential Constant Specification)** Let $c$ (with possible subscripts) stand for a constant and $A$ stand for an axiom.

(1) The *largest constant specification that is not directly self-referential*, notation $\mathcal{CS}^{\not\circlearrowleft_1}$, is defined in J4 and LP as

$$\mathcal{CS}^{\not\circlearrowleft_1} := \{c : A \mid c \not\triangleright A\},$$

and in JT as

$$\mathcal{CS}^{\not\circlearrowleft_1} := \{!^n c : \cdots : c : A \mid c \not\triangleright A\}.$$

(2) The *largest constant specification that is not self-referential modulo renaming*, notation $\mathcal{CS}^{\not\circlearrowleft}$, is defined in J4 and LP as

$$\mathcal{CS}^{\not\circlearrowleft} := \{c_i : A \mid c_j \not\triangleright A \text{ for any } j \geq i\},$$

and in JT as

$$\mathcal{CS}^{\not\circlearrowleft} := \{!^n c_i : \cdots : c_i : A \mid c_j \not\triangleright A \text{ for any } j \geq i\}$$

(Note here that we employ the notational convention of [13] for JT, hence $c_1, c_2, ...$ here are initial elements of series of constants, and the second constant in the series of $c_1$ is written as $!c_1$ that contains an occurrence of $c_1$.) ⊣

Obviously, for a justification logic Y, if $\mathsf{Y}(\mathcal{CS}_0) \vdash \phi$ for a constant specification $\mathcal{CS}_0$ that is not directly self-referential, then $\mathsf{Y}(\mathcal{CS}^{\not\circlearrowleft_1}) \vdash \phi$. What less obvious is the case in (2), which we will explain after the coming definition.

**Definition 69 (Non-self-referential Fragments)** Let $\mathsf{X} \in \{\mathsf{T}, \mathsf{K4}, \mathsf{S4}\}$, and let $\mathsf{Y} \in \{\mathsf{JT}, \mathsf{J4}, \mathsf{LP}\}$ respectively.

(1) The *non-directly-self-referential fragment* of $\mathsf{X}$, notation $\mathsf{X}^{\not\circlearrowleft_1}$, is defined by

$$\mathsf{X}^{\not\circlearrowleft_1} := \{\mathsf{X} \vdash \phi \,|\, \text{there is a realizer } r \text{ s.r. } \mathsf{Y}(\mathcal{CS}^{\not\circlearrowleft_1}) \vdash \phi^r\}.$$

(2) The *non-self-referential fragment* of $\mathsf{X}$, notation $\mathsf{X}^{\not\circlearrowleft}$, is defined by

$$\mathsf{X}^{\not\circlearrowleft} := \{\mathsf{X} \vdash \phi \,|\, \text{there is a realizer } r \text{ s.r. } \mathsf{Y}(\mathcal{CS}^{\not\circlearrowleft}) \vdash \phi^r\}.$$

$\dashv$

Observe that, if a modal theorem has a realization that calls for a constant specification that is not self-referential, that constant specification cannot have any subset of the form

$$\{c_1 : A_1(c_2), c_2 : A_2(c_3), \cdots, c_n : A_n(c_1)\}.$$

Then, there must be a way of renaming constants so that no constant occur in an axiom prefixed by a constant with equal or larger subscript. After this renaming of constants, the resulting constant specification will be a subset of $\mathcal{CS}^{\not\circlearrowleft}$ defined in Definition 68(2). In Definition 69(2), in order to be included in the non-self-referential fragment, a nodal theorem can take any realizer to be realized. The free choice of realizer automatically covers the renaming of constants necessary for Definition 68(2) to make sense.

The following theorem is obvious.

**Theorem 70** *If* $\mathsf{X} \in \{\mathsf{T}, \mathsf{K4}, \mathsf{S4}\}$, *then* $\mathsf{X}^{\not\circlearrowleft} \subseteq \mathsf{X}^{\not\circlearrowleft_1} \subsetneq \mathsf{X}$.

**Proof.** Since $\mathcal{CS}^{\not\circlearrowleft} \subseteq \mathcal{CS}^{\not\circlearrowleft_1}$ holds for each justification logic, by definition, we see that $\mathsf{X}^{\not\circlearrowleft} \subseteq \mathsf{X}^{\not\circlearrowleft_1}$.

To see that $\mathsf{X}^{\not\circlearrowleft_1} \subsetneq \mathsf{X}$, first observe that $\mathsf{X}^{\not\circlearrowleft_1} \subseteq \mathsf{X}$ holds by definition. By Kuznets' results in Table 2.1, we see that $\mathsf{X}^{\not\circlearrowleft_1} \not\supseteq \mathsf{X}$, which implies that non-equation part. $\dashv$

In the terminology defined in this section, Theorem 42 can be reformulated as follows.

**Corollary 71** *If* $\mathsf{X} \in \{\mathsf{T}, \mathsf{K4}, \mathsf{S4}\}$*, then* $\mathsf{X}^{\otimes} \subseteq \mathsf{X}^{\not\circlearrowleft}$.

**Proof.** By Theorem 42. $\dashv$

With this corollary, we have the following:

**Theorem 72** *For* $\mathsf{X} \in \{\mathsf{S4}, \mathsf{T}, \mathsf{K4}\}$*, any axiom of* $\mathsf{X}$ *is also in* $\mathsf{X}^{\not\circlearrowleft}$ *and* $\mathsf{X}^{\not\circlearrowleft_1}$.

**Proof.** By Theorem 65, all instances of axiom schemes of $\mathsf{X}$ are included in $\mathsf{X}^{\otimes}$. Since $\mathsf{X}^{\otimes} \subseteq \mathsf{X}^{\not\circlearrowleft}$ by Corollary 71, they are also included in $\mathsf{X}^{\not\circlearrowleft}$.

Then by Theorem 70, $\mathsf{X}^{\not\circlearrowleft} \subseteq \mathsf{X}^{\not\circlearrowleft_1}$, and hence all instances of axiom schemes of $\mathsf{X}$ are included in $\mathsf{X}^{\not\circlearrowleft_1}$ $\dashv$

The closure of non-self-referential fragments under Necessitation is actually already shown as Corollary 41. We reformulate it in the current terminology.

**Corollary 73 (Closure under Necessitation)** *If* $\mathsf{X} \in \{\mathsf{T}, \mathsf{K4}, \mathsf{S4}\}$*, then both* $\mathsf{X}^{\not\circlearrowleft_1}$ *and* $\mathsf{X}^{\not\circlearrowleft}$ *are closed under the rule of Necessitation.*

**Proof.** By Corollary 41 and Definition 69. $\dashv$

Having followed a similar approach as in Section 3.1, we are now ready to show that non-self-referential fragments fail to close under $MP$.

**Theorem 74 (Non-admissibility of** $MP$**)** *If* $\mathsf{X} \in \{\mathsf{T}, \mathsf{K4}, \mathsf{S4}\}$*, then neither* $\mathsf{X}^{\not\circlearrowleft_1}$ *nor* $\mathsf{X}^{\not\circlearrowleft}$ *is closed under* $MP$.

**Proof.** We verify for $\mathsf{X}^{\not\circlearrowleft}$, and the case for $\mathsf{X}^{\not\circlearrowleft_1}$ is quite similar.

By Theorem 72, $\mathsf{X}^{\not\circlearrowleft}$ contains all axioms of $\mathsf{X}$. By Corollary 73, $\mathsf{X}^{\not\circlearrowleft}$ is closed under Necessitation, hence Necessitation is admissible in $\mathsf{X}^{\not\circlearrowleft}$.

Suppose with the seek of a contradiction that $\mathsf{X}^{\not\circlearrowleft}$ is closed under $MP$, then $MP$ is also admissible in $\mathsf{X}^{\not\circlearrowleft}$, which implies that $\mathsf{X} \subseteq \mathsf{X}^{\not\circlearrowleft}$, which contradicts with Theorem 70. $\dashv$

We do not give an example pair of modal theorems that can be realized non(-direct)-self-referentially, while the result of applying $MP$ on them cannot. Finding such an example is not hard, since directly self-referential modal theorems given by Kuznets and presented in Table 2.1 have Hilbert-style modal proofs with short lengths. Among applications of $MP$ in the proof for each theorems, there must be a desired one.

**Corollary 75 (Non-normalities)** *If* $\mathsf{X} \in \{\mathsf{T}, \mathsf{K4}, \mathsf{S4}\}$*, then neither* $\mathsf{X}^{\not\circlearrowleft_1}$ *nor* $\mathsf{X}^{\not\circlearrowleft}$ *is a normal modal logic.*

**Proof.** By Theorem 74 and Definition 6. $\dashv$

In a Kripke model for modal logic, an implication is satisfied on a state if and only if either the antecedent is not satisfied, or the succedent is satisfied. That is to say, satisfiability on a state in a Kripke model is closed under $MP$. The failure of closure under $MP$ of non-self-referential fragment indicates that on the modal layer, self-referentiality cannot be characterized in Kripke semantics.

## 3.3 Between Fragments of Variant Modal Logics

Both Section 3.1 and Section 3.2 present results for fragmentS of a specified modal logic. This section takes care of relations between the same kind of fragments of different modal logics. We start by the following theorem.

**Theorem 76** *(A)* $\mathsf{K4}^{\not\circlearrowleft_1} \subseteq \mathsf{S4}^{\not\circlearrowleft_1}$*;*

*(B)* $\mathsf{T}^{\not\circlearrowleft_1} \subseteq \mathsf{S4}^{\not\circlearrowleft_1}$.

**Proof.** (A) For any $\phi \in \mathsf{K4}^{\not\circlearrowleft_1}$, there must be a realizer $r$ s.t. $\mathsf{J4}(\mathcal{CS}^{\not\circlearrowleft_1}) \vdash \phi^r$.

First observe that each J4-axiom is also an LP-axiom. Any $AN.1$ rule application in the proof has the form of $c:A$ where $c \rhd A$. The J4-axiom $A$ is also an LP-axiom, and hence $c:A$ can also be introduced by the $AN.1$ rule in LP. Besides, $c:A$ remains as a non-self-referential formula as in J4. Finally, both justification logic share the rule $MP$.

In summary, the proof of $\phi^r$ in $\mathsf{J4}(\mathcal{CS}^{\not\circlearrowleft_1})$ is also a proof in $\mathsf{LP}(\mathcal{CS}^{\not\circlearrowleft_1})$. Therefore, $\phi \in \mathsf{S4}^{\not\circlearrowleft_1}$.

(B) For any $\phi \in \mathsf{T}^{\not\circlearrowleft_1}$, there must be a realizer $r$ s.t. $\mathsf{JT}(\mathcal{CS}^{\not\circlearrowleft_1}) \vdash \phi^r$. The cases for axiom and $MP$ are similar to that in (A), whereas the case for $AN$ is harder, since JT enjoys $AN.2$ instead of $AN.1$.

For each $AN.2$ application

$$AN.2 \qquad \vdash !^n c : !^{n-1} c : \cdots : c : A,$$

substitute this line by

$$\begin{array}{ll}
AN.1 & \vdash c:A \\
A3 & \vdash c:A \to !c:c:A \\
MP & \vdash !c:c:A \\
 & \vdots \\
A3 & \vdash !^{n-1}c:\cdots:c:A \to !^n c:!^{n-2}c:\cdots:c:A \\
MP & \vdash !^n c:\cdots:c:A.
\end{array}$$

The resulting proof is then a proof in $\mathsf{LP}(\mathcal{CS}^{\circlearrowleft_1})$, which implies that $\phi \in \mathsf{S4}^{\circlearrowleft_1}$. ⊣

This theorem says, in the power of avoiding directly self-referentiality, $\mathsf{S4}$ is not weaker than either $\mathsf{T}$ or $\mathsf{K4}$. Recall that we have a similar result on the power of avoiding the necessity of a prehistoric-loop in Section 3.1.

The other direction is interesting only if the formula is indeed a theorem of the smaller logic, i.e., whether or not $\mathsf{S4}^{\circlearrowleft_1} \cap \mathsf{T} \subseteq \mathsf{T}^{\circlearrowleft_1}$, and whether or not $\mathsf{S4}^{\circlearrowleft_1} \cap \mathsf{K4} \subseteq \mathsf{K4}^{\circlearrowleft_1}$. This is still open for the $\mathsf{T}$ case, but is shown to fail for the $\mathsf{K4}$ case by what follows. Generally speaking, we find a theorem

$$\Diamond\Box p \to \Diamond\Box\Diamond p,$$

shared by all of $\mathsf{T}$, $\mathsf{K4}$, and $\mathsf{S4}$, whereas it is further included in both $\mathsf{T}^{\circlearrowleft_1}$ and $\mathsf{S4}^{\circlearrowleft_1}$, but not in $\mathsf{K4}^{\circlearrowleft_1}$.

We first show the following:

**Theorem 77** *(A)* $\mathsf{K4} \vdash \Diamond\Box p \to \Diamond\Box\Diamond p$;

*(B1)* $\Diamond\Box p \to \Diamond\Box\Diamond p \in \mathsf{T}^{\otimes}$,

*(B2)* $\Diamond\Box p \to \Diamond\Box\Diamond p \in \mathsf{T}^{\circlearrowleft}$,

*(B3)* $\Diamond\Box p \to \Diamond\Box\Diamond p \in \mathsf{T}^{\circlearrowleft_1}$;

*(C1)* $\Diamond\Box p \to \Diamond\Box\Diamond p \in \mathsf{S4}^{\otimes}$,

*(C2)* $\Diamond\Box p \to \Diamond\Box\Diamond p \in \mathsf{S4}^{\circlearrowleft}$,

*(C3)* $\Diamond\Box p \to \Diamond\Box\Diamond p \in \mathsf{S4}^{\circlearrowleft_1}$.

**Proof.** Since $\Diamond$ is defined as an abbreviation of $\neg\Box\neg$, we have (after adding annotations to positive $\Box$'s for convenience):

$$\Diamond\Box p \to \Diamond\Box\Diamond p \equiv \neg\Box_a\neg\Box p \to \neg\Box\neg\Box_b\neg\Box\neg p.$$

(A) We present a proof of

$$\Rightarrow \neg\Box_a\neg\Box p \to \neg\Box\neg\Box_b\neg\Box\neg p$$

in G34.[3]

$$\begin{array}{rcll}
p & \Rightarrow & p & (Ax)\\
p, \neg p & \Rightarrow & & (L\neg)\\
p, \Box p, \neg p, \Box\neg p & \Rightarrow & \neg\Box\neg p & (W)\\
\Box p, \Box\neg p & \Rightarrow & \Box_b\neg\Box\neg p & (4\Box)_{b.1}\\
\Box p & \Rightarrow & \Box_b\neg\Box\neg p, \neg\Box\neg p & (R\neg)\\
\Box p, \neg\Box_b\neg\Box\neg p & \Rightarrow & \neg\Box\neg p & (L\neg)\\
p, \Box p, \neg\Box_b\neg\Box\neg p, \Box\neg\Box_b\neg\Box\neg p & \Rightarrow & \neg\Box\neg p & (W)\\
\Box p, \Box\neg\Box_b\neg\Box\neg p & \Rightarrow & \Box_b\neg\Box\neg p & (4\Box)_{b.2}\\
\Box p, \neg\Box_b\neg\Box\neg p, \Box\neg\Box_b\neg\Box\neg p & \Rightarrow & & (L\neg)\\
\neg\Box_b\neg\Box\neg p, \Box\neg\Box_b\neg\Box\neg p & \Rightarrow & \neg\Box p & (R\neg)\\
\Box\neg\Box_b\neg\Box\neg p & \Rightarrow & \Box_a\neg\Box p & (4\Box)_a\\
 & \Rightarrow & \Box_a\neg\Box p, \neg\Box\neg\Box_b\neg\Box\neg p & (R\neg)\\
\neg\Box_a\neg\Box p & \Rightarrow & \neg\Box\neg\Box_b\neg\Box\neg p & (L\neg)\\
 & \Rightarrow & \neg\Box_a\neg\Box p \to \neg\Box\neg\Box_b\neg\Box\neg p & (R\to)
\end{array}$$

Note that this proof has a prehistoric loop $b \prec b$ of length 1, introduced by the only $(4\Box)_a$.

As we will see in Corollary 82, this sequent is not loop-free provable in G34.

(B1) It is sufficient to give a proof of

$$\Rightarrow \neg\Box_a\neg\Box p \to \neg\Box\neg\Box_b\neg\Box\neg p$$

in G3t$^{\otimes}$.

A desired proof is as follows:

$$\begin{array}{rcll}
p & \Rightarrow & p & (Ax)\\
p, \neg p & \Rightarrow & & (L\neg)\\
p, \Box\neg p & \Rightarrow & & (L\Box W)\\
p & \Rightarrow & \neg\Box\neg p & (R\neg)\\
\Box p & \Rightarrow & \Box_b\neg\Box\neg p & (K\Box)\\
 & \Rightarrow & \neg\Box p, \Box_b\neg\Box\neg p & (R\neg)\\
\neg\Box_b\neg\Box\neg p & \Rightarrow & \neg\Box p & (L\neg)\\
\Box\neg\Box_b\neg\Box\neg p & \Rightarrow & \Box_a\neg\Box p & (K\Box)\\
\neg\Box_a\neg\Box p, \Box\neg\Box_b\neg\Box\neg p & \Rightarrow & & (L\neg)\\
\neg\Box_a\neg\Box p & \Rightarrow & \neg\Box\neg\Box_b\neg\Box\neg p & (R\neg)\\
 & \Rightarrow & \neg\Box_a\neg\Box p \to \neg\Box\neg\Box_b\neg\Box\neg p & (R\to)
\end{array}$$

[3] We omit lines between sequents here.

whose prehistoric graph has two vertices $a, b$, and only one edge $b \prec a$ introduced by the second $(K\Box)$.

(B2) By (B1) and Corollary 71.

(B3) By (B2) and Theorem 70.

(C1) By (B1) and Theorem 64(C).

(C2) By (B1) and Corollary 71.

(C3) By (B3) and Theorem 70. ⊣

Then we verify the fact that $\Diamond\Box p \to \Diamond\Box\Diamond p \notin \mathsf{K4}^{\wp_1}$, using a construction of counter $\mathcal{CS}^{\wp_1}$-Fitting model in $\mathsf{J4}$.

