Self-referentiality in Justification Logic

Masterarbeit der
Philosophisch-naturwissenschaftlichen Fakultät
der Universität Bern

vorgelegt von

Nathan Sebastian Gass

2019

Leiter der Arbeit:

Prof. Dr. Thomas Studer
Institut für Informatik und angewandte
Mathematik

Contents

Chapter 1

Introduction

Sergei Artemov first introduced the Logic of Proofs (LP) in Artemov (1995), where he introduced proof terms as explicit and constructive replacements of the $\Box$ modality in S4.[1] Later more applications of this explicit notations were discovered for different epistemic logics (Brezhnev 2001). So Artemov (2008) introduced the more general notion of justification logics where justification terms take over the role of the proof terms in LP. In any justification logic $t{:}A$ is read as t is a justification of A, leaving open what exactly that entails. Using different axioms and different operators, various different justification logic counterparts where developed for the different modal systems used in epistemic logic (K, T, K4, S4, K45, KD45, S5, etc.).

In justification logics it is possible for a term t to be a justification for a formula $A(t)$ containing t itself, i.e. for the assertion $t{:}A(t)$ to hold. Prima facie this seems suspicious from a philosophical standpoint as well for more formal mathematical reasons. Such a self-referential sentence is for example impossible with an arithmetic proof predicate using standard Gödel numbers as the Gödel number of a proof is always greater than any number referenced in it as discussed by Roman Kuznets (2010). In the same paper, the author argues that there is nothing inherently wrong with self-referential justifications if we understand the justifications as valid reasoning templates or schemes which of course then can be used on themselves.

Kuznets studied the topic of self-referentiality at the logic-level. He discovered theorems of S4, D4, T and K4 which need a self-referential constant specification to be realized in their justification logic counterparts (Kuznets 2010). Junhua Yu on the other hand studied self-referentiality at the theorem level. He discovered prehistoric cycles as a necessary condition for self-referential S4 theorems (Yu 2010) and later expanded that results to the modal logics T and K4 (Yu 2014). He also conjectured that the condition is actually sufficient for self-referential S4 theorems. In this paper I will concentrate on that topic, that is prehistoric cycles as necessary and sufficient condition for self-referential theorems in S4.

[1] Although Artemov (1995) is the first comprehensive presentation of the logic of proofs, this paper will mostly reference and use the notations of the presentation in Artemov (2001).

This paper is divided in three parts. In the first part I introduce the modal logic S4 and its justification counterpart LP as well as two Gentzen systems for S4 and LP used in the later parts. The second part reproduces Yu's main theorem, i.e. that prehistoric cycles are a necessary condition for self-referential theorems in S4. The third part goes beyond Yu's original paper by adapting the notion of prehistoric cycles to Gentzen systems with cut rules and finally to a Gentzen system for LP. This allows to study prehistoric cycles directly in LP which leads to the two main results of this paper. Firstly, with the standard definition of self-referential theorems, prehistoric cycles are not a sufficient condition. Second, with an expansion on the definition of self-referential theorems, prehistoric cycles become sufficient for self-referential theorems.

Chapter 2

LP and S4

2.1 Preliminaries

As the results and concepts in this paper are mostly purely syntactical, I will also limit this brief introduction to the modal logic S4 and its justification counterpart LP to the syntactic side. For the semantic side, Artemov already gives an arithmetical interpretation for LP in his original papers (Artemov 1995, 2001). Later Melvin Fitting (2005) developed semantics for LP based on Kripke models. A more comprehensive discussion of semantics as well as syntax of LP and S4 can be found in Studer and Kuznets (n.d. ch 1 and 2).

Definition 1. [Syntax of S4] The language of S4 is given by $A := \bot \mid P \mid A_0 \wedge A_1 \mid A_0 \vee A_1 \mid A_0 \to A_1 \mid \Box A \mid \Diamond A$. By using the known derived definitions for $\wedge$, $\vee$ and $\Diamond$ we can reduce that to the minimal language $A := \bot \mid P \mid A_0 \to A_1 \mid \Box A$.

Definition 2. [Syntax of LP] The language of LP consists of terms given by $t := c \mid x \mid t_0 \cdot t_1 \mid t_0 + t_0 \mid !t$ and formulas given by $A := \bot \mid P \mid A_0 \to A_1 \mid t{:}A$.

A Hilbert style system for LP is given by the following Axioms and the rules modus ponens and axiom necessitation. (Artemov 2001, 8)

- $A0$: Finite set of axiom schemes of classical propositional logic
- $A1$: $t{:}F \to F$ (Reflection)
- $A2$: $s{:}(F \to G) \to (t{:}F \to (s \cdot t){:}G)$ (Application)
- $A3$: $t{:}F \to !t{:}(t{:}F)$ (Proof Checker)
- $A4$: $s{:}F \to (s+t){:}F$, $t{:}F \to (s+t){:}F$ (Sum)

- $R1$: $F \to G, F \vdash G$ (Modus Ponens)
- $R2$: $A \vdash c{:}A$, if A is an axiom $A0 - A4$ and c a constant (Axiom Necessitation)

A Hilbert style derivation d from a set of assumptions Γ is a sequence of formulas $A_0, ...A_n$ such that any formula is either an instance of an axiom A0-A4, a formula $A \in \Gamma$ or derived from earlier formulas by a rule R1 or R2. The notation $\Gamma \vdash_{\mathrm{LP}} A$ means that a LP derivation from assumptions Γ ending in A

exists. We also write $\vdash_{\mathrm{LP}} A$ or $\mathrm{LP} \vdash A$ if a LP derivation for A without any assumptions exists.

When formulating such derivations, we will introduce propositional tautologies without derivation and use the term propositional reasoning for any use of modus ponens together with a propositional tautology. This is of course correct as axioms A0 together with the modus ponens rule R1 are a complete Hilbert style system for classical propositional logic. Its easy to see by a simple complete induction on the proof length that this derivations do not use any new terms not already occurring in the final propositional tautology.

Definition 3. [Constant Specification] A *constant specification* CS is a set of of formulas of the form $c{:}A$ with c a constant and A an axiom A0-A4.

Every LP derivation naturally generates a finite constant specification of all formulas derived by axiom necessitation (R2). For a given constant specification CS, LP(CS) is the logic with axiom necessitation restricted to that CS. $\mathrm{LP}_0 := \mathrm{LP}(\emptyset)$ is the logic without axiom necessitation. A constant specification CS is injective if for each constant c there is at most one formula $c{:}A \in \mathrm{CS}$.

We end this brief introduction with some generic lemmas and theorems about LP, which will be used in later proofs. Especially in chapter 3.3 it will be important to closely track the used justification variables, which is why most results come with a corollary limiting the possible use of justification variables and terms.

Lemma 1. *[Substitution] If $\Gamma \vdash_{\mathrm{LP(CS)}} A$ with a derivation d, then also $\Gamma' \vdash_{\mathrm{LP(CS')}} A'$ with a derivation d' acquired by replacing all occurrences of a variable x by a term t in Γ, CS and d.*

Proof. Trivial induction over the derivation d. ■

Theorem 1. *[Deduction Theorem] If $\Gamma, A \vdash_{\mathrm{LP(CS)}} B$, then $\Gamma \vdash_{\mathrm{LP(CS)}} A \to B$ (Artemov 2001, 9)*

Proof. From a proof d for $A, \Gamma \vdash_{\mathrm{LP}} B$ we inductively construct a proof d' for $\Gamma \vdash_{\mathrm{LP}} A \to B$ as follows:

1. case: $B \equiv A$, then $A \to B \equiv A \to A$ is a propositional tautology and derivable from axioms A0 and modus ponens.

2. case: B is an assumption or an axiom A0-A4. Then d' is the derivation B, $B \to (A \to B)$, $A \to B$.

3. case: $B \equiv c{:}B_0$ is derived by axiom necessitation. Then d' is the derivation B_0, $c{:}B_0$, $c{:}B_0 \to (A \to c{:}B_0)$, $A \to c{:}B_0$.

4. case: B is derived by modus ponens. So there are derivations d_l and d_r for the premises $C \to B$ and C. By induction hypothesis, there are derivations d_l' and d_r' for $A \to (C \to B)$ and $A \to C$. The derivation d' is $(A \to (C \to B)) \to ((A \to C) \to (A \to B))$, d_l', $(A \to C) \to (A \to B)$, d_r', $A \to B$ ■

Corollary 1.1. *The deduction d' for $\Gamma \vdash_{\mathrm{LP(CS)}} A \to B$ only uses variables x also occurring in the deduction d for $A, \Gamma \vdash_{\mathrm{LP(CS)}} B$.*

Proof. As constructed in the main proof, the new deduction d' only uses subformulas from d and does not introduce any new terms. ■

Theorem 2. *[Lifting Lemma] If $x_1{:}A_1, \cdots, x_n{:}A_n \vdash_{\mathrm{LP}} B$, then there is a term t s.t. $x_1{:}A_1, \cdots, x_n{:}A_n \vdash_{\mathrm{LP}} t{:}B$. (Artemov 2001, 9)*

Proof. From a proof d for $x_1{:}A_1, \cdots, x_n{:}A_n \vdash_{\mathrm{LP}} B$ we inductively construct a term t and a proof d' for $x_1{:}A_1, \cdots, x_n{:}A_n \vdash_{\mathrm{LP}} t(x_1, \cdots, x_n){:}B$ as follows:

1. case: $B \equiv x_i{:}A_i$ is an assumption. Then $t := !x_i$ and d' is the derivation $x_i{:}A_i$, $x_i{:}A_i \to !x_i{:}x_i{:}A_i$, $!x_i{:}x_i{:}A_i$.

2. case: B is an axiom A0-A4. Then $t := c$ for a new constant c and d' is the derivation B, $c{:}B$.

3. case: $B \equiv c{:}B_0$ is derived by axiom necessitation. Then $t := !c$ and d' is the derivation B_0, $c{:}B_0$, $c{:}B_0 \to !c{:}c{:}B_0$, $!c{:}c{:}B_0$ as B_0 is an axiom.

4. case: B is derived by modus ponens. So there are derivations d_l and d_r for the premises $C \to B$ and C. By induction hypothesis, there are terms t_l and t_r and derivations d_l' and d_r' for $t_l{:}(C \to B)$ and $t_r{:}C$. Set $t := t_l \cdot t_r$ and the derivation d' is $t_l{:}(C \to B) \to (t_r{:}C \to t_l \cdot t_r{:}B)$, d_l', $t_r{:}C \to t_l \cdot t_r{:}B$, d_r', $t_l \cdot t_r{:}B$ ■

Corollary 2.1. *If $x_1{:}A_1, \cdots, x_n{:}A_n \vdash_{\mathrm{LP(CS)}} B$ based on an injective constant specification CS, then there is a term t and a injective constant specification $\mathrm{CS}' \supset \mathrm{CS}$ s.t. $x_1{:}A_1, \cdots, x_n{:}A_n \vdash_{\mathrm{LP(CS')}} t{:}B$.*

Proof. The proof is exactly the same as for the main theorem, except in the 4. case. In that case we just have to reuse a constant c from CS for the exact same axiom, if it already exists or else add the new constant c to the new constant specification CS'. ■

Corollary 2.2. *The deduction d' for $x_1{:}A_1, \cdots, x_n{:}A_n \vdash_{\mathrm{LP(CS')}} t(x_1, \cdots, x_n){:}B$ and the constant specification of the new constants $\mathrm{CS}' \setminus \mathrm{CS}$ only uses variables x also occurring in the deduction d for $x_1{:}A_1, \cdots, x_n{:}A_n \vdash_{\mathrm{LP}} B$.*

Proof. As constructed in the main proof, the new deduction d' only uses true subformulas and variables already occurring in d. Moreover it only introduces new constants c for axioms A occurring in d. Therefore no new variables are introduced in d' or CS'. ■

2.2 Gentzen Systems

In the following text capital greek letters Γ, Δ are used for multisets of formulas, latin letters P, Q for atomic formulas and latin letters A, B for arbitrary formulas. We also use the following short forms:

- $\Box\Gamma := \{\Box A \mid A \in \Gamma\}$
- $\Gamma, A := \Gamma \cup \{A\}$
- $\Gamma, \Delta := \Gamma \cup \Delta$
- $\bigwedge\Gamma := A_0 \wedge ... \wedge A_n$ and $\bigvee\Gamma := A_0 \vee ... \vee A_n$ for the formulas $A_i \in \Gamma$ in an arbitrary but fixed order.

Throughout this text, we will use the G3s calculus from Troelstra and Schwichtenberg (2000, 287) for our examples with additional rules $(\neg \supset)$ and $(\supset \neg)$ as we are only concerned with classical logic (see figure 2.1). For proofs on the other hand we will use a minimal subset of that system given by figure 2.2 using the standard derived definitions for $\neg$, $\vee$, $\wedge$ and $\Diamond$. All the missing rules from the full system are admissible in the minimal system using this definitions and the theorems therefore carry over to the full G3s system.

Figure 2.1: Full G3s

$$P, \Gamma \supset \Delta, P \ (Ax)\ (P \text{ atomic}) \qquad \bot, \Gamma \supset \Delta \ (\bot \supset)$$

$$\frac{\Gamma \supset \Delta, A}{\neg A, \Gamma \supset \Delta}\ (\neg \supset) \qquad \frac{A, \Gamma \supset \Delta}{\Gamma \supset \Delta, \neg A}\ (\supset \neg)$$

$$\frac{A, B, \Gamma \supset \Delta}{A \wedge B, \Gamma \supset \Delta}\ (\wedge \supset) \qquad \frac{\Gamma \supset \Delta, A \qquad \Gamma \supset \Delta, B}{\Gamma \supset \Delta, A \wedge B}\ (\supset \wedge)$$

$$\frac{A, \Gamma \supset \Delta \qquad B, \Gamma \supset \Delta}{A \vee B, \Gamma \supset \Delta}\ (\vee \supset) \qquad \frac{\Gamma \supset \Delta, A, B}{\Gamma \supset \Delta, A \vee B}\ (\supset \vee)$$

$$\frac{\Gamma \supset \Delta, A \qquad B, \Gamma \supset \Delta}{A \to B, \Gamma \supset \Delta}\ (\to\supset) \qquad \frac{A, \Gamma \supset \Delta, B}{\Gamma \supset \Delta, A \to B}\ (\supset\to)$$

$$\frac{A, \Box A, \Gamma \supset \Delta}{\Box A, \Gamma \supset \Delta}\ (\Box \supset) \qquad \frac{\Box\Gamma \supset \Diamond\Delta, A}{\Gamma', \Box\Gamma \supset \Diamond\Delta, \Delta', \Box A}\ (\supset \Box)$$

$$\frac{A, \Box\Gamma \supset \Diamond\Delta}{\Diamond A, \Gamma', \Box\Gamma \supset \Diamond\Delta, \Delta'}\ (\Diamond \supset) \qquad \frac{\Gamma \supset \Delta, A, \Diamond A}{\Gamma \supset \Delta, \Diamond A}\ (\supset \Diamond)$$

In Artemov (2001, 14), a Gentzen-Style system LPG is introduced for the logic of proofs LP using explicit contraction and weakening rules, i.e. based on G1c as defined in Troelstra and Schwichtenberg (2000, 61). Later we will follow Cornelia Pulver (2010) instead and use G3lp with the structural rules absorbed. For now the system from figure 2.3 closely resembling the only hinted at "LPG_0 + Lifting Lemma Rule" system from Yu (2010) is actually the most practical for our purpose. The reason for this is that it exactly mirrors the rules of G3s. Other than LPG_0 from Yu (2010) and the original Gentzen style systems from Artemov (2001, 14), it does not deconstruct justification terms but falls back on the Hilbert style definition of LP to introduce terms already fully constructed. I will call this system G3lift to differentiate it from the later used system G3lp.

Definition 4. [G3lift Preproof] A *G3lift preproof* is a proof tree using the rules of G3lift, but where the (lift) rule may be used without fulfilling the necessary precondition on the new term t.

In all this rules, arbitrary formulas which occur in the premises and the conclu-

Figure 2.2: Minimal G3s

$$P, \Gamma \supset \Delta, P \ (Ax)\ (P \text{ atomic}) \qquad \bot, \Gamma \supset \Delta\ (\bot \supset)$$

$$\frac{\Gamma \supset \Delta, A \qquad B, \Gamma \supset \Delta}{A \to B, \Gamma \supset \Delta}\ (\to\supset) \qquad \frac{A, \Gamma \supset \Delta, B}{\Gamma \supset \Delta, A \to B}\ (\supset\to)$$

$$\frac{A, \Box A, \Gamma \supset \Delta}{\Box A, \Gamma \supset \Delta}\ (\Box \supset) \qquad \frac{\Box\Gamma \supset A}{\Gamma', \Box\Gamma \supset \Box A, \Delta}\ (\supset \Box)$$

Figure 2.3: G3lift

$$P, \Gamma \supset \Delta, P \ (Ax)\ (P \text{ atomic}) \qquad \bot, \Gamma \supset \Delta\ (\bot \supset)$$

$$\frac{\Gamma \supset \Delta, A \qquad B, \Gamma \supset \Delta}{A \to B, \Gamma \supset \Delta}\ (\to\supset) \qquad \frac{A, \Gamma \supset \Delta, B}{\Gamma \supset \Delta, A \to B}\ (\supset\to)$$

$$\frac{A, t{:}A, \Gamma \supset \Delta}{t{:}A, \Gamma \supset \Delta}\ ({:} \supset) \qquad \frac{t_1{:}B_1, ..., t_n{:}B_n \supset A}{t_1{:}B_1, ..., t_n{:}B_n, \Gamma \supset \Delta, t{:}A}\ (\text{lift})$$

$$\text{where } t_1{:}B_1, ..., t_n{:}B_n \vdash_{\text{LP}} t{:}C$$

sion (denoted by repeated multisets Γ, $\Box\Gamma$, Δ and $\Diamond\Delta$) are called side formulas. Arbitrary formulas which only occur in the conclusion (denoted by new multisets Γ, Δ, Γ', Δ') are called weakening formulas.[1] The remaining single new formula in the conclusion is called the principal formula of the rule. The remaining formulas in the premises are called active formulas. Active formulas are always used as subformulas of the principal formula. Active formulas which are also strict subformulas of other active formulas of the same rule as used in $(:\supset)$ and $(\Box\supset)$ are contraction formulas.