**Theorem 78** $\Diamond\Box p \to \Diamond\Box\Diamond p \notin \mathsf{K4}^{\wp_1}$.

**Proof.** Any possible realization of $\neg\Box_a\neg\Box p \to \neg\Box\neg\Box_a\neg\Box\neg p$ has the form of

$$\neg t_a : \neg t_1 : p \to \neg t_2 : \neg t_b : \neg t_3 : \neg p.$$

Recall that in Definition 21 and Lemma 22, we introduced the notation $*_X$ to denote the closure of $X$ under conditions there, and showed that $*_X$ is the smallest evidence function generated by initial set $X$. Let

$$\begin{aligned}
*^0 &:= *_{\mathcal{CS}^{\varphi_1}}\ ,\\
*^1 &:= *_{\mathcal{CS}^{\varphi_1}\cup\{t_1:p\}}\ ,\\
*^2 &:= *_{\mathcal{CS}^{\varphi_1}\cup\{t_1:p,t_2:\neg t_b:\neg t_3:\neg p\}}\ ,\\
*^3 &:= *_{\mathcal{CS}^{\varphi_1}\cup\{t_2:\neg t_b:\neg t_3:\neg p\}}\ ,\\
*^{i\setminus j} &:= *^i \setminus *^j \quad \text{(this needs not to be an evidence function itself);}\\
T_0(p) &:= \{p\},\\
T_{n+1}(p) &:= \{t:\phi \mid t \text{ is a term and } \phi \in T_n(p)\},\\
T(p) &:= \bigcup_{i\in\omega} T_i(p).
\end{aligned}$$

Let $\mathfrak{M} := (W, R, E, V)$ be a J4 Fitting model where $W = \{u, v\}$, $R = \{(u, v)\}$, $E(u) = *^u = *^3$, $E(v) = *^v = *^2$, $V$ is arbitrary. In what follows, we will show that

$$\mathfrak{M}, u \nVdash \neg t_a : \neg t_1 : p \to \neg t_2 : \neg t_b : \neg t_3 : \neg p\ .$$

**Lemma 79** *In a Fitting model, if formula $t : \phi$ is true at one state, then it is also true at any accessible state.*

**Proof.** Assume $w_1 \Vdash t : \phi$ and $w_1Rw_2$, it is sufficient to show that $w_2 \Vdash t : \phi$. Take any $w_3$ accessible from $w_2$, by transitivity, $w_1Rw_3$ and hence $w_3 \Vdash \phi$. Note that $\phi \in *^{w_1}(t)$, by monotonicity, we see that $\phi \in *^{w_2}(t)$. ⊣

**Claim 80** *For any $s \rhd t_b$:*

*(A) If $\phi \in *^0(s)$, then*

*(i) $\mathsf{J4}(\mathcal{CS}^{\varphi_1}) \vdash \phi$,*

*(ii) $t_b \ntriangleright \phi$;*

*(B) If $\phi \in *^{1\backslash 0}(s)$, then*

*(iii) For any $\mathcal{CS}^{\varphi_1}$-model $\mathfrak{M}' := (W', R', E', V')$ with $u' \in W'$, if $V'(p) = W'$ and $\mathfrak{M}', u' \Vdash T(p)$, then $\mathfrak{M}', u' \Vdash \phi$,*

*(iv) $t_b \not\rhd \phi$;*

*(C) If $\phi \in *^{2\backslash 1}(s)$, then*

*(v) $t_b \rhd \phi$,*

*(vi) If $\phi$ is an implication, then $\phi \equiv \neg t_b : \neg t_3 : \neg p$,*

*(vii) $\phi \not\equiv t_b : \neg t_3 : \neg p$.*

**Proof.** First observe that (iii), (iv) are satisfied if $\phi \equiv p$, and (v), (vi), (vii) are satisfied if $\phi \equiv \neg t_b : \neg t_3 : \neg p$.

We prove the lemma by an induction on $s$.

(1) $s \equiv c$ for a constant $c$:

(1A) Assume $\phi \in *^0(c)$. (1A1) Via initial set. In this case $c : \phi \in \mathcal{CS}^{\varphi_1}$, so $\phi$ is an axiom and (i) holds. Suppose that $t_b \rhd \phi$, then $c \equiv s \rhd t_b \rhd \phi$, which makes $\mathcal{CS}^{\varphi_1}$ directly self-referential, a contradiction. (1A2) Via closure. Impossible since $s \equiv c$ is a constant.

(1B) Assume $\phi \in *^{1\backslash 0}(c)$. (1B1) Via initial set. In this case $\phi \equiv p$, hence both (iii) and (iv) hold. (1B2) Via closure. Impossible.

(1C) Assume $\phi \in *^{2\backslash 1}(c)$. (1C1) Via initial set. In this case $\phi \equiv \neg t_b : \neg t_3 : \neg p$, hence all of (v), (vi), and (vii) hold. (1C2) Via closure. Impossible.

(2) $s \equiv x$ for a variable $x$:

(2A) Assume $\phi \in *^0(x)$. (2A1) Via initial set. In this case $x : \phi \in \mathcal{CS}^{\varphi_1}$, which is impossible. (2A2) Via closure. Impossible since $s \equiv x$ is a variable.

(2B) Assume $\phi \in *^{1\backslash 0}(x)$. (2B1) Via initial set. In this case $\phi \equiv p$. (2B2) Via closure. Impossible.

(2C) Assume $\phi \in *^{2\setminus 1}(c)$. (2C1) Via initial set. In this case $\phi \equiv \neg t_b : \neg t_3 : \neg p$. (2C2) Via closure. Impossible.

(3) $s \equiv !s_1$ for a term $s_1$:

(3A) Assume $\phi \in *^0(!s_1)$. (3A1) Via initial set. In this case $!s_1 : \phi \in \mathcal{CS}^{\varphi_1}$, which is impossible. (3A2) Via closure. So $\phi \equiv s_1 : \psi$ for some formula $\psi \in *^0(s_1)$. By IH(i), $\mathsf{J4}(\mathcal{CS}^{\varphi_1}) \vdash \psi$. Take an arbitrary $\mathcal{CS}^{\varphi_1}$-model $\mathfrak{M}'$ with a state $u'$ in it, let $*^{u'} = E'(u')$ be the evidence function on $u'$. Since $*^{u'}$ admits $\mathcal{CS}^{\varphi_1}$, and $*^0$ is the smallest $\mathcal{CS}^{\varphi_1}$-evidence function, from $\psi \in *^0(s_1)$ we see that $\psi \in *^{u'}(s_1)$. For any state $v'$ in $\mathfrak{M}'$ accessible from $u'$, by soundness, $\mathfrak{M}', v' \Vdash \psi$. So we have $\mathfrak{M}', u' \Vdash s_1 : \psi$, and then $\mathsf{J4}(\mathcal{CS}^{\varphi_1}) \vdash s_1 : \psi$ by completeness, (i) holds. By IH(ii), $t_b \not\trianglerighteq \psi$. Since $s_1$ is a proper subterm of $s$ and hence of $t_b$, we see that $t_b \not\trianglerighteq s_1$. Therefore, $t_b \not\trianglerighteq s_1 : \psi$, (ii) holds.

(3B) Assume $\phi \in *^{1\setminus 0}(!s_1)$. (3B1) Via initial set. In this case $\phi \equiv p$. (3B2) Via closure. So $\phi \equiv s_1 : \psi$ for some formula $\psi \in *^{1\setminus 0}(s_1) \subseteq *^1(s_1)$. Take an arbitrary $\mathcal{CS}^{\varphi_1}$-model $\mathfrak{M}' = (W', R', E', V')$ with $u' \in W'$, assume $V'(p) = W'$ and $\mathfrak{M}', u' \Vdash T(p)$. Since $t_1 : p \in T(p)$, we know $\mathfrak{M}', u' \Vdash t_1 : p$, and hence $p \in *^{u'}(t_1)$, where $*^{u'} = E'(u')$ is the evidence function on state $u'$. Note that $\mathfrak{M}'$ is a $\mathcal{CS}^{\varphi_1}$-model, which implies that $*^{u'}$ admits $\mathcal{CS}^{\varphi_1}$. Now we see $*^1(s_1) \subseteq *^{u'}(s_1)$ from the fact that $*^1$ is the smallest $\mathcal{CS}^{\varphi_1} \cup \{t_1 : p\}$-evidence function. So $\psi \in *^{u'}(s_1)$ follows from $\psi \in *^1(s_1)$. Now take any state $v'$ in $\mathfrak{M}'$ s.t. $u'R'v'$. From $V'(p) = W'$ we see $\mathfrak{M}', v' \Vdash p$, i.e., $\mathfrak{M}', v' \Vdash T_0(p)$. For any $i > 0$, formulas in $T_i(p)$ all have the form of $t : \alpha$. By Lemma 79, $\mathfrak{M}', u' \Vdash T_i(p)$ implies $\mathfrak{M}', v' \Vdash T_i(p)$. Since $\mathfrak{M}', u' \Vdash T(p)$, we see that $\mathfrak{M}', v' \Vdash T(p)$. By IH(iii), $\mathfrak{M}', v' \Vdash \psi$. This, together with earlier proved fact $\psi \in *^{u'}(s_1)$ gives $\mathfrak{M}', u' \Vdash s_1 : \psi$, (iii) holds. By IH(iv), $t_b \not\trianglerighteq \psi$. Since $s \trianglerighteq t_b$ and $s_1$ is a proper subterm of $s$, $t_b \not\trianglerighteq s_1$. Therefore, $t_b \not\trianglerighteq s_1 : \psi$, (iv) holds.

(3C) Assume $\phi \in *^{2\setminus 1}(!s_1)$. (3C1) Via initial set. In this case $\phi \equiv \neg t_b : \neg t_3 : \neg p$. (3C2)

Via closure. So $\phi \equiv s_1 : \psi$ for some formula $\psi \in *^{2\backslash 1}(s_1)$. By IH(v), $t_b \rhd \psi$, which implies $t_b \rhd s_1 : \psi$, (v) holds. As $s_1 : \psi$ is not an implication, (vi) holds. Since $s \rhd t_b$ and $s_1$ is a proper subterm of $s$, we see that $s_1 \not\equiv t_b$. Therefore, $s_1 : \psi \not\equiv t_b : \neg t_3 : \neg p$, (vii) holds.

(4) $s \equiv s_1 \cdot s_2$ for some terms $s_1$ and $s_2$:

(4A) Assume $\phi \in *^0(s_1 \cdot s_2)$. (4A1) Via initial set. In this case $s_1 \cdot s_2 : \phi \in \mathcal{CS}^{\varphi_1}$, which is impossible. (4A2) Via closure. So there is a formula $\tau$ s.r. $\tau \to \phi \in *^0(s_1)$ and $\tau \in *^0(s_2)$. By IH(i), $\mathsf{J4}(\mathcal{CS}^{\varphi_1}) \vdash \tau \to \phi$ and $\mathsf{J4}(\mathcal{CS}^{\varphi_1}) \vdash \tau$, which implies $\mathsf{J4}(\mathcal{CS}^{\varphi_1}) \vdash \phi$, (i) holds. By IH(ii), $t_b \not\rhd \tau \to \phi$, which implies $t_b \not\rhd \phi$, (ii) holds.

(4B) Assume $\phi \in *^{1\backslash 0}(s_1 \cdot s_2)$. (4B1) Via initial set. In this case $\phi \equiv p$. (4B2) Via closure. So there is a formula $\tau$ s.r. $\tau \to \phi \in *^1(s_1)$ and $\tau \in *^1(s_2)$. (4B21) $\tau \to \phi \in *^0(s_1)$ and $\tau \in *^0(s_2)$. Then $\phi \in *^0(s_1 \cdot s_2)$, a contradiction. (4B22) $\tau \to \phi \in *^0(s_1)$ and $\tau \in *^{1\backslash 0}(s_2)$. Take an arbitrary $\mathcal{CS}^{\varphi_1}$-model $\mathfrak{M}' = (W', R', E', V')$ with $u' \in W'$, assume $V'(p) = W'$ and $\mathfrak{M}', u' \Vdash T(p)$. By IH(iii), $\mathfrak{M}', u' \Vdash \tau$. By IH(i) and soundness, $\mathfrak{M}', u' \Vdash \tau \to \phi$. So $\mathfrak{M}', u' \Vdash \phi$, (iii) holds. By IH(iv), $t_b \not\rhd \tau \to \phi$, which implies $t_b \not\rhd \phi$, (iv) holds. (4B23) $\tau \to \phi \in *^{1\backslash 0}(s_1)$ and $\tau \in *^0(s_2)$. Similar as (4B22), using IH(iii,iv) on $s_1$ and (i) on $s_2$. (4B24) $\tau \to \phi \in *^{1\backslash 0}(s_1)$ and $\tau \in *^{1\backslash 0}(s_2)$. Similar as (4B22), using IH(iii) on both $s_1$ and $s_2$, and (iv) on $s_1$.

(4C) Assume $\phi \in *^{2\backslash 1}(s_1 \cdot s_2)$. (4C1) Via initial set. In this case $\phi \equiv \neg t_b : \neg t_3 : \neg p$. (4C2) Via closure. So there is a formula $\tau$ s.r. $\tau \to \phi \in *^2(s_1)$ and $\tau \in *^2(s_2)$. (4C21) $\tau \to \phi \in *^1(s_1)$ and $\tau \in *^1(s_2)$. Then $\phi \in *^1(s_1 \cdot s_2)$, a contradiction. (4C22) $\tau \to \phi \in *^1(s_1)$ and $\tau \in *^{2\backslash 1}(s_2)$. If $\tau \to \phi \in *^0(s_1)$, then by IH(ii), $t_b \not\rhd \tau \to \phi$. Otherwise $\tau \to \phi \in *^{1\backslash 0}(s_1)$, then by IH(iv), $t_b \not\rhd \tau \to \phi$. In summary, we have $t_b \not\rhd \tau \to \phi$ in both cases. Now IH(v) offers $t_b \rhd \tau$, contradicts with what we have. (4C23) $\tau \to \phi \in *^{2\backslash 1}(s_1)$ and $\tau \in *^1(s_2)$. As an implication, by IH(vi), $\tau \to \phi \equiv \neg t_b : \neg t_3 : \neg p$, which implies $\tau \equiv t_b : \neg t_3 : \neg p$. Consider the model $\mathfrak{M}' = (W', R', E', V')$, where $W' = \{u', v'\}$, $R' = \{(u', v')\}$, $E'(u') = *^{u'} = E'(v') = *^{v'}$ being the full evidence function,[4] and $V'(p) = W'$. Note that $\mathfrak{M}'$ is a $\mathcal{CS}^{\varphi_1}$-model as the full evidence function admits $\mathcal{CS}^{\varphi_1}$. By an east observation we see that $\mathfrak{M}', u' \Vdash T(p)$. Since $\neg p \in *^{v'}(t_3)$ and $v'$ has no accessible state, $\mathfrak{M}', v' \Vdash t_3 : \neg p$. Thus $\mathfrak{M}', v' \nVdash \neg t_3 : \neg p$, and by $u'R'v'$, we have $\mathfrak{M}', u' \nVdash t_b : \neg t_3 : \neg p$, i.e., $\mathfrak{M}', u' \nVdash \tau$. If $\tau \in *^0(s_2)$, then by IH(i) and soundness, $\mathfrak{M}', u' \Vdash \tau$, a contradiction. Otherwise $\tau \in *^{1\backslash 0}(s_2)$. Since $V'(p) = W'$ and $\mathfrak{M}', u' \Vdash T(p)$, by IH(iii), we have $\mathfrak{M}', u' \Vdash \tau$, a contradiction. In summary, both cases lead to contradictions. (4C24) $\tau \to \phi \in *^{2\backslash 1}(s_1)$ and $\tau \in *^{2\backslash 1}(s_2)$. By IH(vi), as an implication, $\tau \to \phi \equiv \neg t_b : \neg t_3 : \neg p$, and hence $\tau \equiv t_b : \neg t_3 : \neg p$. This contradicts with $\tau \not\equiv t_b : \neg t_3 : \neg p$, concluded by IH(vii).

(5) $s \equiv s_1 + s_2$ for some terms $s_1$ and $s_2$:

(5A) Assume $\phi \in *^0(s_1 + s_2)$. (5A1) Via initial set. In this case $s_1 + s_2 : \phi \in \mathcal{CS}^{\varphi_1}$, which is impossible. (5A2) Via closure. W.l.o.g., $\phi \in *^0(s_1)$, and all desired follow from IH.

(5B) Assume $\phi \in *^{1\backslash 0}(s_1 + s_2)$. (5B1) Via initial set. In this case $\phi \equiv p$. (5B2) Via closure. In this case $\phi \notin *^0(s_1) \cup *^0(s_2)$, and w.l.o.g., $\phi \in *^1(s_1)$. So $\phi \in *^{1\backslash 0}(s_1)$, and all desired follow from IH.

(5C) Assume $\phi \in *^{2\backslash 1}(s_1 + s_2)$. (5C1) Via initial set. In this case $\phi \equiv \neg t_b : \neg t_3 : \neg p$. (5C2) Via closure. In this case $\phi \notin *^1(s_1) \cup *^1(s_2)$, and w.l.o.g., $\phi \in *^2(s_1)$. So $\phi \in *^{2\backslash 1}(s_1)$, and all desired follow from IH. ⊣

**Claim 81** $\neg t_3 : \neg p \notin *^v(t_b)$.

**Proof.** Consider the model $\mathfrak{M}' = (W', R', E', V')$, where $W' = \{u'\}$, $R' = \varnothing$, $E'(u') = *^{u'}$ being the full evidence function, and $V'(p) = W'$. Note that $\mathfrak{M}'$ is a $\mathcal{CS}^{\varphi_1}$-model as the full evidence function admits $\mathcal{CS}^{\varphi_1}$. It is easy to see that $\mathfrak{M}', u' \Vdash T(p)$ and $\mathfrak{M}', u' \nVdash \neg t_3 : \neg p$.

[4] Recall that an evidence function $*$ is full, if $\phi \in *(t)$ for any term $t$ and formula $\phi$.

Assume with the seek of a contradiction that $\neg t_3 : \neg p \in *^v(t_b)$. Since $*^v = *^2$, we have $\neg t_3 : \neg p \in *^2(t_b) = *^0(t_b) \cup *^{1\backslash 0}(t_b) \cup *^{2\backslash 1}(t_b)$.

Case (1), $\neg t_3 : \neg p \in *^0(t_b)$. By Claim 80(Ai) and soundness, $\mathfrak{M}', u' \Vdash \neg t_3 : \neg p$, a contradiction.

Case (2), $\neg t_3 : \neg p \in *^{1\backslash 0}(t_b)$. By Claim 80(Biii), $\mathfrak{M}', u' \Vdash \neg t_3 : \neg p$, a contradiction.

Case (3), $\neg t_3 : \neg p \in *^{2\backslash 1}(t_b)$. By Claim 80(Cvi), as an implication, $\neg t_3 : \neg p \equiv \neg t_b : \neg t_3 : \neg p$, a contradiction.

In summary, none case is possible, and our assumption cannot be true. ⊣

Now we are ready to show that $\mathfrak{M}, u \nVdash \neg t_a : \neg t_1 : p \to \neg t_2 : \neg t_b : \neg t_3 : \neg p$.

By Claim 81, $\neg t_3 : \neg p \notin *^v(t_b)$, which implies $\mathfrak{M}, v \nVdash t_b : \neg t_3 : \neg p$. By our definition of $*^u$, we have $\neg t_b : \neg t_3 : \neg p \in *^u(t_2)$. Since $v$ is the only state accessible from $u$ and $\mathfrak{M}, v \Vdash \neg t_b : \neg t_3 : \neg p$, we have $\mathfrak{M}, u \Vdash t_2 : \neg t_b : \neg t_3 : \neg p$, and hence $\mathfrak{M}, u \nVdash \neg t_2 : \neg t_b : \neg t_3 : \neg p$.