Formally, a Gentzen style proof is denoted by $\mathcal{T} = (T, R)$, where $T := \{S_0, ..., S_n\}$ is the set of occurrences of sequents, and $R := \{(S_i, S_j) \in T \times T \mid S_i \text{ is the conclusion of a rule which has } S_j \text{ as a premise}\}$. The only root sequent of $\mathcal{T}$ is denoted by S_r. A leaf sequent S is a sequent without any premises, i.e $S \not R S'$ for all $S' \in T$.

A path in a proof tree is a list of related sequent occurrences $S_0R...RS_n$. A root path is a path starting at the root sequent S_r. A root-leaf path is a root path ending in a leaf sequent.[2] A root path is fully defined by the last sequent S. So we will use root path S to mean the unique path $S_rRS_0R...RS$ from the root S_r to the sequent S. $T{\upharpoonright}S$ denotes the subtree of T with root S. The transitive closure of R is denoted by R^+ and the reflexive-transitive closure is denoted by R^*.

Consistent with the notation for the Hilbert style system LP, the notation $G \vdash \Gamma \subset \Delta$ is used if there exists a Gentzen style proof tree with the sequent $\Gamma \subset \Delta$ as root in the system G.

Definition 5. [Correspondence] The subformula (symbol) occurrences in a proof correspond to each other as follows:

- Every subformula (symbol) occurrence in a side formula of a premise directly corresponds to the same occurrence of that subformula (symbol) in the same side formula in the conclusion.
- Every active formula of a premise directly correspond to the topmost subformula occurrence of the same formula in the principal formula of the conclusion.
- Every subformula (symbol) occurrence in an active formula of a premise directly corresponds to the same occurrence of that subformula (symbol) in the corresponding subformula in the principal formula of the rule.
- Two subformulas (symbols) correspond to each other by the transitive reflexive closure of direct correspondence.

As by definition correspondence is reflexive and transitive, we get the following definition for the equivalence classes of correspondence:

Definition 6. [Family] A family is an equivalence class of $\Box$ occurrences which respect to correspondence.

[1] Notice that weakening formulas only occur in axioms and the rules $(\supset\Box)$, $(\Diamond\supset)$ and (lift), which are also the only rules which restrict the possible side formulas.

[2] Yu uses the term path for a root path and branch for a root-leaf path. As this terminology is ambiguous we adopted the slightly different terminology given here.

For the following lemma and all the other results in this paper concerning correspondence, we fix a proof tree $\mathcal{T} = (T, R)$ and consider correspondence according to this complete proof tree even when talking about subtrees $T{\upharpoonright}S$ of $\mathcal{T}$.

Lemma 2. *[Subformula Property] Any subformula (symbol) occurrence in a partial Gentzen style (pre-)proof $T{\upharpoonright}S$ in the systems G3lift and G3s corresponds to* at least one *subformula (symbol) occurrence of the sequent S of $T{\upharpoonright}S$.*

Any subformula (symbol) occurrence in a complete Gentzen style (pre-)proof T in the systems G3lift and G3s corresponds to exactly *one subformula (symbol) occurrence in the root sequent S_r of T.*

Proof. As defined, occurrences of subformulas only correspond directly via a rule of the proof tree. Checking the rules of G3lift and G3s, we see that each subformula occurrence in a premise always corresponds directly to exactly one subformula occurrence in the conclusion.

So for any occurrence in the proof tree we get a unique path of corresponding occurrences up to the root sequent of that tree, which proofs the first part of the theorem. For the second part, notice that two occurrences correspond indirectly if and only if their paths to the root sequent merge at some point in the proof tree. So occurrences in the root sequent itself can not correspond to each other. ■

2.3 Adequacy of G3lift

We will show in this chapter that G3lift is adequate by showing it is equivalent to the Hilbert system LP from Artemov (2001) as introduced in chapter 2.1.

Theorem 3. *[Soundness of G3lift]* G3lift $\vdash \Gamma \supset \Delta \Rightarrow \Gamma \vdash_{\mathrm{LP}} \bigvee \Delta$

Proof. We construct a LP derivation d of $\bigvee \Delta$ by structural induction over the proof tree $\mathcal{T} = (T, R)$ for $\Gamma \supset \Delta$.

1. case: $\Gamma \supset \Delta$ is an axiom (Ax) with atomic formula P. Then it has the form $P, \Gamma' \supset \Delta', P$ and P, $P \to \bigvee \Delta' \vee P$, $\bigvee \Delta' \vee P$ is the required LP derivation.

2. case: $\Gamma \supset \Delta$ is an axiom $(\bot \supset)$. Then it has the form $\bot, \Gamma' \supset \Delta$ and $\bot$, $\bot \to \bigvee \Delta$, $\bigvee \Delta$ is the required LP derivation.

3. case: $\Gamma \supset \Delta$ is derived by a $(\to\supset)$ rule. Then it has the form $A \to B, \Gamma' \supset \Delta$ and the the premises are $\Gamma' \supset \Delta, A$ and $B, \Gamma' \supset \Delta$. By the induction hypothesis there exists LP derivations d_L and d_R for $\Gamma' \vdash_{\mathrm{LP}} \bigvee \Delta \vee A$ and $B, \Gamma' \vdash_{\mathrm{LP}} \bigvee \Delta$. By the deduction theorem (thm. 1) there exists a LP derivation d_R' for $\Gamma' \vdash_{\mathrm{LP}} B \to \bigvee \Delta$. Using d_R', the assumption $A \to B$ and propositional reasoning, we get $(A \to B), \Gamma' \vdash_{\mathrm{LP}} A \to \bigvee \Delta$. By appending d_L and propositional reasoning we get the final $(A \to B), \Gamma' \vdash_{\mathrm{LP}} \bigvee \Delta$

4. case: $\Gamma \supset \Delta$ is derived by a $(\supset\to)$ rule. Then it has the form $\Gamma \supset \Delta', A \to B$ and the premise is $A, \Gamma \supset \Delta', B$. By the induction hypothesis there exists a LP derivation d for $A, \Gamma \vdash_{\mathrm{LP}} \bigvee \Delta' \vee B$. From the deduction theorem (thm. 1)

we get $\Gamma \vdash_{\mathrm{LP}} A \rightarrow (\bigvee \Delta' \vee B)$. By propositional reasoning we get the final $\Gamma \vdash_{\mathrm{LP}} \bigvee \Delta' \vee (A \rightarrow B)$.

5. case: $\Gamma \supset \Delta$ is derived by a $(:\supset)$ rule. Then it has the form $t{:}A, \Gamma' \supset \Delta$ and the premise is $A, t{:}A, \Gamma' \supset \Delta$. By the induction hypothesis there exists a LP derivation d for $A, t{:}A, \Gamma' \vdash_{\mathrm{LP}} \bigvee \Delta$. By adding $t{:}A, t{:}A \rightarrow A, A$ to the beginning of d we get the necessary derivation d' for $t{:}A, \Gamma' \vdash_{\mathrm{LP}} \bigvee \Delta$.

6. case: $\Gamma \supset \Delta$ is is derived by a (lift) rule. Then it has the form $t_1{:}A_1, ..., t_n{:}A_n \vdash_{\mathrm{LP}} t{:}A$ and by the precondition on t there exists a derivation of $t_1{:}A_1, ..., t_n{:}A_n \vdash_{\mathrm{LP}} t{:}A$. ■

Corollary 3.1. *The deduction d for $\Gamma \vdash_{\mathrm{LP}} \bigvee \Delta$ only uses variables x which also occur in the proof tree $\mathcal{T} = (T, R)$ for* G3lift $\vdash \Gamma \supset \Delta$ *or any deduction d_t for $t_1{:}A_1, ..., t_n{:}A_n \vdash_{\mathrm{LP}} t{:}A$ used in case 6.*

Proof. As constructed in the main proof, the deduction d only uses true subformulas and variables already occurring in T. For cases 1 and 2 this is immediate as the given derivations directly use subformulas from T. In cases 3, 4 and 5, the starting derivations do not use new variables by induction hypothesis and the use of the deduction theorem does not introduce new variables by corollary 1.1. The derivation for case 6 is already included in the corollary and therefore trivially does not use new variables. ■

The proofs for the following standard lemmas for G3lift exactly mirror the proofs for the same lemmas for G3s, as G3lift exactly mirrors G3s. The only trivial difference is that the precondition of the (lift) rule has to be checked whenever a different/new (lift) rule gets introduced. Because of this and also because the following result is just included for completeness and not actually used for the main theorems of this text, we will omit the proofs here and refer the reader to the proofs for G3s in Pulver (2010, 40ff.) as well as later in this text.

Lemma 3. *[Weakening for G3lift]* G3lift $\vdash \Gamma \supset \Delta \Rightarrow$ G3lift $\vdash \Gamma, \Gamma' \supset \Delta, \Delta'$

Lemma 4. *[Inversion for G3lift]*

- G3lift $\vdash \Gamma \supset \Delta, A \rightarrow B \Rightarrow$ G3lift $\vdash A, \Gamma \supset \Delta, B$
- G3lift $\vdash A \rightarrow B, \Gamma \supset \Delta \Rightarrow$ G3lift $\vdash \Gamma \supset \Delta, A$ *and* G3lift $\vdash B, \Gamma \supset \Delta$
- G3lift $\vdash t : A, \Gamma \supset \Delta \Rightarrow$ G3lift $\vdash A, t : A, \Gamma \supset \Delta$
- G3lift $\vdash \Gamma \supset \Delta, t : A \Rightarrow$ G3lift $\vdash \Gamma \supset \Delta, A$

Lemma 5. *[Contraction for G3lift]*

- G3lift $\vdash A, A, \Gamma \supset \Delta \Rightarrow$ G3lift $\vdash A, \Gamma \supset \Delta$
- G3lift $\vdash \Gamma \supset \Delta, A, A \Rightarrow$ G3lift $\vdash \Gamma \supset \Delta, A$

Lemma 6. *[Cut Elimination for G3lift] If* G3lift $\vdash A, \Gamma \supset \Delta$ *and* G3lift $\vdash \Gamma' \supset \Delta', A$ *then* G3lift $\vdash \Gamma, \Gamma' \supset \Delta, \Delta'$.

Lemma 7. G3lift $\vdash A, \Gamma \supset \Delta, A$ *for any LP formula A.*

Theorem 4. *[Completeness of G3lift]* $\Gamma \vdash_{\mathrm{LP}} A \Rightarrow$ G3lift $\vdash \Gamma \supset A$

Proof. By complete induction over the length of the derivation d for $\Gamma \vdash_{\mathrm{LP}} A$.

1. case A is an axiom A0. By the completeness of G3c included in G3lift there exists a derivation of $\Gamma \supset A$ and $\supset A$ using the subset G3c and lemma 7 for the base cases.

2. case A is an axiom $A1 - A4$. As the derivations in figure 2.4 show, $\supset A$ can be derived for each axiom using lemma 7 for the base cases. $\Gamma \supset A$ follows from weakening.

Figure 2.4: G3lift proofs for LP axioms

$$\dfrac{\dfrac{F, t{:}F \supset F}{t{:}F \supset F}\,(:\supset)}{\supset t{:}F \to F}\,(\supset\to)
\qquad
\dfrac{\dfrac{\dfrac{\dfrac{F, t{:}F \supset F}{t{:}F \supset F}\,(:\supset)}{t{:}F \supset t{:}F}\,(\text{lift})}{t{:}F \supset !t{:}t{:}F}\,(\text{lift})}{\supset t{:}F \to !t{:}t{:}F}\,(\supset\to)$$

$$\dfrac{\dfrac{\dfrac{F, t{:}F \supset F}{t{:}F \supset F}\,(:\supset)}{t{:}F \supset (s+t)F}\,(\text{lift})}{\supset t{:}F \to (s+t)F}\,(\supset\to)
\qquad
\dfrac{\dfrac{\dfrac{F, s{:}F \supset F}{s{:}F \supset F}\,(:\supset)}{s{:}F \supset (s+t)F}\,(\text{lift})}{\supset s{:}F \to (s+t)F}\,(\supset\to)$$

$$\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{F, s{:}(F \to G), t{:}F \supset G, F}{s{:}(F \to G), t{:}F \supset G, F}\,(:\supset) \qquad G, s{:}(F \to G), t{:}F \supset G}{F \to G, s{:}(F \to G), t{:}F \supset G}\,(\to\supset)}{s{:}(F \to G), t{:}F \supset G}\,(:\supset)}{s{:}(F \to G), t{:}F \supset s \cdot t{:}G}\,(\text{lift})}{s{:}(F \to G) \supset t{:}F \to s \cdot t{:}G}\,(\supset\to)}{\supset s{:}(F \to G) \to (t{:}F \to s \cdot t{:}G)}\,(\supset\to)$$

3. case $A \in \Gamma$ is an assumption. We get the required proof for $A, \Gamma' \supset A$ directly from lemma 7.

4. case $A \equiv c{:}A_0$ is derived by rule R1 (Axiom Necessitation). Then A_0 is an axiom and there is a G3lift proof for $\supset A_0$ by induction hypothesis. Appending a (lift) rule gives a G3lift proof for $\Gamma \supset c{:}A_0$.

5. case A is derived by rule R0 (Modus Ponens). By induction hypothesis, we have G3lift proofs for $\Gamma \supset B \to A$ and $\Gamma \supset B$ for the premises of the modus ponens rule. By the inversion lemma we get a G3lift proof for $B, \Gamma \supset A$ and by cut elimination and contraction we get the required proof for $\Gamma \supset A$. ■

2.4 Annotated S4

As we have already seen, all symbol occurrences in a Gentzen style proof can be divided in disjoint equivalence classes of corresponding symbol occurrences. In this text we will be mainly concerned with the equivalence classes of $\Box$

occurrences, called families, and their polarities as defined below. I will therefore define annotated formulas, sequents and proof trees in this chapter which make the families and polarities of $\Box$ occurrences explicit in the notation and usable in definitions.

Definition 7. [Polarity] Assign *positive* or *negative polarity* relative to A to all subformulas occurrences B in A as follows:

- The only occurrence of A in A has positive polarity.
- If an occurrence $B \to C$ in A already has a polarity, then the occurrence of C in $B \to C$ has the same polarity and the occurrence of B in $B \to C$ has the opposite polarity.
- If an occurrence $\Box B$ already has a polarity, then the occurrence of B in $\Box B$ has the same polarity.

Similarly all occurrences of subformulas in a sequent $\Gamma \supset \Delta$ get assigned a *polarity* as follows:

- An occurrence of a subformula B in a formula A in Γ has the opposite polarity relative to the sequent $\Gamma \supset \Delta$ as the same occurrence B in the formula A has relative to A.
- An occurrence of a subformula B in a formula A in Δ has the same polarity relative to the sequent $\Gamma \supset \Delta$ as the same occurrence B in the formula A has relative to A.

This gives the subformulas of a sequent $\Gamma \supset \Delta$ the same polarity as they would have in the equivalent formula $\bigwedge \Gamma \to \bigvee \Delta$. Also notice that for the derived operators all subformulas have the same polarity, except for $\neg$ which switches the polarity for its subformula.

The rules of S4 respect the polarities of the subformulas, so that all corresponding occurrences of subformulas have the same polarity throughout the proof. We therefore assign positive polarity to families of positive occurrences and negative polarity to families of negative occurrences. Moreover, positive families in a S4 proof which have occurrences introduced by a $(\supset \Box)$ rule are called principal positive families or simply principal families. The remaining positive families are called non-principal positive families.[3]

The following definition of annotated formulas and proofs as well as the definition of a realization function based on it in the next chapter is heavily inspired by Fittings use of explicit annotations in Fitting (2009). Other than Fitting, I allow myself to treat symbols $\boxplus_i$, $\boxminus_i$ directly as mathematical objects and define functions on them, instead of encoding the symbols as natural numbers.

For the following definition, we use an arbitrary fixed enumeration for all different classes of families. That is, we enumerate all principal positive families as $p_0, ..., p_{n_p}$, all non-principal positive families as $o_0, ..., o_{n_o}$ and all negative families as $n_0, ..., n_{n_n}$. Given a S4 proof T we then annotate the formulas A in the proof in the following way:

[3] This is the same terminology as used in Yu (2010). In many texts principal families are called essential families following the original text (Artemov 2001).

Definition 8. [Annotated Proof]

$\mathrm{an}_T(A)$ is defined recursively on all occurrences of subformulas A in a proof T as follows:

- If A is the occurrence of an atomic formula P or $\bot$, then $\mathrm{an}_T(A) := A$.
- If $A = A_0 \to A_1$, then $\mathrm{an}_T(A) := \mathrm{an}_T(A_0) \to \mathrm{an}_T(A_1)$
- If $A = \Box A_0$ and the $\Box$ belongs to a principal positive family p_i, then $\mathrm{an}_T(A) := \boxplus_i \mathrm{an}_T(A_0)$.
- If $A = \Box A_0$ and the $\Box$ belongs to a non-principal positive family o_i, then $\mathrm{an}_T(A) := \boxdot_i \mathrm{an}_T(A_0)$.
- If $A = \Box A_0$ and the $\Box$ belongs to a negative family n_i, then $\mathrm{an}_T(A) := \boxminus_i \mathrm{an}_T(A_0)$.

Similarly we define annotated formulas without the context of a proof tree by distinguishing all $\Box$ occurrences as separate families and dropping the distinction between principal positive and non-principal positive. This leads to the following definition:

Definition 9. [Annotated Formula]

$\mathrm{an}_A(B)$ is defined recursively on all occurrences of subformulas B in a formula A as follows:

- If B is the occurrence of an atomic formula P or $\bot$, then $\mathrm{an}_A(B) := B$.
- If $B = B_0 \to B_1$, then $\mathrm{an}_A(B) := \mathrm{an}_A(B_0) \to \mathrm{an}_A(B_1)$
- If $B = \Box B_0$ and has positive polarity in A, then $\mathrm{an}_A(B) := \boxplus_i \mathrm{an}_A(B_0)$ for a new $\boxplus_i$.
- If $B = \Box B_0$ and has negative polarity in A, then $\mathrm{an}_A(B) := \boxminus_i \mathrm{an}_A(B_0)$ for a new $\boxminus_i$.

So for example the annotated version of $\Box((R \to \Box R) \to \bot) \to \bot$ is $\boxminus_0((R \to \boxplus_0 R) \to \bot) \to \bot$.