It is therefore sufficient to show that $\mathfrak{M}, u \Vdash \neg t_a : \neg t_1 : p$. By our definition of $*^v$, we have $p \in *^v(t_1)$. Since $v$ has no accessible state, $\mathfrak{M}, v \Vdash t_1 : p$. So $\mathfrak{M}, v \nVdash \neg t_1 : p$, and since $uRv$, we have $\mathfrak{M}, u \nVdash t_a : \neg t_1 : p$, which implies the desired $\mathfrak{M}, u \Vdash \neg t_a : \neg t_1 : p$.

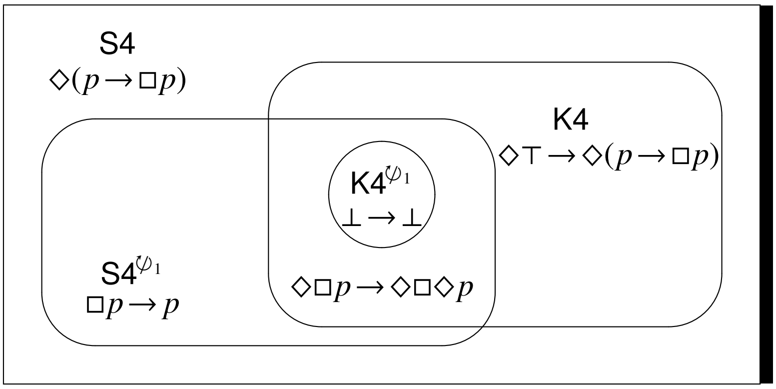
Have verified $\mathfrak{M}, u \nVdash \neg t_a : \neg t_1 : p \to \neg t_2 : \neg t_b : \neg t_3 : \neg p$, we can draw the conclusion that $\Diamond\Box p \to \Diamond\Box\Diamond p$ cannot be realized in $\mathsf{J4}$ unless a directly self-referential constant specification is called, since otherwise, the realized formula would be a $\mathsf{J4}(\mathcal{CS}^{\varphi_1})$-theorem, and hence would have no counter $\mathcal{CS}^{\varphi_1}$-models by Theorem 24. ⊣

As a quick corollary, we have:

**Corollary 82** $\Diamond\Box p \to \Diamond\Box\Diamond p \notin \mathsf{K4}^{\otimes}$.

**Proof.** Following Theorem 78, by Theorem 70, $\Diamond\Box p \to \Diamond\Box\Diamond p \notin \mathsf{K4}^{\varphi}$. Now Corollary 71 offers what desired. ⊣

Figure 3.1: Direct Self-referentiality in $\mathsf{K4}$ and $\mathsf{S4}$.



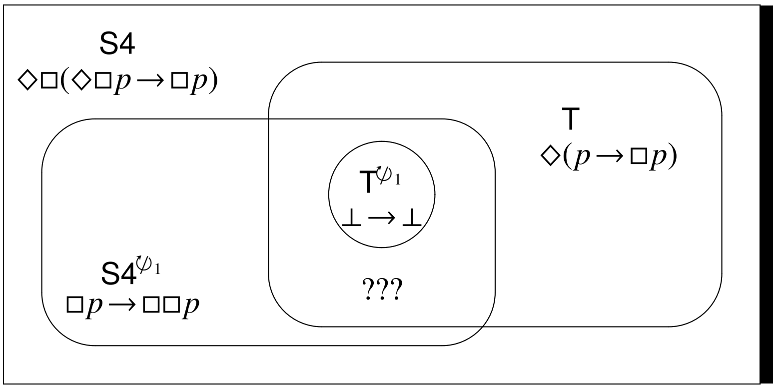

Combined with Theorem 77, we see that $\Diamond\Box p \to \Diamond\Box\Diamond p$ is prehistoric-loop-free provable in both $\mathsf{T}$ and $\mathsf{S4}$, but not so in $\mathsf{K4}$. Note that this is an analogue of the result on self-referentiality of this formula, and hence the agreement of loop-free provability and non-self-referentiality is still possible.

At the end of this chapter, we present two figures.

Figure 3.1 shows the relationship between modal logics $\mathsf{K4}$, $\mathsf{S4}$, and their non-directly-self-referential fragments. In this figure, inclusions of sets are obvious, except the fact that $\mathsf{K4}^{\wp_1} \subseteq \mathsf{S4}^{\wp_1}$ which follows from Theorem 76. The formula $\Diamond(p \to \Box p)$ is Kuznets' instance for direct self-referentiality in $\mathsf{S4}$ in [14]. This is not a $\mathsf{K4}$-theorem, which draws Kuznets' later work [45] where $\Diamond\top \to \Diamond(p \to \Box p)$ is found. Results in those papers are summarized in Table 2.1. Facts about $\bot \to \bot$ and $\Box p \to p$ follow from Theorem 72. Facts about $\Diamond\Box p \to \Diamond\Box\Diamond p$ are found in this section, which completes this figure.

In the Figure 3.2 for $\mathsf{T}$, inclusions of sets are obvious otherwise follow from Theorem 76. The formula $\Diamond(p \to \Box p)$ is due to Kuznets [45]. Facts about $\bot \to \bot$ and $\Box p \to \Box\Box p$ follow from Theorem 72. The fact that $\Diamond\Box(\Diamond\Box p \to \Box p)$ is not provable in $\mathsf{T}$ is obvious, and its direct self-referentiality in $\mathsf{S4}$ follows from Theorem 88 in Section 4.2. It is still

Figure 3.2: Direct Self-referentiality in T and S4.




open that whether is the space marked by question marks is empty or not.

Given observations of this section, we know that it makes so sense to talk about non-self-referential fragments of modal logic in general, and we have to specify which modal/justification logic pair we are talking about.

# Chapter 4

# Self-referentiality in the BHK Semantics

In Section 1.2, we presented a brief introduction to the provability semantics of intuitionistic logic, also known as Brouwer–Heyting–Kolmogorov (BHK) semantics. The propositional part of this semantics is formulated in Gödel–Artemov approach that employs the realization of $\mathsf{S4}$ in $\mathsf{LP}$. Thus, Artemov's question about necessity of self-referentiality also makes sense in this formulation of BHK semantics.

Kuznets' result in [14] (Theorem 16 in this thesis) shows that as an $\mathsf{S4}$-thoerem, $\neg\Box\neg(p \to \Box p) \notin \mathsf{S4}^{\not\circlearrowleft_1}$, from which we know that direct self-referentiality is necessary for the provability semantics of $\mathsf{S4}$. This does not answer Artemov's question for BHK semantics, since by an easy observation of Gödel's modal embedding $(\cdot)^{\triangle}$ in Section 1.2, we notice that $\neg\Box\neg(p \to \Box p)$ is not the image of any $\mathsf{IPC}$-theorem under $(\cdot)^{\triangle}$. The same for other known embeddings (cf. [30], [17], and [63]). The first four sections in this chapter present the candidate's results in [70, 71] that there is no "basic" way to embed $\mathsf{IPC}$ into $\mathsf{S4}^{\not\circlearrowleft_1}$.

In Section 4.1, we define the class of basic embeddings, a natural extension of Gödel's modal embedding.

Then Section 4.2 provides a set of directly self-referential $\mathsf{S4}$-theorems.

In Section 4.3, we see that the double-negation of each intuitionistically invalid tautology falls into the set found in Section 4.2 by any basic embedding. That is to say, no basic embedding can map all $\mathsf{IPC}$-theorems into $\mathsf{S4}^{\wp_1}$.

In Section 4.4, we see that even the purely implicational fragment of $\mathsf{IPC}$ cannot be mapped into $\mathsf{S4}^{\wp_1}$ by any basic embedding.

Finally in Section 4.5, we apply the idea of Chapter 3 on fragments to $\mathsf{IPC}$.

## 4.1 Basic Embeddings

This Section contains materials reused from the candidate's published paper [71], with the kind permission from Elsevier, the owner of the copyright of [71]. This Section contains materials reused from the candidate's published paper [70], with the kind permission from Springer, the owner of the copyright of [70].

There are several known ways to faithfully embed IPC in S4, with minor differences between them (cf. [30], [49], [17], and [63]). Three of them are presented in Table 4.1, where $(\cdot)^\circ$ and $(\cdot)^\Box$ are contained in [63], and $(\cdot)^\triangle$ is the well-known Gödel's "$\Box$ each subformula" embedding.[1]

**Fact 83 ([63])** *Embeddings in Table 4.1 are all faithful, i.e., for each* $(\cdot)^\times \in \{(\cdot)^\circ, (\cdot)^\Box, (\cdot)^\triangle\}$, $\mathsf{IPC} \vdash \phi$ *iff* $\mathsf{S4} \vdash \phi^\times$. *Differences between these embeddings are minor, since by induction, we see that* $\mathsf{S4} \vdash \phi^\Box \leftrightarrow \phi^\triangle$ *and* $\mathsf{S4} \vdash \Box\phi^\times \leftrightarrow \Box\phi^+$ *for any propositional formula* $\phi$, *and any* $(\cdot)^\times, (\cdot)^+ \in \{(\cdot)^\circ, (\cdot)^\Box, (\cdot)^\triangle\}$.

Though differences between embeddings $(\cdot)^\circ$, $(\cdot)^\Box$, and $(\cdot)^\triangle$ are minor, they can affect realizations in LP. When an S4-theorem is realized, each $\Box$ is replaced by a term, hence

[1] In [30], Gödel presented a version of embedding along with some alternative options, say, whether or not to add $\Box$'s in the $\wedge$ case. The version Gödel presented has $\neg$, but not $\bot$, as a primitive connective, and hence looks slightly different. Except for this, if we take the alternative option that adds $\Box$'s in the $\wedge$ case, then the only difference between the resulting version and $(\cdot)^\triangle$ is the outermost $\Box$.

Table 4.1: Three Faithful Embeddings of IPC in S4.

| | $(\cdot)^\circ$ | $(\cdot)^\Box$ | $(\cdot)^\triangle$ |
|---|---|---|---|
| $\bot$ | $\bot$ | $\bot$ | $\Box\bot$ |
| $p$ | $p$ | $\Box p$ | $\Box p$ |
| $\alpha \wedge \beta$ | $\alpha^\circ \wedge \beta^\circ$ | $\alpha^\Box \wedge \beta^\Box$ | $\Box(\alpha^\triangle \wedge \beta^\triangle)$ |
| $\alpha \vee \beta$ | $\Box\alpha^\circ \vee \Box\beta^\circ$ | $\alpha^\Box \vee \beta^\Box$ | $\Box(\alpha^\triangle \vee \beta^\triangle)$ |
| $\alpha \to \beta$ | $\Box\alpha^\circ \to \beta^\circ$ | $\Box(\alpha^\Box \to \beta^\Box)$ | $\Box(\alpha^\triangle \to \beta^\triangle)$ |

the resulting realization is quite sensitive to the exact syntactical form of the modal formula. It is therefore reasonable to consider possible embeddings of IPC in S4 in a more general setting.

In what follows, we consider the class of "basic" embeddings. This class initially appears in the candidate's [70] with only the "unpolarized version," and then appears in full in the candidate's [71] under the name "polarized basic embedding." Since polarities make no difference in results that will come in this chapter, we simply use "basic embedding" in the general sense of "polarized basic embedding" from [71].

**Definition 84 (Basic Embeddings)** (1) A mapping from intuitionistic propositional formulas to modal formulas, denoted by $(\cdot)^{\times}$, is a *basic potential embedding* if it, together with $(\cdot)^{\times}_{+}$ and $(\cdot)^{\times}_{-}$, satisfy:

$$\begin{cases} p^{\times}_{+} \equiv \Box^{h_+} p, \\ p^{\times}_{-} \equiv \Box^{h_-} p, \\ \bot^{\times}_{+} \equiv \Box^{i_+} \bot, \\ \bot^{\times}_{-} \equiv \Box^{i_-} \bot, \\ (\phi \odot \psi)^{\times}_{+} \equiv \Box^{j_{\odot +}}(\Box^{k_{\odot +}} \phi^{\times}_{+} \odot \Box^{l_{\odot +}} \psi^{\times}_{+}) \text{ for } \odot \in \{\wedge, \vee\}, \\ (\phi \odot \psi)^{\times}_{-} \equiv \Box^{j_{\odot -}}(\Box^{k_{\odot -}} \phi^{\times}_{-} \odot \Box^{l_{\odot -}} \psi^{\times}_{-}) \text{ for } \odot \in \{\wedge, \vee\}, \\ (\phi \to \psi)^{\times}_{+} \equiv \Box^{j_{\to +}}(\Box^{k_{\to +}} \phi^{\times}_{-} \to \Box^{l_{\to +}} \psi^{\times}_{+}), \\ (\phi \to \psi)^{\times}_{-} \equiv \Box^{j_{\to -}}(\Box^{k_{\to -}} \phi^{\times}_{+} \to \Box^{l_{\to -}} \psi^{\times}_{-}), \text{ and} \\ \phi^{\times} \equiv \phi^{\times}_{+}. \end{cases}$$

In these equations, the $p$, $\bot$, $\odot$, or $\to$ at the left side is called the *source* of the one at the right side. Natural numbers $h_+, h_-, i_+, i_-, j_{\odot+}, j_{\odot-}, k_{\odot+}, k_{\odot-}, l_{\odot+}, l_{\odot-}$ for $\odot \in \{\wedge, \vee, \to\}$ are called *parameters* of the embedding. There are 22 parameters for each basic potential embedding. By saying that $(\cdot)^{\times}$ is a basic (potential) embedding, we assume that those notations are all reserved for the parameters, in the way indicated above. Notations $j_+$, $k_+$, and $l_+$ are also reserved as abbreviates of $j_{\to+}$, $k_{\to+}$, and $l_{\to+}$, respectively. Similarly for $j_-$, $k_-$, and $l_-$.

(2) A basic potential embedding $(\cdot)^{\times}$ is *unpolarized*, if it satisfies $\phi^{\times}_{+} \equiv \phi^{\times}_{-}$, or equiva-

lently, if $h_+ = h_-$, $i_+ = i_-$, ..., $l_{\vee+} = l_{\vee-}$ all hold for the 11 pairs of parameters. In this case we omit + or − subscripts in notations for parameters.

(3) A basic potential embedding $(\cdot)^\times$ is a *basic embedding* if it is faithful, i.e., if it satisfies: $\mathsf{IPC} \vdash \phi$ iff $\mathsf{S4} \vdash \phi^\times$. ⊣

**Fact 85** *The three embeddings in Table 4.1 are all basic embeddings, and are all unpolarized.*

**Proof.** For example, $(\cdot)^\square$ is the basic embedding for which $h = j = 1$ and

$i = k = l = j_\wedge = k_\wedge = l_\wedge = j_\vee = k_\vee = l_\vee = 0.$ ⊣

**Remark:** As the reader may have observed, if we consider the embedding of IPC into S4 by its own, it only matters whether a parameter has value 0, or a positive value, since $\mathsf{S4} \vdash \square\phi \leftrightarrow \square\square\phi$. However, since the realization of an S4-theorem in LP is quite sensitive to the exact syntactical form, we do not use this observation to simply our definition of basic embedding. Another benefit of our general definition is, it can also be applied to modal embeddings between other logics, e.g., from Visser's basic propositional logic BPL to K4 (cf.[64]). Though not included in this thesis, it will be interesting to see which basic embeddings are faithful and which are not, for a particular pair of logics.

We will consider embeddings of some specific formulas under a basic potential embedding, and it is straightforward but tedious to calculate them. Some examples of calculations are given in Example 86, and when using a calculation of this kind, we will say "by calculation."

**Example 86** *Let* $(\cdot)^\times$ *be a basic potential embedding, then*

$$
\begin{aligned}
(\neg\phi)^\times &\equiv (\phi\to\bot)^\times_+ \equiv \Box^{j_+}(\Box^{k_+}\phi^\times_- \to \Box^{l_+}\bot^\times_+) \equiv \Box^{j_+}(\Box^{k_+}\phi^\times_- \to \Box^{i_++l_+}\bot),\\
(\neg\neg\phi)^\times &\equiv \Box^{j_+}(\Box^{k_+}(\neg\phi)^\times_- \to \Box^{i_++l_+}\bot) \equiv \Box^{j_+}(\Box^{k_+}\Box^{j_-}(\Box^{k_-}\phi^\times_+ \to \Box^{i_-+l_-}\bot)\to\Box^{i_++l_+}\bot)\\
&\equiv \Box^{j_+}(\Box^{j_-+k_+}(\Box^{k_-}\phi^\times_+ \to \Box^{i_-+l_-}\bot)\to\Box^{i_++l_+}\bot),\\
((\phi\to q)\to q)^\times &\equiv \Box^{j_+}(\Box^{k_+}(\phi\to q)^\times_- \to \Box^{l_+}q^\times_+) \equiv \Box^{j_+}(\Box^{k_+}\Box^{j_-}(\Box^{k_-}\phi^\times_+ \to \Box^{l_-}q^\times_-)\to\Box^{h_++l_+}q)\\
&\equiv \Box^{j_+}(\Box^{j_-+k_+}(\Box^{k_-}\phi^\times_+ \to \Box^{h_-+l_-}q)\to\Box^{h_++l_+}q).
\end{aligned}
$$

*By definition, these also give the embedding images under* $(\cdot)^\times_+$. *For the embedding images under* $(\cdot)^\times_-$, *simply read each parameter as the one with opposite polarity symbol (e.g., read* $j_+$ *as* $j_-$ *and vice versa) in the equations above.*

Some basic potential embeddings are faithful, some are not. We have the following lemma, which provide some necessary conditions on parameters.

**Lemma 87** *For any basic embedding:*

*(A)* $j_+ + j_- + k_+ + k_- > 0$*,*

*(B)* $j_- + k_+ > 0$*,*

*(C)* $j_+ + k_- > 0$*, and*

*(D)* $h_- + k_+ > 0$*.*

**Proof.** (A) Let $(\cdot)^\times$ be a basic potential embedding with $j_+ = j_- = k_+ = k_- = 0$. By Example 86,

$$(\neg\neg\phi)^\times \equiv \Box^0(\Box^{0+0}(\Box^0\phi^\times_+ \to \Box^{i_-+l_-}\bot)\to\Box^{i_++l_+}\bot) \equiv (\phi^\times_+ \to \Box^{i_-+l_-}\bot)\to\Box^{i_++l_+}\bot.$$

Under S4-provability, the formula above equals to $(\phi^\times_+ \to \bot)\to\bot$, then to $\phi^\times_+$, and to $\phi^\times$. Thus, S4 $\vdash (\neg\neg\phi)^\times \leftrightarrow \phi^\times$. Let $\phi$ be $\neg\neg p\to p$, then S4 $\vdash (\neg\neg(\neg\neg p\to p))^\times \leftrightarrow (\neg\neg p\to p)^\times$. Note that IPC $\vdash \neg\neg(\neg\neg p\to p)$ and IPC $\nvdash \neg\neg p\to p$.

Assume with the seek of a contradiction that $(\cdot)^\times$ is faithful. Then $\mathsf{S4} \vdash (\neg\neg(\neg\neg p \to p))^\times$ and $\mathsf{S4} \nvdash (\neg\neg p \to p)^\times$. Thus we have a contradiction.

(B) Let $(\cdot)^\times$ be a basic potential embedding that satisfies $j_- = k_+ = 0$.

Case (i), $j_+ + l_+ > 0$. By calculation,

$$\begin{aligned}((p\to q)\to(p\to q))^\times &\equiv \Box^{j_+}(\Box^{j_-+k_+}(\Box^{k_-}p_+^\times\to\Box^{l_-}q_-^\times)\to\Box^{j_++l_+}(\Box^{k_+}p_-^\times\to\Box^{l_+}q_+^\times))\\ &\equiv \Box^{j_+}((\Box^{k_-}p_+^\times\to\Box^{l_-}q_-^\times)\to\Box^{j_++l_+}(p_-^\times\to\Box^{l_+}q_+^\times)),\end{aligned}$$

which is not a theorem of $\mathsf{S4}$ no matter what value $h_+$ and $h_-$ have. Since $\mathsf{IPC} \vdash (p\to q)\to(p\to q)$, we see that $(\cdot)^\times$ is not faithful.

Case (ii), $j_+ = l_+ = 0$. By calculation,

$$\begin{aligned}(((p\to q)\to r)\to((p\to q)\to r))^\times &\equiv \Box^{j_+}(\Box^{j_-+k_+}(\Box^{j_++k_-}(\Box^{k_+}p_-^\times\to\Box^{l_+}q_+^\times)\to\Box^{l_-}r_-^\times)\\ &\quad\to\Box^{j_++l_+}(\Box^{j_-+k_+}(\Box^{k_-}p_+^\times\to\Box^{l_-}q_-^\times)\to\Box^{l_+}r_+^\times))\\ &\equiv (\Box^{k_-}(p_-^\times\to q_+^\times)\to\Box^{l_-}r_-^\times)\to((\Box^{k_-}p_+^\times\to\Box^{l_-}q_-^\times)\to r_+^\times).\end{aligned}$$

By (A), $j_+ + j_- + k_+ + k_- > 0$. Given $j_+ = j_- = k_+ = 0$ by assumption, we have $k_- > 0$, and hence the formula above is not provable in $\mathsf{S4}$ no matter what value $h_+$ and $h_-$ have. Since $\mathsf{IPC} \vdash ((p\to q)\to r)\to((p\to q)\to r)$, the embedding $(\cdot)^\times$ is not faithful.

In summary, $(\cdot)^\times$ is not faithful in both cases.

(C) Let $(\cdot)^\times$ be a basic potential embedding that satisfies $j_+ = k_- = 0$. By calculation,

$$\begin{aligned}(((p\to q)\to p)\to p)^\times &\equiv \Box^{j_+}(\Box^{j_-+k_+}(\Box^{j_++k_-}(\Box^{k_+}p_-^\times\to\Box^{l_+}q_+^\times)\to\Box^{l_-}p_-^\times)\to\Box^{l_+}p_+^\times)\\ &\equiv \Box^{j_-+k_+}((\Box^{k_+}p_-^\times\to\Box^{l_+}q_+^\times)\to\Box^{l_-}p_-^\times)\to\Box^{l_+}p_+^\times.\end{aligned}$$

By (A), we have $j_- + k_+ > 0$, and hence the formula above is provable in $\mathsf{S4}$ no matter what value $h_+$ and $h_-$ have. Since $\mathsf{IPC} \nvdash ((p\to q)\to p)\to p$, we know that $(\cdot)^\times$ is not faithful.

(D) Let $(\cdot)^\times$ be a basic potential embedding that satisfies $h_- = k_+ = 0$. Suppose with the seek of a contradiction that $(\cdot)^\times$ is faithful.

Case (i), $h_+ + j_+ + l_+ > 0$. Since $\mathsf{IPC} \vdash p \to (q \to p)$, we should have $\mathsf{S4} \vdash (p \to (q \to p))^\times$.

By calculation,

$$\begin{aligned}(p \to (q \to p))^\times &\equiv \Box^{j_+}(\Box^{h_-+k_+} p \to \Box^{j_++l_+}(\Box^{h_-+k_+} q \to \Box^{h_++l_+} p)) \\ &\equiv \Box^{j_+}(p \to \Box^{j_++l_+}(q \to \Box^{h_++l_+} p)),\end{aligned}$$

If this is an $\mathsf{S4}$-theorem, then $\mathsf{S4} \vdash p \to \Box^{j_++l_+}(q \to \Box^{h_++l_+} p)$.

If $j_+ + l_+ > 0$, then $\mathsf{S4} \vdash p \to \Box(q \to \Box^{h_++l_+} p)$, which is not the case.

So $j_+ = l_+ = 0$, and hence we should have $\mathsf{S4} \vdash p \to (q \to \Box^{h_++l_+} p)$. Since $h_+ + j_+ + l_+ > 0$ by assumption of this case, we have $h_+ > 0$, and hence we should have $\mathsf{S4} \vdash p \to (q \to \Box p)$, which is not the case.

Neither subcase is possible, we have a contradiction.

Case (ii), $h_+ = j_+ = l_+ = 0$. Since $\mathsf{IPC} \vdash ((((p \to q) \to p) \to p) \to q) \to q$, we should have $\mathsf{S4} \vdash (((((p \to q) \to p) \to p) \to q) \to q)^\times$.

By calculation,

$$\begin{aligned}&(((((p \to q) \to p) \to p) \to q) \to q)^\times \\ \equiv\ & \Box^{j_+}(\Box^{j_-+k_+}( \\ &\quad \Box^{j_++k_-}(\Box^{j_-+k_+}(\Box^{j_++k_-}(\Box^{h_-+k_+} p \to \Box^{h_++l_+} q) \to \Box^{h_-+l_-} p) \to \Box^{h_++l_+} p) \to \Box^{h_-+l_-} q \\ &\qquad\qquad ) \to \Box^{h_++l_+} q) \\ \equiv\ & \Box^{j_-}(\Box^{k_-}(\Box^{j_-}(\Box^{k_-}(p \to q) \to \Box^{l_-} p) \to p) \to \Box^{l_-} q) \to q\end{aligned}$$

Since $j_+ = 0$ and $k_+ = 0$, we have $k_- > 0$ and $j_- > 0$ by (C) and (B), respectively. Thus, if the formula above is an S4-theorem, then

$$\mathsf{S4} \vdash \Box(\Box(\Box(\Box(p \to q) \to \Box^{l_-} p) \to p) \to \Box^{l_-} q) \to q,$$

which is not the case, a contradiction.

In summary, each case leads to a contradiction, and hence $(\cdot)^\times$ is not faithful. ⊣

At the end of this section, we have some discussions about the basic embedding.

Gödel's embedding is often presented as "□-prefixing each subformula" (i.e., $(\cdot)^\vartriangle$ in our notation), although he also introduced some variants in [30]. In McKinsey and Tarski's [49], faithfulness was shown for three different embeddings $T$, $T'$, and $T''$; among them $T$ can be roughly described as $(\cdot)^\Box$ with $\neg$ and $\Diamond$ as primitives instead of $\bot$ and $\Box$. The embedding $(\cdot)^\circ$ originates from Girard's embedding of IPC in classical linear logic [28]. Chagrov and Zakharyashchev presented eight embeddings in [17], with different primitives and different manners of definition. In [70], the notion of basic embedding first appears as its unpolarized version, and is then extended to the general case in [71], which we are following with some notational alternatives.

A natural question is: *how general is our "basic" embedding*? Our answer follows. Inevitably, a modal embedding has to be defined with $\wedge$, $\vee$, and $\to$ as primitives, as they are independent of each other. Being alternatives of each other we have $\bot$ and $\neg$ for connectives, and $\Box$ and $\Diamond$ for modalities. Since BHK semantics is a provability semantics, it is more natural to have $\Box$ (provable) instead of $\Diamond$, though they are no essential technical differences. Among known embeddings cited above, any embedding defined with $\bot$ and $\Box$ as primitives is basic, and an embedding defined with $\neg$ or $\Diamond$ as a primitive may work differently (in some cases we may still find the parameters $i_+, i_-$ so that $(\neg\phi)^\times \equiv (\phi \to \bot)^\times$, while in some other cases, because of the parameters already given to $\to$ and $\neg$, this is not possible, like in embeddings $T'$ and $T''$ from [49]). It should be emphasized that, in Section 4.4, we will show the direct self-referentiality of the purely implicational fragment of $\mathsf{IPC}$. As only implication is involved, it no longer matters whether $\perp$ or $\neg$ is taken as primitive.

## 4.2 A Class of Directly Self-referential S4-formulas

This Section contains materials reused from the candidate's published paper [71], with the kind permission from Elsevier, the owner of the copyright of [71]. This Section contains materials reused from the candidate's published paper [70], with the kind permission from Springer, the owner of the copyright of [70].

(Direct) self-referentiality of S4 was first proved by Kuznets in [14] (cf. Theorem 16 in Section 2.1) where the example formula (2.1) was given. In order to show a similar result on the IPC layer, we adapt Kuznets' method from [14], and find a natural class of S4-formulas that require direct self-referentiality in LP.

Recall that we have following notations and conventions. For formulas $\phi$ and $\psi$, by $\phi \equiv \psi$ we mean $\phi$ and $\psi$ are syntactically identical. For instance, we may write $\neg\phi \equiv \phi \rightarrow \bot$, but $\bot \not\equiv \bot \wedge \bot$. For terms $s, t$ and formula $\phi$, by $s \rhd \phi$ (or $s \rhd t$), we mean $s$ occurs in $\phi$ (or $t$, respectively). Let $\Box^i \phi$ be an abbreviation of $\underbrace{\Box \cdots \Box}_{i} \phi$ for any natural number $i$ (note that $\Box^0 \phi \equiv \phi$). For pairwise distinct atoms $p_1, ..., p_m$, and formulas $\phi, \psi_1, ..., \psi_m$, let $\phi[\frac{p_e}{\psi_e}]_{e=1}^m$ be the result of simultaneously substituting $p_1, ..., p_m$ in $\phi$ by $\psi_1, ..., \psi_m$, respectively.

The following theorem gives a class of S4-formulas that cannot be realized in LP without calling for a directly self-referential constant specification.

**Theorem 88** *Assume the following:*

*(a) All of $m, u, w, z, n_1, ..., n_m$ are natural numbers, and $w, n_1, ..., n_m > 0$.*

*(b) Both $\rho$ and $\zeta$ are propositional formulas with $\bot, \wedge, \vee$ as their only connectives.*[2] *All of $\theta, \xi_1, ..., \xi_m$ are modal formulas.*

[2]We have chosen $\bot, \rightarrow$ as our primitive connectives for classical propositional formulas. Here by saying "with $\bot, \wedge, \vee$ as their only connectives," we actually mean $\rho$ and $\zeta$ can be presented using these three (primitive or defined) connectives only. Note that $\{\bot, \wedge, \vee\}$ is not a complete set of classical propositional connectives.

*(c) $P = \{p_1, ..., p_m\}$ is a set of atoms where each of $p_1, ..., p_m$ occurs in $\rho$ exactly once. Each of $\xi_1, ..., \xi_m$ does not have $\Box$ as its main connective, and does not contain any occurrence of any atom from $P$.*

*(d)* $\mathsf{S4} \nvdash \rho[\frac{p_e}{\Box^{n_e}\xi_e}]_{e=1}^m$.

*Then any realization of $\Box^u(\Box^w(\rho[\frac{p_e}{\Box^{n_e}\xi_e}]_{e=1}^m \to \theta) \to \Box^z\zeta)$ has a counter $\mathcal{CS}^{\varphi_1}$-model.*

**Proof.** Since $\rho$ and $\zeta$ are propositional formulas, we see that $\rho^r \equiv \rho$ and $\zeta^r \equiv \zeta$. Thus, any possible realization of $\Box^u(\Box^w(\rho[\frac{p_e}{\Box^{n_e}\xi_e}]_{e=1}^m \to \theta) \to \Box^z\zeta)$ has the form

$$t_u^0 : \cdots : t_1^0 : (t_w^1 : \cdots : t_1^1 : (\rho[\frac{p_e}{t_{n_e}^{\xi_e} : \cdots : t_1^{\xi_e} : \xi_e^r}]_{e=1}^m \to \theta^r) \to t_z^4 : \cdots : t_1^4 : \zeta).$$

We have the following abbreviations:

$$\begin{array}{rcl}
I & := & \{1, ..., m\} \text{ (the set of indexes of elements of } P), \\
\mathfrak{S} & := & [\frac{p_e}{t_{n_e}^{\xi_e} : \cdots : t_1^{\xi_e} : \xi_e^r}]_{e=1}^m \ , \\
\delta & := & \rho\mathfrak{S} \to \theta^r, \\
X & := & \{t_1^1 : \delta, t_2^1 : t_1^1 : \delta, \ \dots, t_w^1 : \cdots : t_1^1 : \delta\}, \\
X^- & := & \{\delta, t_1^1 : \delta, \ \dots, t_{w-1}^1 : \cdots : t_1^1 : \delta\}, \\
*^0 & := & *_{\mathcal{CS}^{\varphi_1}} \ , \\
*^1 & := & *_{\mathcal{CS}^{\varphi_1} \cup X} \ , \\
*^{1\backslash 0} & := & *^1 \setminus *^0 \text{ (this needs not to be an evidence function itself).}
\end{array}$$

Here $w > 0$ ensures that $X \neq \emptyset$, and hence by the definition of $X^-$, we know that for any $\psi \in \mathcal{F}$:

$$t : \psi \in X \text{ for some } t \in \mathcal{T} \text{ iff } \psi \in X^-. \tag{4.1}$$

We also have

$$\delta \text{ is a subformula of } \psi \text{ for any } \psi \in X \cup X^-. \tag{4.2}$$

Let $v$ be a propositional valuation that falsifies all atoms. The desired counter model is $\mathfrak{M}^r = (*^1, v)$.

For any constant $c$, axiom $A$ s.t. $c : A \in \mathcal{CS}^{\varphi_1}$, we have $c : A \in \mathcal{CS}^{\varphi_1} \cup X$, and hence $A \in *^1(c)$. So $\mathfrak{M}^r$ is indeed a $\mathcal{CS}^{\varphi_1}$-model.

Observe that it is sufficient to show $\mathfrak{M}^r \nvDash \rho\mathfrak{S}$. To see this, suppose that $\mathfrak{M}^r \nvDash \rho\mathfrak{S}$, then $\mathfrak{M}^r \vDash \delta$. By the definition of $*^1$ and of $X$, we have $\delta \in *^1(t^1_1)$, $t^1_1 : \delta \in *^1(t^1_2)$, ..., and $t^1_{w-1} : \cdots : t^1_1 : \delta \in *^1(t^1_w)$. Thus $\mathfrak{M}^r \vDash t^1_w : \cdots : t^1_1 : \delta$. Since $\zeta$ is propositional and has $\bot, \wedge, \vee$ as its only connectives, we see by an induction that $v(\zeta) = 0$. So $\mathfrak{M}^r \nvDash \zeta$, which implies $\mathfrak{M}^r \nvDash t^4_z : \cdots : t^4_1 : \zeta$. Therefore, $\mathfrak{M}^r \nvDash t^1_w : \cdots : t^1_1 : \delta \to t^4_z : \cdots : t^4_1 : \zeta$, and hence $\mathfrak{M}^r \nvDash t^0_u : \cdots : t^0_1 : (t^1_w : \cdots : t^1_1 : \delta \to t^4_z : \cdots : t^4_1 : \zeta)$. What we construct is actually a counter $\mathcal{CS}^{\varphi_1}$-model.

In order to show $\mathfrak{M}^r \nvDash \rho\mathfrak{S}$, we state and prove two claims.

**Claim 89** *Assume $m > 0$. Let $g \in I$, and let $s$ be a subterm of $t^{\xi_g}_1$.*

*(A) For any $\phi \in \mathcal{F}$, if $\phi \in *^0(s)$, then*

*(i) $\mathsf{LP}(\mathcal{CS}^{\varphi_1}) \vdash \phi$,*

*(ii) $t^{\xi_g}_1 \ntriangleright \phi$;*

*(B) For any $\phi \in \mathcal{F}$, if $\phi \in *^{1\setminus 0}(s)$, then*

*(iii) $t^{\xi_f}_1 \triangleright \phi$ for each $f \in I$,*

*(iv) $\phi$ is not prime, and $\phi \equiv \delta$ if $\phi$ is an implication,*

*(v) $\phi \notin H_g = \{t^{\xi_g}_1 : \xi^r_g, t^{\xi_g}_2 : t^{\xi_g}_1 : \xi^r_g, \ldots, t^{\xi_g}_{n_g} : \cdots : t^{\xi_g}_1 : \xi^r_g, \rho\mathfrak{S}\}$.*[3]

**Proof.** Induction on $s$, with the IH that (A) and (B) hold for each proper subterm of $s$.

First observe that if $\phi \in X^-$, then all of (iii), (iv), (v) hold. To see this, suppose $\phi \in X^-$. By (4.2), $\delta$ is a subformula of $\phi$. For any $f \in I$, since $t^{\xi_f}_1 \triangleright \delta$, we have $t^{\xi_f}_1 \triangleright \phi$, (iii) holds.

[3]In case that $\rho$ is an atom $p_g \in P$, we have $t^{\xi_g}_{n_g} : \cdots : t^{\xi_g}_1 : \xi^r_g \equiv \rho\mathfrak{S}$.

There is no prime formula in $X^-$ and $\delta$ is the only implication in $X^-$, therefore (iv) holds. To see that (v) holds, suppose that $\phi \in H_g$, which implies $\phi$ is a proper subformula of $\delta$. Since $\delta$ is a subformula of $\phi$, we have $\delta$ as a proper subformula of $\delta$ itself, a contradiction. Therefore, we have:

$$\text{all of (iii), (iv), (v) follow from } \phi \in X^-. \tag{4.3}$$

(1) $s \equiv c$ for a constant $c$:

(1A) Assume $\phi \in *^0(c)$. (1A1) Via initial set. Then $c{:}\phi \in \mathcal{CS}^{\varphi_1}$. So $\phi$ is an axiom, hence (i) holds. Suppose that $t_1^{\xi_g} \rhd \phi$, then $c \rhd \phi$ since $c$ is a subterm of $t_1^{\xi_g}$, and hence $\mathcal{CS}^{\varphi_1}$ is directly self-referential, a contradiction. Thus, (ii) holds. (1A2) Via closure. This is impossible, since closure conditions in Definition 21 put $\phi$ into $*^0(c)$ only if $c$ is a composite term, which is not the case.