2.5 Realization

LP and S4 are closely related and LP can be understood as an explicit version of S4. The other way around, S4 can be seen as a version of LP with proof details removed or forgotten. We will establish this close relationship in this chapter formally by two main theorems translating valid LP formulas into valid S4 formulas and vice versa. The former is called forgetful projection, the latter is more complex and called realization.

Definition 10. [Forgetful Projection] The *forgetful projection* A° of a LP formula A is the following S4 formula:

- if A is atomic or $\bot$, then $A^\circ := A$.
- if A is the formula $A_0 \to A_1$ then $A^\circ := A_0^\circ \to A_1^\circ$
- if A is the formula $t{:}A_0$ then $A^\circ := \Box A_0$

The definition is expanded to sets, multisets and sequents of LP formulas in the natural way.

Theorem 5. *If* LP $\vdash A$ *then* $S4 \vdash A^{\circ}$.

Proof. If LP $\vdash A$ then G3lift $\vdash A$ with a proof tree $\mathcal{T} = (T, R)$ by completeness of G3lift (thm. 4). The forgetful projection of the sequents of any G3lift rule map directly to the sequents of an equivalent G3s rule, so the proof tree $\mathcal{T}' = (T^{\circ}, R)$ given by replacing all sequents with the forgetful projection of that sequence is a valid G3s proof with root sequent $\supset A^{\circ}$. By the soundness of G3s we have $S4 \vdash A^{\circ}$.[4] ■

In the other direction, one can realize S4 formulas in LP by replacing the $\Box$ occurrences by explicit justification terms as defined below. Of course most of this realizations will not transform a theorem of S4 into a theorem of LP. So the realization theorem will only assert the existence of a specific realization producing a theorem of LP from a theorem of S4. The constructive proof for the realization theorem also provides us with an algorithm to generate one such realization. However, that realization is not necessarily the only possible realization or the simplest one.

Definition 11. [Realization Function] A *realization function* r_A for a formula A is a mapping from the set of different $\Box$ symbols used in $\mathrm{an}_A(A)$ to arbitrary LP terms. Similarly a *realization function* r_T for a proof T is a mapping from the set of different $\Box$ symbols used in $\mathrm{an}_T(T)$ to arbitrary LP terms. (cf Fitting 2009, 371)

Definition 12. [LP-Realization] By an *LP-realization* of a modal formula A we mean an assignment of proof polynomials to all occurrences of the modality in A along with a constant specification of all constants occurring in those proof polynomials. By A^r we understand the image of A under a realization r (Artemov 2001, 25).

A LP-realization of A is fully determined by a realization function r_A (respective a realization function r_T relative to a proof tree for $\supset A$) and a constant specification of all constants occurring in r_A or r_T with $A^r := r_A(\mathrm{an}_A(A))$ respective $A^r := r_T(\mathrm{an}_T(A))$.

If we read $\Box A$ as there exists a proof for A and $t{:}A$ as t is a proof for A, this process seems immediately reasonable. For the example $\neg\Box A$, read as there is no proof of A, and its realization $\neg t{:}A$, read as t is not a proof of A, on the other hand, that process seems wrong at first. But justification logic without any quantifications over proofs is still enough to capture the meaning of $\neg\Box A$ by using Skolem's idea of replacing quantifiers with functions. That is, we realize $\neg\Box A$ using an implicitly all quantified justification variable $\neg x{:}A$. The same example formulated without the derived connective $\neg$ is $x{:}A \to \bot$. That formula can be read as function which produces a contradiction from a given proof x for A.

This last interpretation also hints at the role of complex justification terms using variables in a realization. They define functions from input proofs named by the

[4] A direct proof for this theorem without using G3lift can be found in Studer and Kuznets (n.d. ch 3.1)

variables to output proofs for different formulas. So a realization $x{:}A \to t(x){:}B$ of a S4 formula $\Box A \to \Box B$ actually defines a function $t(x)$ producing a proof for B from a proof x for A. This then is the Skolem style equivalent of the quantified formula $\exists(x)x{:}A \to \exists(y)y{:}B$ which is the direct reading of $\Box A \to \Box B$ (cf Artemov 2008, 497). This discussion implies that we should replace $\Box$ with negative polarity with justification variables, which leads to the following definition of a normal realization:

Definition 13. [Normal] A realization function is *normal* if all symbols for negative families and non-principal positive families are mapped to distinct variables. A LP-realization is *normal* if the corresponding realization function is normal and the CS is injective.

We are now ready to complete the connection between S4 and LP by the following realization theorem giving a constructive way of producing the necessary proof functions to realize a S4 theorem in LP:

Theorem 6. *Realization If $S4 \vdash A$ then* LP $\vdash A^r$ *for some normal LP-realization r.*

Proof. Because of $S4 \vdash A$ and the completeness of G3s, there exists a G3s proof $\mathcal{T} = (T, R)$ of $\supset A$.

For all principal families $\boxplus_i$ in $\mathrm{an}_T(T)$, enumerate the $(\supset \Box)$ rules principally introducing an occurrence of $\boxplus_i$ as $R_{i,0}, ...R_{i,l_i-1}$. We will use $I_{i,0}, ...I_{i,l_i-1}$ to denote the premises and $O_{i,0}, ...O_{i,l_i-1}$ to denote the conclusions of this rules (so for all $i \leq n_p$, $j < l_i$ we have $I_{i,j}RO_{i,j}$). In total there are $N = \Sigma_{i=0}^{n} l_i$ $(\supset \Box)$ rules in the proof T.

Choose an order $\varepsilon(i,j) \to \{1, ..., N\}$ of all the $(\supset \Box)$ rules such that $\varepsilon(i_2, j_2) < \varepsilon(i_1, j_1)$ whenever $O_{i_1,j_1} R^+ O_{i_2,j_2}$ (i.e. rules closer to the root S_r are later in this order).

Define the normal realization function r_T^0 by $r_T^0(\boxplus_i) := u_{i,0} + ... + u_{i,l_i-1}$ and the injective constant specification $\mathrm{CS}^0 := \emptyset$. The rules of the minimal Gentzen systems G3s for S4 all have a direct equivalent in G3lift, so by a trivial induction the proof tree $r_T^0(\mathrm{an}_T(T))$ is a G3lift preproof. However it is not a G3lift proof as none of the (lift) rules fulfill the necessary precondition on the introduced term t.

We therefore define inductively the normal realization functions $r_T^{\varepsilon(i,j)}$ and injective constant specifications $\mathrm{CS}^{\varepsilon(i,j)}$ such that $r_T^{\varepsilon(i,j)}(\mathrm{an}_T(T \restriction O_{i_0,j_0}))$ is a correct G3lift proof based on $\mathrm{CS}^{\varepsilon(i,j)}$ for all (i_0, j_0) such that $\varepsilon(i_0, j_0) \leq \varepsilon(i,j)$.

The rule $R_{i,j}$ has the following annotated form:

$$\frac{\boxminus_{k_0} B_{k_0}, ..., \boxminus_{k_q} B_{k_q} \supset A}{\Gamma', \boxminus_{k_0} B_{k_0}, ..., \boxminus_{k_q} B_{k_q} \supset \boxplus_i A}$$

By the induction hypothesis there exists an injective constant specification $\mathrm{CS}^{\varepsilon(i,j)-1}$ and a normal realization function $r_T^{\varepsilon(i,j)-1}$ such that $r_T^{\varepsilon(i,j)-1}(\mathrm{an}_T(T \restriction O_{i0,j0}))$ is a correct G3lift proof based on $\mathrm{CS}^{\varepsilon(i,j)-1}$ for all (i_0, j_0) such that $\varepsilon(i_0, j_0) < \varepsilon(i, j)$. From this it follows by a trivial induction on the proof tree that $r_T^{\varepsilon(i,j)-1}(\mathrm{an}_T(T \upharpoonright I_{i,j}))$ is also a correct G3lift proof. By soundness of G3lift (thm. 3) we therefore have a $\mathrm{LP}(\mathrm{CS}^{\varepsilon(i,j)-1})$ derivation for $r_T^{\varepsilon(i,j)-1}(\mathrm{an}_T(I_{i,j}))$, which has the following form:

$$r_T^{\varepsilon(i,j)-1}(\boxminus_{k_0} B_{k_0}), ..., r_T^{\varepsilon(i,j)-1}(\boxminus_{k_q} B_{k_q}) \vdash_{\mathrm{LP}(\mathrm{CS}^{\varepsilon(i,j)-1})} r_T^{\varepsilon(i,j)-1}(A) \tag{1}$$

By the corollary 2.1 of the lifting lemma, we get a new proof term $t_{i,j}(x_{k_0}, ..., x_{k_q})$ and a new injective $\mathrm{CS}'^{\varepsilon(i,j)} = \mathrm{CS}^{\varepsilon(i,j)-1} \cup \{c_{i,j,0}{:}A_{i,j,0}, ..., c_{i,j,m_{i,j}}{:}A_{i,j,m_{i,j}}\}$ such that:

$$r_T^{\varepsilon(i,j)-1}(\boxminus_{k_0} B_{k_0}), ..., r_T^{\varepsilon(i,j)-1}(\boxminus_{k_q} B_{k_q}) \vdash_{\mathrm{LP}(\mathrm{CS}'^{\varepsilon(i,j)})} t_{i,j}{:}r_T^{\varepsilon(i,j)-1}(A) \tag{2}$$