(1B) Assume $\phi \in *^{1\backslash 0}(c)$. (1B1) Via initial set. Then $c{:}\phi \in X$ (precisely, $(\mathcal{CS}^{\varphi_1} \cup X) \setminus \mathcal{CS}^{\varphi_1}$, which is a subset of $X$), which implies $\phi \in X^-$ by (4.1). Now all of (iii), (iv), (v) hold, by (4.3). (1B2) Via closure. Impossible, since $s \equiv c$ is a constant.

(2) $s \equiv x$ for a constant $x$:

(2A) Assume $\phi \in *^0(x)$. (2A1) Via initial set. Then $x{:}\phi \in \mathcal{CS}^{\varphi_1}$, which is impossible, since $x$ is not a constant. (2A2) Via closure. Impossible, since $s \equiv x$ is a variable.

(2B) Assume $\phi \in *^{1\backslash 0}(x)$. (2B1) Via initial set. Then $x{:}\phi \in X$, which implies $\phi \in X^-$ (by (4.1)). From (4.3), all of (iii), (iv), (v) hold. (2B2) Via closure. Impossible, since $s \equiv x$ is a variable.

(3) $s \equiv !s_1$ for a term $s_1$, and by IH both (A) and (B) hold for $s_1$:

(3A) Assume $\phi \in *^0(!s_1)$. (3A1) Via initial set. Then $!s_1 : \phi \in \mathcal{CS}^{\varphi_1}$, which is impossible since $s \equiv !s_1$ is not a constant. (3A2) Via closure. According to Definition 21, $\phi \equiv s_1 : \psi$ for some $\psi \in *^0(s_1)$. By IH, $\mathsf{LP}(\mathcal{CS}^{\varphi_1}) \vdash \psi$. For any $\mathcal{CS}^{\varphi_1}$-model $\mathfrak{M}$, by

Theorem 26, $\mathfrak{M} \vDash \psi$. Since $\psi \in *^0(s_1)$ where $*^0$ is the smallest $\mathcal{CS}^{\varphi_1}$-evidence function (by Lemma 22), $\psi$ is also in the image of $s_1$ under the evidence function of $\mathfrak{M}$, and hence $\mathfrak{M} \vDash s_1 : \psi$. By Theorem 26, $\mathsf{LP}(\mathcal{CS}^{\varphi_1}) \vdash s_1 : \psi$, (i) holds. Also by IH, $t_1^{\xi_g} \ntriangleright \psi$. Since $s_1$ is a proper subterm of $t_1^{\xi_g}$, we know $t_1^{\xi_g} \ntriangleright s_1$. Hence, $t_1^{\xi_g} \ntriangleright s_1 : \psi$, (ii) holds.

(3B) Assume $\phi \in *^{1\setminus 0}(!s_1)$. (3B1) Via initial set. Then $!s_1 : \phi \in X$, which implies $\phi \in X^-$ with the help of (4.1). Now (4.3) ensures all of (iii), (iv), and (v). (3B2) Via closure. Then $\phi \equiv s_1 : \psi$ for some $\psi \in *^{1\setminus 0}(s_1)$. For any $f \in I$, by IH, $t_1^{\xi_f} \triangleright \psi$, hence $t_1^{\xi_f} \triangleright s_1 : \psi$, (iii) holds. Since $s_1 : \psi$ is neither a prime formula nor an implication, (iv) holds in a vacuous way. By IH, $\psi \notin H_g = \{t_1^{\xi_g} : \xi_g^r, t_2^{\xi_g} : t_1^{\xi_g} : \xi_g^r, \dots, t_{n_g}^{\xi_g} : \dots : t_1^{\xi_g} : \xi_g^r, \rho\mathfrak{S}\}$, from which we can derive that

$$s_1 : \psi \notin \{t_2^{\xi_g} : t_1^{\xi_g} : \xi_g^r, \dots, t_{n_g}^{\xi_g} : \dots : t_1^{\xi_g} : \xi_g^r\}. \tag{4.4}$$

To see that (v) holds, it is then sufficient to show that $s_1 : \psi \neq t_1^{\xi_g} : \xi_g^r$ and $s_1 : \psi \neq \rho\mathfrak{S}$. The former holds, since $s_1$ is a proper subterm of $t_1^{\xi_g}$. For the latter, we reason by subcases. (3B21) $\rho \equiv \bot$ or $\rho \equiv p$ for some atom $p \notin P$. By assumption (c) of Theorem 88, $P = \varnothing$. Hence $m = 0$, which violates the assumption of this claim. (3B22) $\rho \equiv p$ for some atom $p \in P$, then by assumptions of Theorem 88, $\rho \equiv p \equiv p_1$, $P = \{p_1\}$, $m = 1$, and $I = \{1\}$. Thus we have $g = 1$, and hence

$$\rho\mathfrak{S} \equiv p_1[\frac{p_e}{t_{n_e}^{\xi_e} : \dots : t_1^{\xi_e} : \xi_e^r}]_{e=1}^1 \equiv t_{n_1}^{\xi_1} : \dots : t_1^{\xi_1} : \xi_1^r,$$

while we have verified $s_1 : \psi \neq t_{n_1}^{\xi_1} : \dots : t_1^{\xi_1} : \xi_1^r$ at (4.4). (3B23), if none of the first two subcases hold, then by assumption (b) of Theorem 88, $\rho$ has main connective $\wedge$ or $\vee$, so is $\rho\mathfrak{S}$. Therefore, $s_1 : \psi \neq \rho\mathfrak{S}$. In summary, $s_1 : \psi \neq \rho\mathfrak{S}$ in all subcases, (v) holds.

(4) $s \equiv s_1 \cdot s_2$ for some terms $s_1$ and $s_2$, while by IH (A) and (B) hold for both $s_1$ and $s_2$:

(4A) Assume $\phi \in *^0(s_1 \cdot s_2)$. (4A1) Via initial set. Then $(s_1 \cdot s_2) : \phi \in \mathcal{CS}^{\varphi_1}$, which is impossible since $s \equiv s_1 \cdot s_2$ is not a constant. (4A2) Via closure. Then there is some $\tau$ s.t. $\tau \to \phi \in *^0(s_1)$ and $\tau \in *^0(s_2)$. By IH, we have $\mathsf{LP}(\mathcal{CS}^{\varphi_1}) \vdash \tau \to \phi$ and $\mathsf{LP}(\mathcal{CS}^{\varphi_1}) \vdash \tau$, then $\mathsf{LP}(\mathcal{CS}^{\varphi_1}) \vdash \phi$, (i) holds. By IH, $t_1^{\xi_g} \ntriangleright \tau \to \phi$, which implies $t_1^{\xi_g} \ntriangleright \phi$, (ii) holds.

(4B) Assume $\phi \in *^{1\backslash 0}(s_1 \cdot s_2)$. (4B1) Via initial set. Then $s_1 \cdot s_2 : \phi \in X$, which implies $\phi \in X^-$ with the help of (4.1). All of (iii), (iv), (v) are ensured by (4.3). (4B2) Via closure. So there is a formula $\tau$ s.r. $\tau \to \phi \in *^1(s_1)$ and $\tau \in *^1(s_2)$. There are four subcases, but none of them are possible. (4B21) $\tau \to \phi \in *^0(s_1)$ and $\tau \in *^0(s_2)$. Then $\phi \in *^0(s_1 \cdot s_2)$ by Definition 21, a contradiction. (4B22) $\tau \to \phi \in *^0(s_1)$ and $\tau \in *^{1\backslash 0}(s_2)$. By IH, $t_1^{\xi_g} \ntriangleright \tau \to \phi$. But also by IH, $t_1^{\xi_g} \triangleright \tau$, hence $t_1^{\xi_g} \triangleright \tau \to \phi$, a contradiction. (4B23) $\tau \to \phi \in *^{1\backslash 0}(s_1)$ and $\tau \in *^0(s_2)$. By IH, $t_1^{\xi_g} \ntriangleright \tau$. Since $\tau \to \phi$ is an implication, by IH, $\tau \to \phi \equiv \delta$, so $\tau \equiv \rho\mathfrak{S}$. Note that $t_1^{\xi_g} \triangleright \rho\mathfrak{S}$; we have a contradiction. (4B24) $\tau \to \phi \in *^{1\backslash 0}(s_1)$ and $\tau \in *^{1\backslash 0}(s_2)$. By IH, we know that $\tau \notin H_g$. As in subcase (b.3), by applying IH on $\tau \to \phi$, we have $\tau \equiv \rho\mathfrak{S} \in H_g$, a contradiction. In summary, (4B2) is impossible.

(5) $s \equiv s_1 + s_2$ for some terms $s_1$ and $s_2$, while by IH (A) and (B) hold for both $s_1$ and $s_2$:

(5A) Assume $\phi \in *^0(s_1 + s_2)$. (5A1) Via initial set. Then $(s_1 + s_2) : \phi \in \mathcal{CS}^{\varphi_1}$, which is impossible since $s \equiv s_1 + s_2$ is not a constant. (5A2) Via closure. Then w.l.o.g., $\phi \in *^0(s_1)$, and hence both (i) and (ii) follow from IH.

(5B) Assume $\phi \in *^{1\backslash 0}(s_1 + s_2)$. (5B1) Via initial set. Then $s_1 + s_2 : \phi \in X$, which implies $\phi \in X^-$. From (4.3), we have all of (iii), (iv), (v). (5B2) Via closure. Then $\phi \notin *^0(s_1)$, $\phi \notin *^0(s_2)$, and w.l.o.g., $\phi \in *^1(s_1)$. Thus $\phi \in *^{1\backslash 0}(s_1)$, which, by IH, gives all of (iii), (iv), and (v). ⊣

**Claim 90** *Assume $m > 0$. For any $f \in I$, if $\mathfrak{M}^r \vDash p_f\mathfrak{S}$, then* $\mathsf{S4} \vdash \Box^{n_f}\xi_f$.

**Proof.** Since $\mathfrak{M}^r \vDash p_f\mathfrak{S}$, we have $\mathfrak{M}^r \vDash t^{\xi_f}_{n_f} : \cdots : t^{\xi_f}_1 : \xi^r_f$, which implies $\xi^r_f \in *^1(t^{\xi_f}_1)$. Here the existence of term $t^{\xi_f}_1$ is ensured by $n_f > 0$ from assumption (a) of Theorem 88.

Assume with the seek of a contradiction that $\xi^r_f \in *^{1\backslash 0}(t^{\xi_f}_1)$. Then $\xi^r_f$ cannot be a prime formula, as stated in Claim 89(B)(iv). Again by Claim 89(B)(iv), $\xi^r_f$ is not an implication, since otherwise we have $\xi^r_f \equiv \delta$ while $\delta \equiv \rho\mathfrak{S} \to \theta^r$ has $\xi^r_f$ as a proper subformula, a contradiction. Also, $\xi^r_f$ cannot have the form of $t : \psi$, since $\xi_f$ has main connective other than $\Box$ by assumption (c) of Theorem 88. In summary, $\xi^r_f$ cannot be any formula, a contradiction. Thus the assumption is false, which implies $\xi^r_f \in *^0(t^{\xi_f}_1)$.

By Claim 89(A)(i), we have $\mathsf{LP}(\mathcal{CS}^{\varphi_1}) \vdash \xi^r_f$. By Theorem 8, we see that $\mathsf{S4} \vdash \xi_f$, and then $\mathsf{S4} \vdash \Box^{n_f}\xi_f$. ⊣

We are now ready to show $\mathfrak{M}^r \nvDash \rho\mathfrak{S}$.

If $m = 0$, then $\rho$ has no atoms from $P$, which implies $\rho\mathfrak{S} \equiv \rho$. Hence $\rho\mathfrak{S} \equiv \rho$ consists of only $\bot$, $\wedge$, $\vee$ as connectives, and positive atoms that are falsified by $v$. By induction, we see that $\mathfrak{M}^r \nvDash \rho\mathfrak{S}$.

If $m > 0$, define the propositional valuation $v_{\mathfrak{M}^r}$ as follows:

$$v_{\mathfrak{M}^r}(p) = \begin{cases} 0 & \text{if } p \notin P \ ; \\ 0 & \text{if } p \equiv p_f \in P \text{ and } \mathfrak{M}^r \nvDash t^{\xi_f}_{n_f} : \cdots : t^{\xi_f}_1 : \xi^r_f \ ; \\ 1 & \text{if } p \equiv p_f \in P \text{ and } \mathfrak{M}^r \vDash t^{\xi_f}_{n_f} : \cdots : t^{\xi_f}_1 : \xi^r_f \ . \end{cases}$$

For any atom $p$, since $v(p) = 0$, we have $\mathfrak{M}^r \nvDash p$. Note that

$$p\mathfrak{S} \equiv \begin{cases} p & \text{if } p \notin P \\ t^{\xi_f}_{n_f} : \cdots : t^{\xi_f}_1 : \xi^r_f & \text{if } p \equiv p_f \in P \end{cases},$$

and hence for any atom $p$, we have $v_{\mathfrak{M}^r}(p) = 1$ iff $\mathfrak{M}^r \vDash p\mathfrak{S}$. An induction on subformulas of $\rho$ shows that $v_{\mathfrak{M}^r}(\rho) = 1$ iff $\mathfrak{M}^r \vDash \rho\mathfrak{S}$. Since $\rho$ is a propositional formula with all connectives belonging to $\{\bot, \wedge, \vee\}$, all its subformulas are positive, and for any subformula $\psi$, $v_{\mathfrak{M}^r}(\psi) = 1$ implies that $\bigwedge\{p \mid v_{\mathfrak{M}^r}(p) = 1\} \to \psi$ is a tautology.[4]

[4] By the definition of $v_{\mathfrak{M}^r}$, we see that $v_{\mathfrak{M}^r}(p) = 1$ only if $p$ belongs to the finite set $P$. Therefore, we

With the seek of a contradiction, assume $\mathfrak{M}^r \vDash \rho\mathfrak{S}$. Then $v_{\mathfrak{M}^r}(\rho) = 1$, and hence

$$\bigwedge\{p \mid v_{\mathfrak{M}^r}(p) = 1\} \to \rho$$

is a tautology. Since $v_{\mathfrak{M}^r}(p) = 0$ for any $p \notin P$, we have

$$\{p \mid v_{\mathfrak{M}^r}(p) = 1\} = \{p_f \mid f \in I, v_{\mathfrak{M}^r}(p_f) = 1\}.$$

Thus $\bigwedge\{p_f \mid f \in I, v_{\mathfrak{M}^r}(p_f) = 1\} \to \rho$ is a tautology, and hence so is

$$\bigwedge \left\{ p_f[\frac{p_e}{\Box^{n_e}\xi_e}]_{e=1}^m \mid f \in I, \mathfrak{M}^r \vDash p_f\mathfrak{S} \right\} \to \rho[\frac{p_e}{\Box^{n_e}\xi_e}]_{e=1}^m.$$

Therefore, $\mathsf{S4} \vdash \bigwedge\{\Box^{n_f}\xi_f \mid f \in I, \mathfrak{M}^r \vDash p_f\mathfrak{S}\} \to \rho[\frac{p_e}{\Box^{n_e}\xi_e}]_{e=1}^m$. By Claim 90, for any $f \in I$, if $\mathfrak{M}^r \vDash p_f\mathfrak{S}$, then $\mathsf{S4} \vdash \Box^{n_f}\xi_f$. Thus $\mathsf{S4} \vdash \rho[\frac{p_e}{\Box^{n_e}\xi_e}]_{e=1}^m$, which contradicts with assumption (d) of this theorem.

In summary, we have proved $\mathfrak{M}^r \nvDash \rho\mathfrak{S}$ in both cases. ⊣

Some interesting special cases of Theorem 88 are gathered in the following corollary.

**Corollary 91** *Let $\sigma$ be a modal formula s.t.* $\mathsf{S4} \nvdash \sigma$.

*(A) Let $\zeta$ be a propositional formula with $\bot, \wedge, \vee$ as its only connectives and let $\theta$ be a modal formula, then any realization of*

$$\Box^u(\Box^w(\Box^x\sigma \to \theta) \to \Box^z\zeta) \qquad (w, x > 0)$$

*has a counter $\mathcal{CS}^{\varphi_1}$-model.*

*(B) If* $\mathsf{S4} \vdash \Diamond\Box\sigma$ *and* $(w, x > 0)$*, then*

$$\Box^u(\Box^w(\Box^x\sigma \to \Box^y\bot) \to \Box^z\bot) \in \mathsf{S4} \setminus \mathsf{S4}^{\varphi_1}.$$

*(C) If* $\mathsf{S4} \vdash \Diamond\Box\sigma$*, then* $\Diamond\Box\sigma \notin \mathsf{S4}^{\varphi_1}$.

are not using infinite conjunction, and $\bigwedge\{p \mid v_{\mathfrak{M}^r}(p) = 1\}$ can be seen as an abbreviation of a finite formula. By the same reason, all conjunctions applied to sets in this proof are abbreviations of finite formulas. Also, $\bigwedge \varnothing \equiv \top$ by standard definition.

**Proof.** (A) As a modal formula, $\sigma$ has the form of $\Box^{x'}\sigma_0$ for some natural number $x'$, where $\sigma_0$ has main connective other than $\Box$. Take a fresh (propositional) atom $p_1$. Let $\rho \equiv p_1$, $P = \{p_1\}$, $m = 1$, $n_1 = x + x'$, and $\xi_1 \equiv \sigma_0$, we have

$$\Box^u(\Box^w(\Box^x\sigma \to \theta) \to \Box^z\zeta) \equiv \Box^u(\Box^w(\rho[\frac{p_e}{\Box^{n_e}\xi_e}]_{e=1}^m \to \theta) \to \Box^z\zeta).$$

As assumed in Theorem 88: $w > 0$ is given; $n_1 = x + x' > 0$ since $x > 0$; $\zeta$ is propositional and has only $\bot, \wedge, \vee$ as its connectives; $\theta$ is a modal formula; so is $\xi_1 \equiv \sigma_0$ which further has main connective other than $\Box$; $\mathsf{S4} \nvdash \rho[\frac{p_e}{\Box^{n_e}\xi_e}]_{e=1}^m \equiv p_1[\frac{p_1}{\Box^{x+x'}\sigma_0}] \equiv \Box^x\sigma$ since $\mathsf{S4} \nvdash \sigma$. Thus the desired counter $\mathcal{CS}^{\varphi_1}$-model is offered by Theorem 88.

(B) Since $\mathsf{S4} \vdash \Diamond\Box\sigma \equiv \Box(\Box\sigma \to \bot) \to \bot$, $\mathsf{S4} \vdash \Box(\Box\sigma \to \Box^y\bot) \to \Box^z\bot$. Since $w, x > 0$, we further have $\mathsf{S4} \vdash \Box^w(\Box^x\sigma \to \Box^y\bot) \to \Box^z\bot$, and thus $\mathsf{S4} \vdash \Box^u(\Box^w(\Box^x\sigma \to \Box^y\bot) \to \Box^z\bot)$. Let $\theta \equiv \Box^y\bot$ and $\zeta \equiv \bot$, by (A), there is a counter $\mathcal{CS}^{\varphi_1}$-model $\mathfrak{M}^r$ for any possible realization of $\Box^u(\Box^w(\Box^x\sigma \to \Box^y\bot) \to \Box^z\bot)$.

Suppose we were able to realize $\Box^u(\Box^w(\Box^x\sigma \to \Box^y\bot) \to \Box^z\bot)$ (denote the realization temporarily by $\kappa$) with a constant specification $\mathcal{CS}'$ that is not directly self-referential. That is, we have $\mathsf{LP}(\mathcal{CS}') \vdash \kappa$. By Definition 15, for any $c : A \in \mathcal{CS}'$, $c \not\triangleright A$. That is to say, $\mathcal{CS}' \subseteq \mathcal{CS}^{\varphi_1}$ and hence the proof is also an $\mathsf{LP}(\mathcal{CS}^{\varphi_1})$-proof. By Theorem 26, $\mathfrak{M} \models \kappa$ for any $\mathcal{CS}^{\varphi_1}$-model $\mathfrak{M}$, which contradicts the existence of the counter $\mathcal{CS}^{\varphi_1}$-model $\mathfrak{M}^r$. Therefore, any realization of $\Box^u(\Box^w(\Box^x\sigma \to \Box^y\bot) \to \Box^z\bot)$ calls for a directly self-referential constant specification.

(C) A trivial case of (B) where $w = x = 1$ and $u = y = z = 0$. ⊣

Corollary 91 gives a natural class of self-referential $\mathsf{S4}$-theorems that are interesting by their own. Modal theorems in $\mathsf{T}$, $\mathsf{K4}$, and $\mathsf{D4}$ that require direct self-referentiality (w.r.t. realizations in justification logics $\mathsf{JT}$, $\mathsf{J4}$, and $\mathsf{JD4}$, respectively) are given in [45] [46], where $\mathsf{T}$ and $\mathsf{D4}$ share the same instance with $\mathsf{S4}$ (i.e., formula (2.1) in Theorem 16), and the instance in $\mathsf{K4}$ is $\neg\Box\bot \to \neg\Box\neg(p \to \Box p)$, which is also a $\mathsf{T}$-theorem. Note that $\mathsf{S4} \nvdash \sigma$ implies $\mathsf{T} \nvdash \Diamond\Box\sigma$, and hence Corollary 91(C) gives a natural class of self-referential modal instances that are not $\mathsf{T}$-theorems. For example, the $\mathsf{S4}$-theorem $\Diamond\Box(\Diamond\Box p \to \Box p)$, which is not provable in $\mathsf{T}$, is not a member of $\mathsf{S4}^{\wp_1}$ (cf. Figure 3.2 in Section 3.3).

## 4.3 A Class of Directly Self-referential IPC-theorems

This Section contains materials reused from the candidate's published paper [71], with the kind permission from Elsevier, the owner of the copyright of [71]. This Section contains materials reused from the candidate's published paper [70], with the kind permission from Springer, the owner of the copyright of [70].

Self-referentiality in IPC is defined as follows:

**Definition 92 (Self-referentiality of IPC-theorems)** An IPC-theorem $\phi$ is said to be *self-referential* (*directly self-referential*), if for each (faithful) basic embedding $(\cdot)^\times$, the image $\phi^\times$ is self-referential in S4. $\dashv$

In this section we show the following theorem.

**Theorem 93 (Direct Self-referentiality of IPC)** *If $\alpha$ is a tautology such that IPC $\nvdash \alpha$, then $\neg\neg\alpha$ is a directly self-referential IPC-theorem.*

**Proof.** Glivenko's Theorem (cf. [60]) ensures IPC $\vdash \neg\neg\alpha$.

Take any basic embedding $(\cdot)^\times$. By Example 86,

$$(\neg\neg\alpha)^\times \equiv \Box^{j_+}(\Box^{j_-+k_+}(\Box^{k_-}\alpha^\times \to \Box^{i_-+l_-}\bot) \to \Box^{i_++l_+}\bot).$$

Since CPC $\vdash \alpha$, we see that $\alpha$ is not a prime formula, and hence $\alpha \equiv \alpha_1 \odot \alpha_2$ where $\odot \in \{\wedge, \vee, \to\}$. Note that $\alpha$ is a positive subformula of $\neg\neg\alpha$, so we have:

$$\alpha^\times \equiv (\alpha_1 \odot \alpha_2)^\times \equiv \Box^{j_{\odot+}}(\Box^{k_{\odot+}}\alpha_{1\pm}^\times \odot \Box^{l_{\odot+}}\alpha_2^\times),$$

where $\alpha_{1\pm}^\times$ is $\alpha_{1+}^\times$ if $\cdot \in \{\wedge, \vee\}$, and is $\alpha_{1-}^\times$ if $\cdot \in \{\to\}$. Abbreviate $\Box^{k_{\odot+}}\alpha_{1\pm}^\times \odot \Box^{l_{\odot+}}\alpha_2^\times$ by $\sigma$, we have $\alpha^\times \equiv \Box^{j_{\odot+}}\sigma$, and hence

$$(\neg\neg\alpha)^\times \equiv \Box^{j_+}(\Box^{j_-+k_+}(\Box^{j_{\odot+}+k_-}\sigma \to \Box^{i_-+l_-}\bot) \to \Box^{i_++l_+}\bot).$$

Since $\mathsf{IPC} \nvdash \alpha$, we know that $\mathsf{S4} \nvdash \alpha^{\times} \equiv \Box^{j_{\odot+}}\sigma$, which implies

$$\mathsf{S4} \nvdash \sigma. \tag{4.5}$$

Case 1, $j_{\odot+} + k_{-} > 0$. Let $u = j_{+}$, $w = j_{-} + k_{+}$, $x = j_{\odot+} + k_{-}$, $z = i_{+} + l_{+}$, $\theta \equiv \Box^{i_{-}+l_{-}}\bot$, and $\zeta \equiv \bot$. Note that from Lemma 87(B) we know that $j_{-} + k_{+} > 0$ and hence $w > 0$. Also, $j_{\odot+} + k_{-} > 0$ is given and hence $x > 0$. By Corollary 91(A), there is a counter $\mathcal{CS}^{\varphi}$-model for any realization of

$$(\neg\neg\alpha)^{\times} \equiv \Box^{j_{+}}(\Box^{j_{-}+k_{+}}(\Box^{j_{\odot+}+k_{-}}\sigma \to \Box^{i_{-}+l_{-}}\bot) \to \Box^{i_{+}+l_{+}}\bot).$$

Case 2, $j_{\odot+} = k_{-} = 0$. We have $\alpha^{\times} \equiv \sigma$ and

$$(\neg\neg\alpha)^{\times} \equiv \Box^{j_{+}}(\Box^{j_{-}+k_{+}}(\sigma \to \Box^{i_{-}+l_{-}}\bot) \to \Box^{i_{+}+l_{+}}\bot).$$

Since $k_{-} = 0$, by Lemma 87(C), $j_{+} > 0$. So $j_{+} \neq j_{\odot+}$, and hence $\odot \in \{\wedge, \vee\}$. We have a claim here.

**Claim 94** *Let $\phi$ be a positive subformula of $\alpha$ (also that of $\neg\neg\alpha$), and let $\psi$ by a negative subformula of $\phi$. Then there is an implication $\eta$ as a positive subformula of $\phi$ s.t. $\psi$ is a subformula of $\eta$.*

**Proof.** Let $\psi'$ be the maximal negative subformula of $\phi$ that has $\psi$ as a subformula. Note that $\psi'$ can be neither a conjunct (disjunct) of a conjunction (disjunction) nor the succedent of an implication, since otherwise $\psi'$ is a proper subformula of another negative subformula of $\phi$ that has $\psi$ as a subformula. So $\psi'$ can only be the antecedent of a positive implication, denoted by $\eta$. Since $\psi'$ is a negative subformula of $\phi$, we see that $\psi'$ is a proper subformula of $\phi$. Having $\psi'$ as its antecedent, $\eta$ is a subformula of $\phi$. $\dashv$

Let $\mathfrak{d}$ be a negative atom, $\wedge$, $\vee$, or $\to$ in $\sigma \equiv \alpha^{\times}$. The source of $\mathfrak{d}$ is in the scope of a negative subformula $\psi$ of $\alpha$. By Claim 94, there is an implication $\eta$ that is a positive subformula of $\alpha$ and has $\psi$ as a subformula. Since $j_+ > 0$, we know that $\eta^\times$ is a $\Box$-prefixed formula. Therefore, $\psi^\times_-$ is in the scope of a $\Box$, and so is $\mathfrak{d}$. As an intuitionistic propositional formula, $\alpha$ may have atoms and connectives $\bot$, $\wedge$, $\vee$, and $\rightarrow$. By the observation above, each negative one is in the scope of a $\Box$. Since $j_+ > 0$, each positive $\rightarrow$ is in the scope of a $\Box$. As a result, $\alpha^\times \equiv \sigma$ is a $\bot, \wedge, \vee$-combination of positive atoms and $\Box$-prefixed formulas. Let $\iota_1, ..., \iota_m$ be the list of all maximal $\Box$-prefixed subformulas of $\sigma$ where $m$ is the total number of these subformulas. (Syntactically identical subformulas, like all $\Box p$'s in $\Box p \wedge \Box p \wedge \Box p$, are distinguished here, and are all put into the list.) For each $e \in \{1, ..., m\}$, let $\xi_e$ be the maximal non-$\Box$-prefixed subformula of $\iota_e$, and hence we have $\iota_e \equiv \Box^{n_e}\xi_e$ for a positive natural number $n_e$. Let $p_1, ..., p_m$ be $m$ fresh pairwise distinct (propositional) atoms, and let $\rho$ be the result of simultaneously replacing $\iota_1, ..., \iota_m$ in $\sigma$ by $p_1, ..., p_m$ respectively. As a result, $\rho$ is a $\bot, \wedge, \vee$-combination of atoms, and each atom from $\{p_1, ..., p_m\}$ occurs in $\rho$ exactly once. Since all of $p_1, ..., p_m$ are fresh, they do not occur in any of $\xi_1, ..., \xi_m$. Hence we have

$$\sigma \equiv \rho[\frac{p_e}{\Box^{n_e}\xi_e}]_{e=1}^m,$$

and then $\mathsf{S4} \nvdash \rho[\frac{p_e}{\Box^{n_e}\xi_e}]_{e=1}^m$ by (4.5). Let $u = j_+$, $w = j_- + k_+$, $z = i_+ + l_+$, $\theta \equiv \Box^{i_- + l_-}\bot$, and $\zeta \equiv \bot$. Note that from Lemma 87(B) we know that $j_- + k_+ > 0$ and hence $w > 0$. By Theorem 88, there is a counter $\mathcal{CS}^\varphi$-model for any realization of

$$(\neg\neg\alpha)^\times \equiv \Box^{j_+}(\Box^{j_- + k_+}(\rho[\frac{p_e}{\Box^{n_e}\xi_e}]_{e=1}^m \rightarrow \Box^{i_- + l_-}\bot) \rightarrow \Box^{i_+ + l_+}\bot).$$

In summary, we have a counter $\mathcal{CS}^\varphi$-model for any realization of $(\neg\neg\alpha)^\times$ in either case. By Theorem 26, any possible realization of $(\neg\neg\alpha)^\times$ calls for a directly self-referential constant specification. $\dashv$

The set of $\mathsf{IPC}$-theorems shown to require direct self-referentiality in Theorem 93 can be briefly displayed as

$$\neg\neg(\mathsf{CPC} \setminus \mathsf{IPC}).$$

An observation on the Rieger–Nishimura Lattice lists out many instances in this set.

**Definition 95 (Rieger–Nishimura Lattice [10])** For any atom $p$, let

$$\begin{cases} f_0(p) := \bot, \\ f_1(p) := p, \\ f_{n+2}(p) := g_n(p) \vee g_{n+1}(p); \end{cases} \qquad \begin{cases} g_0(p) := p, \\ g_1(p) := \neg p \equiv p \to \bot, \\ g_{n+2}(p) := g_{n+1}(p) \to f_n(p). \end{cases}$$

⊣

**Corollary 96** *Let $p$ be an atom. If*

$$\beta \in B(p) := \{f_n(p) \mid n > 1\} \cup \{g_n(p) \mid n > 2\} \ ,$$

*then $\neg\neg\beta$ is a directly self-referential* IPC*-theorem.*

**Proof.** Observe that any formula in $B(p)$ is a tautology, but not an IPC-theorem. Then use Theorem 93. ⊣

## 4.4 Direct Self-referentiality in the Purely Implicational Fragment of IPC

This Section contains materials reused from the candidate's published paper [71], with the kind permission from Elsevier, the owner of the copyright of [71]. This Section contains materials reused from the candidate's published paper [70], with the kind permission from Springer, the owner of the copyright of [70].

Recall that in the language of intuitionistic propositional logic, all of $\bot, \wedge, \vee, \to$ are primitives. Restricting the language to have $\to$ as the only primitive in Definition 9, and deleting all axiom schemes that contains $\bot$, $\wedge$, or $\vee$, we get *the purely implicational fragment of* IPC, denoted by $\mathsf{IPC}_{\to}$.

IPC is a conservative extension of $\mathsf{IPC}_{\to}$, as presented in the following fact.

**Fact 97 ([63])** *Let $\phi$ be a propositional formula that consists of only propositional atoms and $\to$. If $\mathsf{IPC} \vdash \phi$, then $\mathsf{IPC}_{\to} \vdash \phi$.*

In Theorem 93 we have found a class of directly self-referential IPC-theorems, i.e.,

$$\neg\neg(\mathsf{IPC} \setminus \mathsf{IPC}).$$

Since $\neg$ is defined on $\bot$ and $\to$, all examples we have so far involve $\bot$. In this section, we give a directly self-referential theorem in $\mathsf{IPC}_{\to}$[5] which is a moderation of the Peirce's law.

**Theorem 98 (Direct Self-referentiality of $\mathsf{IPC}_{\to}$)** *The* IPC*-theorem*

$$((((p \to q) \to p) \to p) \to q) \to q$$

*is directly self-referential.*

[5]The question whether there is a self-referential example in $\mathsf{IPC}_{\to}$ was raised by Melvin Fitting.

**Proof.** Take any basic embedding $(\cdot)^\times$. Let $\chi \equiv ((p \to q) \to p) \to p$, then $\gamma \equiv ((((p \to q) \to p) \to p) \to q) \to q$ can be abbreviated as $(\chi \to q) \to q$. By calculation,

$$\chi^\times \equiv (((p \to q) \to p) \to p)^\times \equiv \Box^{j_+}(\Box^{k_+}((p \to q) \to p)^\times_- \to \Box^{l_+} p^\times)$$

since $\chi$ is a positive subformula of $\gamma$.

Let $\sigma \equiv \Box^{k_+}((p \to q) \to p)^\times_- \to \Box^{l_+} p^\times$, we have $\chi^\times \equiv \Box^{j_+}\sigma$, and hence by Example 86,

$$\gamma^\times \equiv ((\chi \to q) \to q)^\times \equiv \Box^{j_+}(\Box^{j_- + k_+}(\Box^{j_+ + k_-}\sigma \to \Box^{h_- + l_-} q) \to \Box^{h_+ + l_+} q). \tag{4.6}$$

Since $\mathsf{IPC} \nvdash \chi$, $\mathsf{S4} \nvdash \chi^\times \equiv \Box^{j_+}\sigma$, and hence $\mathsf{S4} \nvdash \sigma$. Note that $q$ is a propositional formula with no connectives. Let $u = j_+$, $w = j_- + k_+$, $x = j_+ + k_-$, $z = h_+ + l_+$, $\theta \equiv \Box^{h_- + l_-} q$, and $\zeta \equiv q$. Note that from Lemma 87(B) we know that $j_- + k_+ > 0$ and hence $w > 0$. Also, from Lemma 87(C) we know that $j_+ + k_- > 0$ and hence $x > 0$. By Corollary 91(A), any realization of $\Box^{j_+}(\Box^{j_- + k_+}(\Box^{j_+ + k_-}\sigma \to \Box^{h_- + l_-} q) \to \Box^{h_+ + l_+} q)$ has a counter $\mathcal{CS}^{\varphi}$-model. By (4.6), any realization of $\gamma^\times$ has a counter $\mathcal{CS}^{\varphi_1}$-model. By Theorem 26, any realization of $\gamma^\times$ calls for a directly self-referential constant specification. ⊣

Theorem 98 indicates that the BHK semantics of intuitionistic implication involves direct self-referentiality.

At the end of Section 4.1, we have discussed the generality of the basic embedding. Once $\bot$ and $\Box$ are chosen as primitives, the class of basic embeddings is then natural and general. This section takes care of the purely implicational fragment, and hence pushes our result in to a border sense where $\bot$ is not necessarily primitive.

## 4.5 Fragments of IPC

In Chapter 3, we show that both $\mathsf{X}^{\otimes}$, $\mathsf{X}^{\not\circlearrowleft}$ and $\mathsf{X}^{\not\circlearrowleft_1}$ are not closed under $MP$ for each $\mathsf{X} \in \{\mathsf{T}, \mathsf{K4}, \mathsf{S4}\}$. Given a basic embedding $(\cdot)^{\times}$ as a parameter, we can define similar fragments of IPC based on that of S4. In this section, after giving definitions of fragments, we will restrict ourselves to the $\mathsf{IPC}_{\rightarrow}$, show that each axiom there, via any given basic embedding, is loop-free provable in S4, hence non-self-referential realizable in LP. Combined with our results in Section 4.4, we will see that none of these fragments of $\mathsf{IPC}_{\rightarrow}$ are closed under $MP$. Similar results for IPC will then follow from conservativity .

We start with definitions of fragments.

**Definition 99 (Loop-free Provable Fragment of IPC and $\mathsf{IPC}_{\rightarrow}$)** Let $(\cdot)^{\times}$ be a basic embedding.

(1) The *loop-free provable fragment of* IPC *via* $(\cdot)^{\times}$, denoted by $\mathsf{IPC}^{\otimes(\times)}$, is defined as

$$\mathsf{IPC}^{\otimes(\times)} := \{\mathsf{IPC} \vdash \phi \mid \phi^{\times} \in \mathsf{S4}^{\otimes}\}.$$

(2) The *loop-free provable fragment of* $\mathsf{IPC}_{\rightarrow}$ *via* $(\cdot)^{\times}$ is defined as

$$\mathsf{IPC}_{\rightarrow}^{\otimes(\times)} := \{\mathsf{IPC}_{\rightarrow} \vdash \phi \mid \phi^{\times} \in \mathsf{S4}^{\otimes}\}.$$

⊣

**Definition 100 (Non-self-referential Fragment of IPC and $\mathsf{IPC}_{\rightarrow}$)** Let $(\cdot)^{\times}$ be a basic embedding.