Define $r_T^{\varepsilon(i,j)}$ and $\mathrm{CS}^{\varepsilon(i,j)}$ by replacing $u_{i,j}$ with t in $r_T^{\varepsilon(i,j)-1}$ and $\mathrm{CS}'^{\varepsilon(i,j)}$. By the substitution lemma, the assertion (2) still holds for $r_T^{\varepsilon(i,j)}$ and $\mathrm{CS}^{\varepsilon(i,j)}$. The formula $r_T^k(\boxplus_i A)$ has the form $(s_0 + \cdots + s_{j-1} + t_{i,j} + s_{j+1} + \cdots + s_{l_i-1}){:}A$. Therefore $\mathrm{LP}_0 \vdash t_{i,j}{:}A \to r_T^k(\boxplus_i){:}A$ follows from repeated use of $A4$. Together with the substituted proof for (2) we get the precondition required for the final $(\supset{:})$ rule in $r_T^{\varepsilon(i,j)}(\mathrm{an}_T(T \upharpoonright O_{i,j}))$:

$$r_T^{\varepsilon(i,j)-1}(\boxminus_{k_0} B_{k_0}), ..., r_T^{\varepsilon(i,j)-1}(\boxminus_{k_q} B_{k_q}) \vdash_{\mathrm{LP}(\mathrm{CS}^{\varepsilon(i,j)})} r_T^{\varepsilon(i,j)-1}(\boxplus_i A) \tag{3}$$

Moreover, this precondition remains fulfilled for the $(\supset{:})$ rule $R_{i,j}$ in any proof tree $r_T^k(\mathrm{an}_T(T))$ for $k > \varepsilon(i,j)$ again by the substitution lemma.

For the final normal realization function r_T^N and injective constant specification CS^N we have that $r_T^N(\mathrm{an}_T(T))$ is a correct G3lift proof based on CS^N of $\supset r_T(A)$. So by soundness of G3lift (thm. 3) we have $\mathrm{LP} \vdash A^r$ for the normal LP-realization r given by r_T^N and the injective constant specification CS^N. ■

Corollary 6.1. *There exist derivations $d_{i,j}^k$ for the precondition 3 of all rules $R_{i,j}$ in the G3lift proof tree $r_T^k(\mathrm{an}_T(T))$ for any $k \geq \varepsilon(i,j)$ which do not use any new variables not already occurring in $r_T^k(\mathrm{an}_T(T))$.*

Proof. Proof by complete induction over the order $\varepsilon(i,j)$. Given a rule $R_{i,j}$, there exist derivations d_{i_0,j_0}^k which do not use new variables for the precondition of any rule R_{i_0,j_0} in $r_T^k(\mathrm{an}_T(T \upharpoonright I_{i,j}))$ as $\varepsilon(i_0,j_0) < \varepsilon(i,j) \leq k$ for all this rules. Using the exact same steps as in the main proof but using the realization function r_T^k, we get a derivation d for (1) which does not use new variables by the corollary 3.1, a derivation d' for (2) which does not introduce new variables by the corollary 2.2 and finally a derivation $d_{i,j}^k$ for (3) which also does not introduce new variables. ■

Chapter 3

Prehistoric Relations in G3s

3.1 Self-referentiality

As already mentioned in the introduction, the formulation of LP allows for terms t to justify formulas $A(t)$ about themselves. We will see that such self-referential justification terms are not only possible, but actually unavoidable for realizing S4 even at the basic level of justification constants. That is to realize all S4 theorems in LP, we need self-referential constant specifications defined as follows:

Definition 14. [Self-Referential Constant Specification]

- A constant specification CS is *directly self-referential* if there is a constant c such that $c{:}A(c) \in \mathrm{CS}$.
- A constant specification CS is *self-referential* if there is a subset $A \subseteq \mathrm{CS}$ such that $A := \{c_0{:}A(c_1), ..., c_{n-1}{:}A(c_0)\}$.

A constant specification which is not directly self-referential is denoted by CS^*. Similarly a constant specification which is not self-referential at all is denoted by $\mathrm{CS}^{\circledast}$. So CS^* and $\mathrm{CS}^{\circledast}$ stand for a class of constant specifications and not a single specific one. Following Yu (2010), I will use the notation $\mathrm{LP}(\mathrm{CS}^{\circledast}) \vdash A$ if there exists any non-self-referential constant specification CS such that $\mathrm{LP}(\mathrm{CS}) \vdash B$. There does exist a single maximal constant specification CS_{nds} which is not directly self-referential and for any theorem A we have $\mathrm{LP}(\mathrm{CS}^*) \vdash A$ iff $\mathrm{LP}(\mathrm{CS}_{nds}) \vdash A$.

Given that any S4 theorem is realizable in LP with some constant specification, we can carry over the definition of self-referentiality to S4 with the following definition:

Definition 15. [Self-Referential Theorem] A S4 theorem A is (directly) self-referential iff for any LP-realization A^r we have $\mathrm{LP}(\mathrm{CS}^{\circledast}) \nvdash A^r$ (respective $\mathrm{LP}(\mathrm{CS}^*) \nvdash A^r$).

Expanding on a first result for S4 in Brezhnev and Kuznets (2006, 31), Kuznets (2010, 650) explores the topic of self-referentiality on the level of individual modal logics and their justification counterparts. He gives theorems for the modal logics S4, D4, T, and K4 which can only be realized in their justification logic counterpart using directly self-referential constant specifications, i.e. directly self-referential theorems by the above definition. So for S4 in particular, Kuznets gives the theorem $\neg\Box\neg(S \to \Box S)$ and shows that it is directly self-referential.

We will not reproduce this result but use the logically equivalent formula $\neg\Box(P\land\neg\Box P)$ as an example for a self-referential S4 theorem. Notice that it does not directly follow from the above theorem that $\neg\Box(P \land \neg\Box P)$ can only be realized with a self-referential constant specification, as justification terms do not necessary apply to logically equivalent formulas (Artemov and Fitting 2016 ch 1.3). Still it should be fairly straightforward to show that $\neg\Box(P \land \neg\Box P)$ is self-referential by translating justification terms for the outer $\Box$ occurrences in $\neg\Box(P \land \neg\Box P)$ and $\neg\Box\neg(S \to \Box S)$ using the logical equivalence of $P \land \neg\Box P$ and $\neg(S \to \Box S)$.

The following short discussion of the significance of this example is largely based on the more in depth account of Studer and Kuznets (n.d. ch 7.1). The subformula $P \land \neg\Box P$ in our example asserts for an atomic sentence P, for example "it will rain", to be true and unknown. This sentence "It will rain and I do not know that it will rain" is inspired by Moore's paradox and its formalization $P \land \neg\Box P$ is called a Moore sentence. The sentence is easily satisfiable, for example if the weather forecast wrongly predicts no rain. But it is impossible to know that sentence, as is stated by our example theorem $\neg\Box(P \land \neg\Box P)$. Because if one knows the Moore sentence, one also knows the first part of the conjunction, i.e. P. This knowledge then contradicts the second part of the conjunction, $\neg\Box P$.

Looking at the G3s proof for $\neg\Box(P \land \neg\Box P)$ and a realization of that proof in figure 3.1, we can see why a self referential term like t for the propositional tautology $P \land \neg t \cdot x{:}P \to P$ is necessary. In order to prove $\neg\Box(P \land \neg\Box P)$ one needs to disprove $P \land \neg\Box P$ at some point which means one has to prove $\Box P$. The only way to prove $\Box P$ is using $\Box(P \land \neg\Box P)$ as an assumption on the left. This leads to the situation that the proof introduces $\Box$ by a $(\supset \Box)$ rule where the same family already occurs on the left. As the following sections of this chapter will show formally such a situation is actually necessary for the self-referentiality of any S4 formula.

Figure 3.1: proof for $\neg\Box(P \land \neg\Box P)$

$$\begin{array}{cl}
P, \neg\Box P, \Box(P \land \neg\Box P) \supset P & \\ \hline
P \land \neg\Box P, \Box(P \land \neg\Box P) \supset P & (\land \supset) \\ \hline
\Box(P \land \neg\Box P) \supset P & (\Box \supset) \\ \hline
P, \Box(P \land \neg\Box P) \supset \Box P & (\supset \Box) \\ \hline
P, \neg\Box P, \Box(P \land \neg\Box P) \supset & (\neg \supset) \\ \hline
P \land \neg\Box P, \Box(P \land \neg\Box P) \supset & (\land \supset) \\ \hline
\Box(P \land \neg\Box P) \supset & (\Box \supset) \\ \hline
\supset \neg\Box(P \land \neg\Box P) & (\supset \neg)
\end{array}
\qquad
\begin{array}{cl}
P, \neg t \cdot x{:}P, x{:}(P \land \neg t \cdot x{:}P) \supset P & \\ \hline
P \land \neg t \cdot x{:}P, x{:}(P \land \neg t \cdot x{:}P) \supset P & (\land \supset) \\ \hline
x{:}(P \land \neg t \cdot x{:}P) \supset P & ({:}\supset) \\ \hline
P, x{:}(P \land \neg t \cdot x{:}P) \supset t \cdot x{:}P & (lift) \\ \hline
P, \neg t \cdot x{:}P, x{:}(P \land \neg t \cdot x{:}P) \supset & (\neg \supset) \\ \hline
P \land \neg t \cdot x{:}P, x{:}(P \land \neg t \cdot x{:}P) \supset & (\land \supset) \\ \hline
x{:}(P \land \neg t \cdot x{:}P) \supset & ({:}\supset) \\ \hline
\supset \neg x{:}(P \land \neg t \cdot x{:}P) & (\supset \neg)
\end{array}$$

3.2 Prehistoric Relations

In his paper “Prehistoric Phenomena and Self-referentiality”, (Yu 2010) gives a formal definition for the situation described in the last chapter, which he calls a prehistoric loop. In the later paper Yu (2017) adopts the proper graph theoretic term cycle as we do here. Beside that change we will reproduce his definitions of prehistoric relation, prehistoric cycle as well as some basic lemmas about this new notions exactly as they were presented in the original paper.