(1) The *non-self-referential fragment of* IPC *via* $(\cdot)^{\times}$, denoted by $\mathsf{IPC}^{\not\circlearrowleft(\times)}$, is defined as

$$\mathsf{IPC}^{\not\circlearrowleft(\times)} := \{\mathsf{IPC} \vdash \phi \mid \phi^{\times} \in \mathsf{S4}^{\not\circlearrowleft}\}.$$

(2) The *non-self-referential fragment of* $\mathsf{IPC}_{\rightarrow}$ *via* $(\cdot)^{\times}$ is defined as

$$\mathsf{IPC}_{\rightarrow}^{\not\circlearrowleft(\times)} := \{\mathsf{IPC}_{\rightarrow} \vdash \phi \mid \phi^{\times} \in \mathsf{S4}^{\not\circlearrowleft}\}.$$

(3) Similar definitions for the non-directly-self-referential fragments, for which we take $\wp_1$ instead of $\wp$ for notations. $\dashv$

The following theorem is quite obvious.

**Theorem 101** *If* $(\cdot)^\times$ *is a basic embedding, then:*

*(A)* $\mathsf{IPC}^{\otimes(\times)} \subseteq \mathsf{IPC}^{\wp(\times)} \subseteq \mathsf{IPC}^{\wp_1(\times)}$*;*

*(B)* $\mathsf{IPC}_{\rightarrow}^{\otimes(\times)} \subseteq \mathsf{IPC}_{\rightarrow}^{\wp(\times)} \subseteq \mathsf{IPC}_{\rightarrow}^{\wp_1(\times)}$.

**Proof.** (A) Assume $\phi \in \mathsf{IPC}^{\otimes(\times)}$. Then $\phi^\times \in \mathsf{S4}^{\otimes}$. By Corollary 71, $\mathsf{S4}^{\otimes} \subseteq \mathsf{S4}^{\wp}$, hence $\phi^\times \in \mathsf{S4}^{\wp}$ which implies $\phi \in \mathsf{IPC}^{\wp(\times)}$.

Assume $\phi \in \mathsf{IPC}^{\wp(\times)}$. Then $\phi^\times \in \mathsf{S4}^{\wp}$. By Theorem 70, $\mathsf{S4}^{\wp} \subseteq \mathsf{S4}^{\wp_1}$, hence $\phi^\times \in \mathsf{S4}^{\wp_1}$ which implies $\phi \in \mathsf{IPC}^{\wp_1(\times)}$.

(B) Similar to (A). $\dashv$

In our current notation, Theorem 98 can be reformulated as follows:

**Corollary 102** *Denote the* $\mathsf{IPC}$*-theorem* $((((p \rightarrow q) \rightarrow p) \rightarrow p) \rightarrow q) \rightarrow q$ *by* $\gamma$*. For any basic embedding* $(\cdot)^\times$*:*

*(A) (i)* $\gamma \notin \mathsf{IPC}_{\rightarrow}^{\wp_1(\times)}$*, (ii)* $\gamma \notin \mathsf{IPC}_{\rightarrow}^{\wp(\times)}$*, and (iii)* $\gamma \notin \mathsf{IPC}_{\rightarrow}^{\otimes(\times)}$*;*

*(B) (i)* $\gamma \notin \mathsf{IPC}^{\wp_1(\times)}$*, (ii)*$\gamma \notin \mathsf{IPC}^{\wp(\times)}$*, and (iii)* $\gamma \notin \mathsf{IPC}^{\otimes(\times)}$.

**Proof.** (A) (i) directly follows from Theorem 98, then (ii) and (iii) follow from Theorem 101(B).

(B) (i) Suppose with the seek of a contradiction that $\gamma \in \mathsf{IPC}^{\wp_1(\times)}$. By Definition 100, $\mathsf{IPC} \vdash \gamma$ and $\gamma^\times \in \mathsf{S4}^{\wp_1}$. Since $\gamma$ has $\rightarrow$ as its only connective, by Fact 97, $\mathsf{IPC}_{\rightarrow} \vdash \gamma$, and hence $\gamma \in \mathsf{IPC}_{\rightarrow}^{\wp_1(\times)}$, which contradicts (A)(i).

By Theorem 101(A), we further have (ii) and (iii). $\dashv$

We now concentrate our attention to $\mathsf{IPC}_{\to}$, and show that all axioms of it are included in $\mathsf{IPC}_{\to}^{\otimes(\times)}$, for any unpolarized basic embedding $(\cdot)^{\times}$.

**Theorem 103** *For any* $\mathsf{IPC}_{\to}$*-formulas* $\phi, \psi, \chi$ *and unpolarized basic embedding* $(\cdot)^{\times}$*:*

*(A)* $(\phi \to (\psi \to \phi) \in \mathsf{IPC}_{\to}^{\otimes(\times)}$*;*

*(B)* $(\phi \to (\psi \to \chi)) \to ((\phi \to \psi) \to (\phi \to \chi)) \in \mathsf{IPC}_{\to}^{\otimes(\times)}$.

**Proof.** Fact 56 in Section 3.1 will be used many times throughout the proof in order to show that certain applications of rules will not introduce a prehistoric loop. To avoid repetition, we assume that the reader will keep Fact 56 in mind, and refer to it once he is not sure about certain rules.

(A) It is sufficient to show that $\mathsf{G3s}^{\otimes} \vdash \Rightarrow (\phi \to (\psi \to \phi))^{\times}$.

By calculation,

$$(\phi \to (\psi \to \phi))^{\times} \equiv \Box^{j}(\Box^{k}\phi^{\times} \to \Box^{j+l}(\Box^{k}\psi^{\times} \to \Box^{l}\phi^{\times})).$$

If we have $\mathsf{G3s}^{\otimes} \vdash \Box^{k}\phi^{\times} \Rightarrow \Box^{j+l}(\Box^{k}\psi^{\times} \to \Box^{l}\phi^{\times})$, then what follows is the desired loop-free proof:

$$\begin{array}{rcll}
\Box^{k}\phi^{\times} & \Rightarrow & \Box^{j+l}(\Box^{k}\psi^{\times} \to \Box^{l}\phi^{\times}) & \text{assumed to be loop-free} \\
 & \Rightarrow & \Box^{k}\phi^{\times} \to \Box^{j+l}(\Box^{k}\psi^{\times} \to \Box^{l}\phi^{\times}) & (R\to) \\
 & \Rightarrow & \Box^{j}(\Box^{k}\phi^{\times} \to \Box^{j+l}(\Box^{k}\psi^{\times} \to \Box^{l}\phi^{\times})) & (\overrightarrow{R\Box}).
\end{array}$$

Thus, it is sufficient to show $\mathsf{G3s}^{\otimes} \vdash \Box^{k}\phi^{\times} \Rightarrow \Box^{j+l}(\Box^{k}\psi^{\times} \to \Box^{l}\phi^{\times})$.

(1) If $k > 0$, we can take the following:

$$\begin{array}{rcll}
\phi^{\times} & \Rightarrow & \phi^{\times} & \text{loop-free by Lemma 57} \\
\Box^{k}\phi^{\times} & \Rightarrow & \phi^{\times} & (\overrightarrow{L\Box W}) \\
\Box^{k}\psi^{\times}, \Box^{k}\phi^{\times} & \Rightarrow & \Box^{l}\phi^{\times} & (\overrightarrow{R\Box}) \\
\Box^{k}\phi^{\times} & \Rightarrow & \Box^{k}\psi^{\times} \to \Box^{l}\phi^{\times} & (R\to) \\
\Box^{k}\phi^{\times} & \Rightarrow & \Box^{j+l}(\Box^{k}\psi^{\times} \to \Box^{l}\phi^{\times}) & (\overrightarrow{R\Box}),
\end{array}$$

where $(R\Box)$ is applicable since $k > 0$.

(2) Otherwise, $k = 0$, and hence $k_+ = 0$. By Lemma 87, $h_- > 0$ and $j_- > 0$, which implies $h > 0$ and $j > 0$ in unpolarized case.

Now, for an atom $p$, we have $p^\times \equiv \Box^h p$ being $\Box$-prefixed. Also, for any formulas $\alpha$ and $\beta$, we have $(\alpha \to \beta)^\times \equiv \Box^j(\Box^k\alpha^\times \to \Box^l\beta^\times)$ being $\Box$-prefixed. This forms an induction showing that $\theta^\times$ is $\Box$-prefixed for any $\theta$.

Recall that it is sufficient to show $\mathsf{G3s}^\otimes \vdash \Box^k\phi^\times \Rightarrow \Box^{j+l}(\Box^k\psi^\times \to \Box^l\phi^\times)$, which is $\mathsf{G3s}^\otimes \vdash \phi^\times \Rightarrow \Box^{j+l}(\psi^\times \to \Box^l\phi^\times)$ since $k = 0$. The following loop-proof works:

$$\begin{array}{rcll} \phi^\times & \Rightarrow & \phi^\times & \text{loop-free by Lemma 57} \\ \psi^\times, \phi^\times & \Rightarrow & \Box^l\phi^\times & (\overrightarrow{R\Box}) \\ \phi^\times & \Rightarrow & \psi^\times \to \Box^l\phi^\times & (R\to) \\ \phi^\times & \Rightarrow & \Box^{j+l}(\psi^\times \to \Box^l\phi^\times) & (\overrightarrow{R\Box}), \end{array}$$

where $(R\Box)$ is applicable since $\phi^\times$ is $\Box$-prefixed.

(B) It is sufficient to show that $\mathsf{G3s}^\otimes \vdash \Rightarrow ((\phi \to (\psi \to \chi)) \to ((\phi \to \psi) \to (\phi \to \chi)))^\times$. By calculation,

$$\begin{array}{l} ((\phi \to (\psi \to \chi)) \to ((\phi \to \psi) \to (\phi \to \chi)))^\times \equiv \\ \Box^j(\Box^{j+k}(\Box^k\phi^\times \to \Box^{j+l}(\Box^k\psi^\times \to \Box^l\chi^\times)) \to \Box^{j+l}(\Box^{j+k}(\Box^k\phi^\times \to \Box^l\psi^\times) \to \Box^{j+l}(\Box^k\phi^\times \to \Box^l\chi^\times))). \end{array}$$

Taking fresh propositional atoms $p$, $q$, and $r$, we construct a loop-free proof in the following steps.

Step 1. Both

$$\begin{array}{rcll} \langle\Box^k p\rangle, q & \Rightarrow & q & (Ax) \\ \langle\Box^k p\rangle, \Box^l q & \Rightarrow & q & (\overrightarrow{L\Box W}) \end{array}$$

and

$$\Box^k p \quad \Rightarrow \quad \Box^k p, \langle q\rangle \quad \text{loop-free by Lemma 57}$$

are loop-free. Continuing on these two conclusions by

$$\dfrac{\dfrac{\langle\Box^k p\rangle, \Box^l q \Rightarrow q \qquad \Box^k p \Rightarrow \Box^k p, \langle q\rangle}{\Box^k p \to \Box^l q, \Box^k p \Rightarrow q}(L\to)}{\Box^{j+k}(\Box^k p \to \Box^l q), \Box^k p \Rightarrow q}(\overrightarrow{L\Box W})$$

we have a proof that is also loop-free, since for $(L\rightarrow)$, among each pair of unified formulas, there is one being isolated.

Step 2. Case (1) If $k > 0$, then we continue our construction as follows:

$$\begin{array}{lcll} \Box^{j+k}(\Box^k p \rightarrow \Box^l q), \Box^k p & \Rightarrow & q & \text{as in Step 1} \\ \Box^{j+k}(\Box^k p \rightarrow \Box^l q), \Box^k p & \Rightarrow & \Box^k q, \langle r \rangle & (\overrightarrow{R\Box}), \end{array}$$

where $(R\Box)$ is applicable since $k > 0$ by assumption. Case (2) Otherwise, $k = 0$. In this case, $\Box^k q \equiv q$, and hence what we have constructed by Step 1 is a loop-free proof of $\Box^{j+k}(\Box^k p \rightarrow \Box^l q), \Box^k p \Rightarrow \Box^k q$. We continue by

$$\begin{array}{lcll} \Box^{j+k}(\Box^k p \rightarrow \Box^l q), \Box^k p & \Rightarrow & \Box^k q & \text{as in Step 1} \\ \Box^{j+k}(\Box^k p \rightarrow \Box^l q), \Box^k p & \Rightarrow & \Box^k q, \langle r \rangle & (W), \end{array}$$

which is loop-free. In both Case (1) and Case (2), we see that

$$\mathsf{G3s}^{\otimes} \vdash \Box^{j+k}(\Box^k p \rightarrow \Box^l q), \Box^k p \Rightarrow \Box^k q, \langle r \rangle.$$

Step 3. Observe that

$$\langle \Box^{j+k}(\Box^k p \rightarrow \Box^l q) \rangle, \langle \Box^k p \rangle, r \quad \Rightarrow \quad r \quad (Ax)$$

is loop-free, and we can combine it with what we have by Step 2 like follows:

$$\dfrac{\dfrac{\langle \Box^{j+k}(\Box^k p \rightarrow \Box^l q) \rangle, \langle \Box^k p \rangle, r \Rightarrow r \qquad \Box^{j+k}(\Box^k p \rightarrow \Box^l q), \Box^k p \Rightarrow \Box^k q, \langle r \rangle}{\Box^k q \rightarrow r, \Box^{j+k}(\Box^k p \rightarrow \Box^l q), \Box^k p \Rightarrow r}(L\rightarrow)}{\Box^{j+l}(\Box^k q \rightarrow r), \Box^{j+k}(\Box^k p \rightarrow \Box^l q), \Box^k p \Rightarrow r}(\overrightarrow{L\Box W})$$

Since among each pair of unified formulas, there is one being isolated, $(L\rightarrow)$ does not introduce a loop here. Thus,

$$\mathsf{G3s}^{\otimes} \vdash \Box^{j+l}(\Box^k q \rightarrow r), \Box^{j+k}(\Box^k p \rightarrow \Box^l q), \Box^k p \Rightarrow r.$$

Step 4. Observe that

$$\langle \Box^{j+k}(\Box^k p \rightarrow \Box^l q) \rangle, \Box^k p \quad \Rightarrow \quad \Box^k p, \langle r \rangle \quad \text{loop-free by Lemma 57}$$

is loop-free. Combining it with what we constructed by Step 3, we have

$$\dfrac{\Box^{j+l}(\Box^k q\to r),\Box^{j+k}(\Box^k p\to\Box^l q),\Box^k p\Rightarrow r\qquad \langle\Box^{j+k}(\Box^k p\to\Box^l q)\rangle,\Box^k p\Rightarrow\Box^k p,\langle r\rangle}{\Box^k p\to\Box^{j+l}(\Box^k q\to r),\Box^{j+k}(\Box^k p\to\Box^l q),\Box^k p\Rightarrow r}(L\to)$$

which is loop-free. Note that other than reasons mentioned above, there is a pair of non-isolated $\Box^k p$'s from antecedents being unified. Since $\Box^k p$'s in antecedents contain no positive $\Box$'s, there are no vertices from these formulas in prehistoric graphs. Therefore, this unification cannot make a loop.

Step 5. Continue the construction like follows:

$$\begin{array}{rcll}
\Box^k p\to\Box^{j+l}(\Box^k q\to r),\Box^{j+k}(\Box^k p\to\Box^l q),\Box^k p & \Rightarrow & r & \text{as in Step 4}\\
\Box^{j+k}(\Box^k p\to\Box^{j+l}(\Box^k q\to r)),\Box^{j+k}(\Box^k p\to\Box^l q),\Box^k p & \Rightarrow & r & (\overrightarrow{L\Box W})\\
\Box^{j+k}(\Box^k p\to\Box^{j+l}(\Box^k q\to r)),\Box^{j+k}(\Box^k p\to\Box^l q) & \Rightarrow & \Box^k p\to r & (R\to)\\
\Box^{j+k}(\Box^k p\to\Box^{j+l}(\Box^k q\to r)),\Box^{j+k}(\Box^k p\to\Box^l q) & \Rightarrow & \Box^{j+l}(\Box^k p\to r) & (\overrightarrow{R\Box})
\end{array}$$

$$\begin{array}{rcll}
\Box^{j+k}(\Box^k p\to\Box^{j+l}(\Box^k q\to r)) & \Rightarrow & \Box^{j+k}(\Box^k p\to\Box^l q)\to\Box^{j+l}(\Box^k p\to r) & (R\to)\\
\Box^{j+k}(\Box^k p\to\Box^{j+l}(\Box^k q\to r)) & \Rightarrow & \Box^{j+l}(\Box^{j+k}(\Box^k p\to\Box^l q)\to\Box^{j+l}(\Box^k p\to r)) & (\overrightarrow{R\Box})
\end{array}$$

$$\begin{array}{rll}
\Rightarrow & \Box^{j+k}(\Box^k p\to\Box^{j+l}(\Box^k q\to r))\to\Box^{j+l}(\Box^{j+k}(\Box^k p\to\Box^l q)\to\Box^{j+l}(\Box^k p\to r)) & (R\to)\\
\Rightarrow & \Box^{j}(\Box^{j+k}(\Box^k p\to\Box^{j+l}(\Box^k q\to r))\to\Box^{j+l}(\Box^{j+k}(\Box^k p\to\Box^l q)\to\Box^{j+l}(\Box^k p\to r))) & (\overrightarrow{R\Box})
\end{array}$$

where $(R\Box)$ is applicable since $j+k>0$ (by Lemma 87). This gives us

$$\mathsf{G3s}^{\otimes}\vdash\Box^{j}(\Box^{j+k}(\Box^k p\to\Box^{j+l}(\Box^k q\to r))\to\Box^{j+l}(\Box^{j+k}(\Box^k p\to\Box^l q)\to\Box^{j+l}(\Box^k p\to r))).$$

Step 6. By Theorem 58, we have

$$\mathsf{G3s}^{\otimes}\vdash\left(\Box^j\left(\begin{array}{l}\Box^{j+k}(\Box^k p\to\Box^{j+l}(\Box^k q\to r))\to\\ \Box^{j+l}(\Box^{j+k}(\Box^k p\to\Box^l q)\to\Box^{j+l}(\Box^k p\to r))\end{array}\right)\right)[p/\phi^\times][q/\psi^\times][r/\Box^l\chi^\times].$$

We are done since

$$\begin{array}{ll}
& \left(\Box^j\left(\begin{array}{l}\Box^{j+k}(\Box^k p\to\Box^{j+l}(\Box^k q\to r))\to\\ \Box^{j+l}(\Box^{j+k}(\Box^k p\to\Box^l q)\to\Box^{j+l}(\Box^k p\to r))\end{array}\right)\right)[p/\phi^\times][q/\psi^\times][r/\Box^l\chi^\times]\\
\equiv & \Box^j\left(\begin{array}{l}\Box^{j+k}(\Box^k\phi^\times\to\Box^{j+l}(\Box^k\psi^\times\to\Box^l\chi^\times))\to\\ \Box^{j+l}(\Box^{j+k}(\Box^k\phi^\times\to\Box^l\psi^\times)\to\Box^{j+l}(\Box^k\phi^\times\to\Box^l\chi^\times))\end{array}\right).
\end{array}$$

$\dashv$

**Corollary 104** *For any* $\mathsf{IPC}_{\to}$*-formulas* $\phi, \psi, \chi$ *and unpolarized basic embedding* $(\cdot)^{\times}$*:*

*(A)* $\phi \to (\psi \to \phi) \in \mathsf{IPC}_{\to}^{\wp(\times)}$*;*

*(B)* $(\phi \to (\psi \to \chi)) \to ((\phi \to \psi) \to (\phi \to \chi)) \in \mathsf{IPC}_{\to}^{\wp(\times)}$*.*

*(C)* $\phi \to (\psi \to \phi) \in \mathsf{IPC}_{\to}^{\wp_1(\times)}$*;*

*(D)* $(\phi \to (\psi \to \chi)) \to ((\phi \to \psi) \to (\phi \to \chi)) \in \mathsf{IPC}_{\to}^{\wp_1(\times)}$*.*

**Proof.** By Theorem 101(B), $\mathsf{IPC}_{\to}^{\otimes(\times)} \subseteq \mathsf{IPC}_{\to}^{\wp(\times)} \subseteq \mathsf{IPC}_{\to}^{\wp_1(\times)}$. Hence (A) and (C) follow from Theorem 103(A), while (B) and (D) follow from Theorem 103(B). ⊣

Now we are ready to present our last theorem in this thesis.