To work with the ($\supset \Box$) rules introducing occurrences of principal families in a G3s proof, we will use the same notation already introduced in the proof of the realization theorem (thm. 6). That is, we enumerate all ($\supset \Box$) rules introducing an occurrence of the principal family p_i as $R_{i,0}, ...R_{i,l_i-1}$ and use $I_{i,0}, ...I_{i,l_i-1}$ to denote the premises of those rules and $O_{i,0}, ...O_{i,l_i-1}$ to denote their conclusions.

Definition 16. [History] In a root-leaf path S of the form $S_r R^* O_{i,j} R I_{i,j} R^* S$ in a G3s−proof T, the path $S_r R^* O_{i,j}$ is called a *history* of the family p_i in the root-leaf path S. The path $I_{i,j} R^* S$ is called a *pre-history* of p_i in the root-leaf path S.[1]

So intuitively every ($\supset \Box$) rule divides a root-leaf path of the proof tree into two parts. The first part from the root of the tree to the conclusion of the ($\supset \Box$) rule of sequents having a copy of that $\Box$ symbol, i.e. the history of that $\Box$ symbol from its formation up to the root sequent. And the second part which predates the formation of that $\Box$ symbol, i.e. all sequents from the leaf up to the premise of that ($\supset \Box$) rule, which do not have a copy of that symbol. The informal notion of “having a copy of that symbol” is not the same as correspondence, as it is not transitively closed. It is possible to have corresponding $\boxplus_i$ occurrences of a family p_i in a prehistory of that same family. It is even possible for $\boxplus_i$ occurrences of the same family p_i to be introduced multiple time in the same root-leaf path and therefore having multiple different prehistoric periods in the same root-leaf path. The proof in 3.1 of our example theorem exhibits both these cases.

As we are especially interested in these cases, that is occurrences of principal families in prehistoric periods, the following definition and lemma give that concept a precise meaning and notation:

Definition 17. [Prehistoric Relation] For any principal positive families p_i and p_h and any root-leaf path S of the form $S_r R^* O_{i,j} R I_{i,j} R * S$ in a S4 proof $\mathcal{T} = (T, R)$:

(1) If $\mathrm{an}_T(I_{i,j})$ has the form $\boxminus_{k_0} B_{k_0}, ..., \boxminus_k B_k(\boxplus_h C), ..., \boxminus_{k_q} B_{k_q} \supset A$, then p_h is a *left prehistoric family* of p_i in S with notation $h \prec^S_L i$.

(2) If $\mathrm{an}_T(I_{i,j})$ has the form $\boxminus_{k_0} B_{k_0}, ..., \boxminus_{k_q} B_{k_q} \supset A(\boxplus_h C)$ then p_h is a *right prehistoric family* of p_i in S with notation $h \prec^S_R i$.

(3) The relation of *prehistoric family* in S is defined by: $\prec^S := \prec^S_L \cup \prec^S_R$. The relation of *(left, right) prehistoric family* in T is defined by: $\prec_L := \bigcup \{\prec^S_L | S \text{ is a leaf}\}$, $\prec_R := \bigcup \{\prec^S_R | S \text{ is a leaf}\}$ and $\prec := \prec_L \cup \prec_R$.

[1] see Yu (2010), 389

Even though both definitions so far use the notion of a prehistory, they do not directly refer to each other. But the following lemma provides the missing connection between these two definitions and therefore explains the common terminology:

Lemma 8. *There is an occurrence of $\boxplus_h$ in a pre-history of p_i in the root-leaf path S iff $h \prec^S i$.*

Proof. ($\Rightarrow$): $\boxplus_h$ occurs in a sequent S' in a pre-history of p_i in the root-leaf path S, so S has the form $S_r R^* O_{i,j} R I_{i,j} R^* S' R^* S$ for some $j < l_i$. By the subformula property, there is an occurrence of $\boxplus_h$ in $I_{i,j}$ as S' is part of $T \upharpoonright I_{i,j}$. If this occurrence is on the left we have $h \prec_L^S i$, if it is on right we have $h \prec_R^S i$. In both cases $h \prec^S i$ holds.

($\Leftarrow$): By definition there is a $I_{i,j}$ in S, where $\boxplus_h$ occurs either on the left (for $h \prec_L^S i$) or on the right (for $h \prec_R^S i$). $I_{i,j}$ is part of the pre-history of $R_{i,j}$ in S. ■

Having introduced the concepts of prehistoric periods and prehistoric relations, we are now ready to define the concept of prehistoric cycles used in Yu's theorem:

Definition 18. [Prehistoric Cycle] In a G3s$-$proof T, the ordered list of principal positive families $p_{i_0}, ..., p_{i_{n-1}}$ with length n is called a *prehistoric cycle* or *left prehistoric cycle* respectively, if we have: $i_0 \prec i_2 \prec ... \prec i_{n-1} \prec i_0$ or $i_0 \prec_L i_2 \prec_L ... \prec_L i_{n-1} \prec_L i_0$.

In our example formula, we have a prehistoric cycle consisting of a single principal family which has a left prehistoric relation to itself. The following lemmas will show that any prehistoric cycle necessarily contains left prehistoric relations and that we can get rid of any right prehistoric relations. That is, if a proof has a prehistoric cycle it also has a left prehistoric cycle:

Lemma 9. *For any principal positive family p_i, $i \not\prec_R i$.*

Proof. Assume for a contradiction that $i \prec_R i$. It follows from the definition of $\prec_R$, that there is a rule $R_{i,j}$ with $\boxplus_i A(\boxplus_i B)$ as the principal formula. By the subformula property $\boxplus_i A(\boxplus_i B)$ corresponds to a subformula in the root sequent. Also by the subformula property there is only one occurrence of $\boxplus_i$ in the root sequent. ■

Lemma 10. *If $k \prec_R j$ and $j \rhd i$, then $k \rhd i$, where $\rhd$ is any one of $\prec$, $\prec_L$, $\prec_R$, $\prec^s$, $\prec^s_L$ or $\prec^s_R$.*

Proof. Since $k \prec_R j$, there is a $\boxplus_k$ occurring in the scope of a principally introduced $\boxplus_j$, that is a $(\supset \Box)$ rule with the principal formula $\boxplus_j A(\boxplus_k B)$. So by the subformula property, all corresponding occurrences of $\boxplus_j$ are part of corresponding occurrences of the subformula $\boxplus_j A(\boxplus_k B)$, with exactly one occurrence in the root sequent S_r. Therefore, wherever $\boxplus_j$ occurs in the proof T, there is a $\boxplus_k$ occurring in the scope of it.

For any $\rhd$, we have $j \rhd i$ because of some occurrence of $\boxplus_j$ in a subformula of the premise of a rule $R_{i,q}$. By the previous statement there is also an occurrence of $\boxplus_k$ in the same scope, and therefore also $k \rhd i$. ■

Lemma 11. *T has a prehistoric cycle iff T has a left prehistoric cycle.*

Proof. The ($\Leftarrow$) direction is trivial. The ($\Rightarrow$) direction is proven by complete induction on the length of the cycle as follow:

$n = 1$: $i_0 \prec i_0$ so either $i_0 \prec_R i_0$ or $i_0 \prec_L i_0$. As $i_0 \prec_R i_0$ is impossible by lemma 9, we have $i_0 \prec_L i_0$ and the cycle already is a left prehistoric cycle.

$n-1 \Rightarrow n$: If $i_k \prec_L i_{k+1 \mod n}$ for all $k < n$, then the cycle already is a left prehistoric cycle and we are finished. Otherwise there is a $k < n$ such that $i_k \prec_R i_{k+1 \mod n} \prec i_{k+2 \mod n}$. By lemma 10 we also have $i_k \prec i_{k+2 \mod n}$ and therefore the sublist of length $n-1$ without $i_{k+1 \mod n}$ is also a prehistoric cycle. So T has a left prehistoric cycle by the induction hypothesis. ■

3.3 Yu's Theorem

Yu's proof for the main theorem of his paper is based on the idea to carefully choose the order $\varepsilon(i,j)$ used in the realization theorem (thm. 6), such that the generated constant specifications $\mathrm{CS}^{\varepsilon(i,j)}$ never contain any provisional variables $u_{x,y}$. With such an order, formulas $c{:}A_{i,j,k}$ introduced during the realization procedure never change after their introduction, and we get a strong limitation of the constants which can occur in such a formula. This limitation finally will show that the generated CS using that order can not be self-referential.