**Theorem 105 (Non-admissibility of** $MP$**)** *For any unpolarized basic embedding* $(\cdot)^{\times}$*:*

*(A) Each of* $\mathsf{IPC}_{\to}^{\otimes(\times)}$*,* $\mathsf{IPC}_{\to}^{\wp(\times)}$*, and* $\mathsf{IPC}_{\to}^{\wp_1(\times)}$ *is not closed under* $MP$*;*

*(B) Each of* $\mathsf{IPC}^{\otimes(\times)}$*,* $\mathsf{IPC}^{\wp(\times)}$*, and* $\mathsf{IPC}^{\wp_1(\times)}$ *is not closed under* $MP$*.*

**Proof.** (A) By Theorem 103 and Corollary 104, we see that all axioms of $\mathsf{IPC}_{\to}$ are included in each of these three fragments. If any of the three fragments is closed under $MP$, then that fragment will include all $\mathsf{IPC}_{\to}$-theorems, which contradicts with Corollary 102(A).

(B) We prove for $\mathsf{IPC}^{\wp_1(\times)}$, and the proof for the other two is quite similar.

Denote $((((p \to q) \to p) \to p) \to q) \to q$ by $\gamma$.

By Theorem 103 and Corollary 104, we see that each axiom of $\mathsf{IPC}_{\to}$ is included in $\mathsf{IPC}_{\to}^{\wp_1(\times)}$. By Definition 100, we see that $\mathsf{IPC}_{\to}^{\wp_1(\times)} \subseteq \mathsf{IPC}^{\wp_1(\times)}$ and hence each axiom of $\mathsf{IPC}_{\to}$ is included in $\mathsf{IPC}^{\wp_1(\times)}$. If $\mathsf{IPC}^{\wp_1(\times)}$ were closed under $MP$, then it would contain any theorem in $\mathsf{IPC}_{\to}$, including $\gamma$, which contradicts Corollary 102(B). ⊣

Recall that in Section 3.2 we show that non-self-referential fragments of modal logics $\mathsf{T}$, $\mathsf{K4}$, and $\mathsf{S4}$ are not closed under $MP$. In this section we get similar results for $\mathsf{IPC}$

and $\mathsf{IPC}_{\rightarrow}$. All these indicates the complexity of self-referentiality, which is explicitly displayed in justification logic, but very hard to be captured in more implicit layers.

# Chapter 5

# Suggested Future Works

There are many interesting open questions related to this thesis. We list some of them in the order from conceptional to technical.

In Chapter 4, we show that the BHK semantics of intuitionistic logic (specifically, of intuitionistic implication) is intrinsically self-referential. Recall that the BHK stipulation for implication says "a proof of $\phi \to \psi$ is a construction that returns a proof of $\psi$ whenever a proof of $\phi$ is given." This stipulation has the notion of "any" applied to the class of "proofs" that is being defined by the stipulation itself. Therefore, this stipulation is *impredicative*. As suggested by Artemov and Kurokawa during a discussion, self-referentiality is likely the cost we pay for the impredicativity of BHK in this framework.

The notion of prehistoric graph presented in Chapter 2 was employed as a tool in the research of self-referentiality. However, this notion is interesting by its own, since it naturally capture dependence relation between modalities in a sequent proof tree. It will not be surprising if this idea is used in researches about substructural logic or theorem proving.

The notion of basic embedding presented in Chapter 4 receives relatively fewer attention in this thesis. We gave several necessary conditions on parameters for an basic embedding to be faithful in terms of mapping IPC into S4. Some sub/super-intuitionistic logics can also be embedded into modal logics, usually by Gödel's "$\Box$-prefixing each sub-formula" embedding, instanced by [64] and [17]. There are also similar works on the substructural perspective like [19]. It is interesting to explorer the power of basic embedding in those pairs of logics, and even in the proof of faithfulness.

In [64], Visser established the basic propositional logic, and gave there an embedding (which is basic in our terminology) that maps it faithfully into K4. As verified by Kuznets, K4 has self-referential theorems, we can ask the question that whether self-referentiality is intrinsic in Visser's logic.

Since Kuznets [14], which gave the first example in S4 that requires self-referentiality, a criterion for self-referentiality has been desired. Now more self-referential theorems in S4 have been found but a criterion is still not known, and there is no answer for basic questions like whether self-referentiality is decidable in modal logic or not. We have seen in Chapter 2 that prehistoric-loop-free provability implies non-self-referential realizability. This is only an one-direction result and hence does not offer a complete criterion. So far, all self-referential and non-self-referential modal theorems found coincide with the conjecture that the other direction also holds.

Related to the question above, another basic question that is still open is, whether or not there is a modal theorem that is self-referential, but not directly self-referential. Our best approximation so far of non-direct-self-referentiality is still bounded by non-self-referentiality, whereas all self-referential instances we have are directly self-referential.

Taking the possible difference between self-referentiality and direct-self-referentiality into account, figures like those at the end of Section 3.3 may look more complex, and may need more instances to be completed.

The family of justification logics has many members and is still growing. While all of our attentions are paid to modal/justification logics smaller than or equal to S4/LP, there are also many instances larger than this pair, some have other operators in terms, and/or need different methods to establish meta-theorems like realization. Since Kuznets' first self-referential instance for S4, it has long been assumed that some "thing" in S4 is the cause for that, and any logic larger than S4 is automatically self-referential, as they also have that "thing." This is perhaps the reason why most attentions, including that of the candidate's, have been drawn to logics smaller than S4. However, as we verified in Section 3.3, a modal formula can be self-referential in one logic while being non-self-referential in a larger logic. It is therefore quite reasonable to turn back, checking whether our self-referential instances for S4 also work for larger logics.

# Bibliography


[1] Sergei N. Artemov: Operational modal logic. Technical Report MSI 95-29. Cornell University (1995)

[2] Sergei N. Artemov: Explicit provability and constructive semantics. *The Bulletin of Symbolic Logic* 7(1), 1-36 (2001)

[3] Sergei N. Artemov: Symmetric logic of proofs. In: Arnon Avron, et al. (eds.): *Trakhtenbrot/Festschrift*. LNCS 4800, 58-71. Springer (2008)

[4] Sergei N. Artemov: The logic of justification. *The Review of Symbolic Logic* 1(4), 477-513 (2008)

[5] Sergei N. Artemov: The ontology of justifications in the logical setting. Technical Report 2011008, Ph.D. Program in Computer Science, the City University of New York (2011)

[6] Sergei N. Artemov, Lev D. Beklemishev: Provability logic. In: Dov M. Gabbay, Franz Guenthner (eds.): *Handbook of Philosophical Logic* v.10. Kluwer Academic Publisher (2002)

[7] Sergei N. Artemov, Tyko Stra*ß*en: The logic of the Gödel proof predicate. In: Georg Gottlob, Alexander Leitsch, Daniele Mundici (eds.): *Computational Logic and Proof Theory: KGC'93 Brno*. LNCS 713, 71-82. Springer (1993)

[8] Sergei N. Artemov, Tatiana Yavorskaya (Sidon): On first-order logic of proofs. *Moscow Mathematical Journal* 1, 475-490 (2001)

[9] Sergei N. Artemov, Tatiana Yavorskaya (Sidon): First-order logic of proofs. Technical Report TR-2011005, CUNY Ph.D. Program in Computer Science (2011)

[10] Nick Bezhanishvili, Dick de Jongh: Intuitionistic logic. ESSLLI'05 Course Notes (2005)

[11] Patrick Blackburn, Maarten de Rijke, Yde Venema: *Model Logic*, Cambridge University Press, Cambridge (2001)

[12] George Boolos: *The Logic of Provability*, Cambridge University Press, Cambridge (1993)

[13] Vladimir Brezhnev: On explicit counterparts of modal logics. Technical Report CFIS 2000-05, Cornell University (2000)

[14] Vladimir Brezhnev, Roman Kuznets: Making knowledge explicit: how hard it is. *Theoretical Computer Science* 357(1-3), 23-34 (2006)

[15] Luitzen Egbertus Jan Brouwer: *Collected Works: Philosophy and Foundations of Mathematics*. Arend Heyting (ed.), v.1. North Holland. (1975)

[16] Kai Brunnler, Remo Goetschi, Roman Kuznets: A syntactic realization theorem for justification logics. *Advances in Modal Logic* 8, 39-58 (2010)

[17] Alexander Chagrov, Michael Zakharyashche: Modal companions of intermediate propositional logics. *Studia Logica* 51, 49-82 (1992)

[18] Alexander Chagrov, Michael Zakharyashche: *Modal Logic*, Oxford Secience Publications, Oxford (1997)

[19] Kosta Došen: Modal translations in substructural logics. *The Journal of Philosophical Logic* 21, 283-336 (1992)

[20] Solomon Feferman, et al (eds.): *Kurt Gödel Collected Works*, vol. 3. Oxford University Press, Oxford, (1995)

[21] David Fernandez: A polynomial translation of $\mathsf{S4}$ into intuitionistic logic. *The Journal of Symbolic Logic* 71(3), 989-1001 (2006)

[22] Melvin Fitting: *Proof Methods for Modal and Intuitionistic Logics*. Reidel, Dordrecht (1983)

[23] Melvin Fitting: The logic of proofs, semantically. *Annals of Pure and Applied Logic* 132(1), 1-25 (2005)

[24] Melvin Fitting: Realizations and $\mathsf{LP}$. *Annals of Pure and Applied Logic* 161(3), 368-387 (2009)

[25] Melvin Fitting: Realization implemented. Technical Report TR-2013005, CUNY PhD Program in Computer Science (2013)

[26] Melvin Fitting: Justification Logics and Realization. Technical Report TR-2014004, CUNY PhD Program in Computer Science (2014)

[27] Gerhard Gentzen: Unstersuchungen über das logische Schliessen. *Mathematische Zeitschrift* 39, 176-210, 405-431 (1935) English translation in: Manfred Egon Szabo (ed.): *The Collected Papers of Gerhard Gentzen*. North-Holland, Amsterdam (1969)

[28] Jean-Yves Girard: Linear logic. *Theoretical Computer Science* 50, 1-102 (1987)

[29] Valery Ivanovich Glivenko: On some points of the logic of Mr. Brouwer (English translation). In: [48].

[30] Kurt Gödel: Eine Interpretation des intuitionistischen Aussagenkalkuls. *Ergebnisse eines mathematischen Kolloquiums* 4, 39-40, (1933) English translation in: Solomon Feferman, et al (eds.): *Kurt Gödel Collected Works*, vol. 1, 301-303. Oxford University Press, Oxford, Clarendon Press, New York (1986)

[31] Kurt Gödel: Lecture at Zilsel's. In: [20], 86-113.

[32] Remo Goetschi, Roman Kuznets: Realization for justification logics via nested sequents: Modularity through embedding. *Annals of Pure and Applied Logic* 163(9), 1271-1298 (2012)

[33] Nicolas D. Goodman: A theory of constructions equivalent to arithmetic. In: John Myhill, et al (eds.): *Intuitionism and Proof Theory*, 101-120. North-Holland, Amsterdam (1970)

[34] Arend Heyting: *Intuitionism. An Introduction*, 3rd ed. Studies in Logic and the Foundations of Mathematics. North-Holland (1971)

[35] Arend Heyting: The formal rules of intuitionistic logic (English translation). In: [48].

[36] Katsumasa Ishii, Ryo Kashima, Kentaro Kukuchi: Sequent calculi for Visser's propositional logics. *Notre Dame Journal of Formal Logic* 42(1), 1-22 (2001)

[37] Stig Kanger: Provability in logic. *Stockholm Studies in Philosophy* 1, Almquivst and Wiksell, Stockholm (1957)

[38] Kentaro Kikuchi, Katsumi Sasaki: A cut-free Gentzen formulation of basic propositional calculus. *Journal of Logic, Language and Information* 12, 213-225 (2003)

[39] Stephen Cole Kleene: On the interpretation of intuitionistic number theory. *Journal of Symbolic Logic* 10, 109-124 (1945)

[40] Stephen Cole Kleene: *Introduction to Metamathematics*. Van Nostrand, Princeton (1952)

[41] Andrey N. Kolmogorov: On the principle of exclude middle. (English translation) In: Jean van Heijenoort (ed.) *From Frege to Gödel. A Source Book in Mathematical Logic, 1879-1931*. Harvard University Press (1967)

[42] Andrey N. Kolmogorov: On the interpretation of intuitionistic logic. In: [48].

[43] Georg Kreisel: Foundations of intuitionistic logic. In: Ernest Nagel, et al (eds.): *Logic, Methodology and Philosophy of Science*, 198-210. Stanford University Press, Stanford (1962)

[44] Hidenori Kurokawa, Hirohiko Kushida: Substructural logic of proofs. In: Leonid Libkin, et al (eds.): *WoLLIC 2013*. LNCS 8071, 194-210. Springer (2013)

[45] Roman Kuznets: Self-referentiality of justified knowledge. In: Edward A. Hirsch, Alexander A. Razborov, Alexei Semenov, Anatol Slissenko (eds.): *CSR 2008*. LNCS 5010, 228-239. Springer (2008)

[46] Roman Kuznets: Self-referential justifications in epistemic logic. *Theory of Computing Systems* 46(4), 636-661 (2010)

[47] Ugo Dal Lago, Luca Roversi, Luca Vercelli: Taming modal impredicativity: Superlazy reduction. In: Sergei N. Artemov, Anil Nerode (eds.): *LFCS 2009*. LNCS 5407, 137-151. Springer (2009)

[48] Paolo Mancosu (ed.): *From Hilbert to Brouwer. The debate on the Foundations of Mathematics in the 1920s*. Oxford University Press, Oxford (1998)

[49] John Charles Chenoweth McKinsey, Alfred Tarski: Some theorems about the sentential calculi of Lewis and Heyting. *The Journal of Symbolic Logic* 13, 1-15 (1948)

[50] Alexey Mkrtychev: Models for the logic of proofs. In: Sergei Adian, Anil Nerode (eds.): *LFCS 1997*. LNCS 1234, 266-275. Springer (1997)

[51] Sara Negri: Proof theory for modal logic. *Philosophy Compass* 6, 523-538 (2011)

[52] Sara Negri, Jan von Plato: *Structural Proof Theory*. Cambridge University Press, Cambridge (2001)

[53] Masao Ohnishi, Kazuo Matsumoto: Gentzen method in modal calculi. *Osaka Mathematical Journal* 9, 113-130 (1957)

[54] Masao Ohnishi, Kazuo Matsumoto: Gentzen method in modal calculi II. *Osaka Mathematical Journal* 11, 115-120 (1957)

[55] Francesco Paoli: *Substructural Logics: a primer*. Kluwer Academic Publishers (2002)

[56] Cornelia Pulver: Self-referentiality in Contraction-free Fragments of Modal Logic S4. Master thesis. Der Universität Bern (2010)

[57] Natalia Rubtsova: Evidence reconstruction of epistemic modal logic S5. In: Dima Grigoriev, John Harrison, Edward A. Hirsch (eds.): *CSR 2006*, LNCS 3967, 313-321. Springer (2006)

[58] Craig Smoryński: *Self-reference and modal logic*. Springer (1985)

[59] Raymond M. Smullyan: *Diagonalization and Self-Reference*. Oxford Logic Guides 27. Oxford Science (1994)

[60] Morten Heine Sørensen, Pawel Urzyczyn: *Lectures on the Curry-Howard Isomorphism*. Elsevier, Amsterdam (2006)

[61] Yasuhito Suzuki, Hiroakira Ono: Hilbert style proof system for BPL. Research Report IS-RR-97-0040F, JAIST (1993)

[62] Anne Sjerp Troelstra, Dirk van Dalen: *Constructivism in Mathematics, an Introduction*. North-Holland, Amsterdam (1988)

[63] Anne Sjerp Troelstra, Helmut Schwichtenberg: *Basic Proof Theory*, 2nd ed. Cambridge University Press, Cambridge (2000)

[64] Albert Visser: A propositional logic with explicit fixed points. *Studia Logica* 40, 155-175 (1981)

[65] Wang, Ren-June: On non-circular proofs and proof realization in modal logic. Technical Report TR-2011012, CUNY PhD Program in Computer Science (2011)

[66] Noson S. Yanofsky: A universal approach to self-referential paradixes, imcompleteness and fixed points. *The Bulletin of Symbolic Logic* 9(3), 362-386 (2003)

[67] Noson S. Yanofsky: *The Outer Limits of Reason: What Science, Mathematics, and Logic Cannot Tell Us*. The MIT Press (2013)

[68] Yu, Junhua: Prehistoric phenomena and self-referentiality. In: Farid Ablayev, Ernst W. Mayr (eds.): *CSR 2010*. LNCS 6072, 384-396. Springer (2010)

[69] Yu, Junhua: On Self-referentiality in Realization: Possible Criteria via Gentzen Calculus (in Chinese). Master Thesis. Tsinghua University (2010)

[70] Yu, Junhua: Self-referentiality in the Brouwer–Heyting–Kolmogorov semantics of intuitionistic logic. In: Sergei N. Artemov, Anil Nerode (eds.): *LFCS 2013*. LNCS 7734, 401-414. Springer (2013)

[71] Yu, Junhua: Self-referentiality of Brouwer–Heyting–Kolmogorov semantics. *Annals of Pure and Applied Logic* 165(1), 371-388 (2014)

[72] Yu, Junhua: Prehistoric graph in modal derivations and self-referentiality. *Theory of Computing Systems* 54(2), 190-210 (2014)

[73] Yu, Junhua: Closure properties of non-self-referential fragments of modal and intuitionistic logics. Manuscript (2014)