Lemma 12. *Any provisional variable $u_{x,y}$, which does not occur in $r^{\varepsilon(i,j)-1}(\mathrm{an}_T(I_{i,j}))$, does not occur in $\mathrm{CS}^{\varepsilon(i,j)}$.*

Proof. By the subformula property for G3lift proofs, $u_{x,y}$ does not occur in $r^{\varepsilon(i,j)-1}(\mathrm{an}_T(T{\upharpoonright}I_{i,j}))$. By the corollary 3.1 using corollary 6.1 for case 6, the derivation $d_{i,j}$ as constructed in the realization proof does not contain $u_{x,y}$. By the corollary 2.2 of the lifting theorem, $\mathrm{CS}'_{i,j}$ and $t_{i,j}$ do not contain $u_{x,y}$. So also $\mathrm{CS}_{i,j}$ constructed by a substitution of $u_{i,j}$ with $t_{i,j}$ does not contain $u_{x,y}$. ■

Lemma 13. *If a G3s−proof T is prehistoric-cycle-free, then we can realize it in such a way that: If $h_2 \prec h_1$, then $\varepsilon(h_2, j_2) < \varepsilon(h_1, j_1)$ for any $j_1 < l_{h_1}$ and $j_2 < l_{h_2}$.*

Proof. For a prehistoric-cycle-free proof T, $\prec$ describes a directed acyclic graph. Therefore there exists a topological order $p_{k_0}, ..., p_{k_{n_p}}$ of the $n_p + 1$ principal positive families $p_0, ..., p_{n_p}$. For any root-leaf path S of the form $S_r R^* O_{i_1,j_1} R^+ O_{i_2,j_2} R^* S$, we have $i_2 \prec i_1$ by lemma 8. So the order $\varepsilon(k_x, j) := j + \Sigma_{w=0}^{x-1} l_{k_w}$ defined for each family p_{k_x} and each $j < l_{k_x}$ by handling the families p_i in the given topological order k_x fulfills the necessary condition to be used in the realization theorem (thm. 6) and at the same time the condition given in this lemma. ■

Lemma 14. *Assume the proof tree is prehistoric-cycle-free. Taken the ε as defined in lemma 13, we have: If $\varepsilon(i_0, j_0) \geq \varepsilon(i,j)$, then for any $k_0 \leq m_{i_0,j_0}$ and any $k \leq m_{i,j}$, c_{i_0,j_0,k_0} does not occur in the unique $A^N_{i,j,k}$ such that $c_{i,j,k}{:}A^N_{i,j,k} \in$ CS^N*

Proof. By the construction in the proof of the realization theorem (thm. 6), $d_{i,j}$ is a derivation of $r_T^{\varepsilon(i,j)-1}(\mathrm{an}_T(I_{i,j}))$. For any $\boxplus_h$ occurring in $I_{i,j}$, we have by definition $h \prec i$, and therefore by lemma 13 $\varepsilon(h, j_h) \leq \varepsilon(i,j)$ for all $j_h < l_h$. So any provisional variable u_{h,j_h} occurring in $r_T^0(\mathrm{an}_T(I_{i,j}))$ is already replaced in $r_T^{\varepsilon(i,j)-1}(\mathrm{an}_T(I_{i,j}))$, which is therefore provisional variable free. So by lemma 12 also $\mathrm{CS}^{\varepsilon(i,j)}$ is provisional variable free and $A_{i,j,k}^N \equiv A_{i,j,k}$ for any $c_{i,j,k}{:}A_{i,j,k}$ introduced in $\mathrm{CS}^{\varepsilon(i,j)}$. As any c_{i_0,j_0,k_0} for any $\varepsilon(i_0,j_0) \geq \varepsilon(i,j)$ is not yet introduced in $r_T^{\varepsilon(i,j)-1}(\mathrm{an}_T(I_{i,j}))$, it does not occur in $A_{i,j,k}$ and therefor also not in $A_{i,j,k}^N \equiv A_{i,j,k}$ ■

With this three lemmas we can finally proof the main result of Yu (2010, 394):

Theorem 7. *[Necessity of Left Prehistoric Cycle for Self-referentiality] If a S4−theorem A has a left-prehistoric-cycle-free G3s−proof, then there is a LP−formula B s.t.* $B^\circ = A$ *and* $\mathrm{LP}(\mathrm{CS}^{\circledast}) \vdash B$

Proof. Given a left-prehistoric-cycle-free G3s−proof T for A, use lemma 13 and the realization theorem (thm. 6) to construct a realization function r_T^N and a constant specification CS^N such that $B := r_T^N(\mathrm{an}_T(A))$ is a realization of A.

Assume for contradiction, that the generated CS^N is self-referential, i.e. there exist constants $c_{i_0,j_0,k_0}, ..., c_{i_{n-1},j_{n-1},k_{n-1}}$ such that for all $x < n$ the unique $c_{i_x,j_x,k_x}{:}A^N_{i_x,j_x,k_x} \in \mathrm{CS}^N$ contains the next constant $c_{i_{x'},j_{x'},k_{x'}}$ with $x' := x+1$ mod n. By lemma 14 we get $\varepsilon(i_{x'}, j_{x'}) < \varepsilon(i_x, j_x)$ for all $x \leq n$. So $\varepsilon(i_n, j_n) < ... < \varepsilon(i_1, j_1) < \varepsilon(i_0, j_0) < \varepsilon(i_n, j_n)$, which is impossible. Therefore the generated CS^N is not self-referential and we have $\mathrm{LP}(\mathrm{CS}^{\circledast}) \vdash B$. ■

Chapter 4

Prehistoric Relations in G3lp

4.1 Cut Rules

In this chapter we will prepare our discussion of prehistoric relations for LP, by first expanding the notion of families and prehistoric relations to the systems G3s + (Cut) and G3s + (□Cut) using cut rules. The (context sharing) cut rule has the following definition (Troelstra and Schwichtenberg 2000, 67):

Definition 19. [(Cut) Rule]

$$\frac{\Gamma \supset \Delta, A \qquad A, \Gamma \supset \Delta}{\Gamma \supset \Delta}\ (\mathrm{Cut})$$

It is necessary to expand the definition of correspondence (def. 5) to (Cut) rules as follows:

Definition 20. [Correspondence for (Cut)]

- The active formulas (and their symbols) in the premises of a (Cut) rule correspond to each other.

The classification and annotations for families of □ do not carry over to G3s + (Cut), as the (Cut) rule uses the cut formula in different polarities for the two premises. We therefore will consider *all* □ families for prehistoric relations in G3s + (Cut) proofs. This leads to the following expanded definition of prehistoric relation:

Definition 21. [Local Prehistoric Relation in G3s + (Cut)] A family $\Box_i$ has a *prehistoric relation* to another family $\Box_j$, in notation $i \prec j$, if there is a $(\supset \Box)$ rule introducing an occurrence of $\Box_j$ with premise S, such that there is an occurrence of $\Box_i$ in S.

Notice that there can be prehistoric relations with □ families which locally have negative polarity, as the family could be part of a cut formula and therefore also occur with positive polarity in the other branch of the cut. On the other hand, adding prehistoric relations with negative families in a cut free G3s proof does not introduce prehistoric cycles, as in G3s a negative family is never introduced by a $(\supset \Box)$ rule and therefore has no prehistoric families itself. In G3s + (Cut) proofs, the subformula property and therefore also lemma 8 no longer hold. That means we can have an occurrence of a family $\Box$ as part of a cut formula in the *global* prehistory of a $(\supset \Box)$ rule, which by the *local* definition is not a local prehistoric family.

To handle terms $s \cdot t$ in the next chapter an additional rule for modus ponens under $\Box$ is necessary. I therefore introduce here the new rule $(\Box\text{Cut})$ as follows:

Definition 22. [$(\Box\text{Cut})$ Rule]

$$\frac{\Gamma \supset \Delta, \Box A, \Box B \qquad \Gamma \supset \Delta, \Box(A \to B), \Box B}{\Gamma \supset \Delta, \Box B}\ (\Box\text{Cut})$$

Again it is also necessary to expand the definition of correspondence (def. 5) for this rule:

Definition 23. [Correspondence for $(\Box\text{Cut})$]

- The topmost $\Box$ occurrence in the active formulas and the principal formula correspond to each other.
- The subformulas A in the active formulas of the premises correspond to each other.
- The subformulas B correspond to each other.

Notice that with this expansion $\Box$ occurrences of the same family no longer are always part of the same subformula $\Box C$ and therefore lemma 10 no longer holds. Also similar to the (Cut) rule, correspondence is expanded to relate negative and positive occurrences of $\Box$ symbols as A is used with different polarities in the two premises.

With the following lemmas and theorems I will establish a constructive proof for $\text{G3s}+(\Box\,\text{Cut}) \vdash \Gamma \supset \Delta \Rightarrow \text{G3s}+(\text{Cut}) \vdash \Gamma \supset \Delta \Rightarrow \text{G3s} \vdash \Gamma \supset \Delta$. Moreover there will be corollaries showing that the constructions do not introduce prehistoric cycles by the new definition 21. As all prehistoric relations by the first definition 17 are included in the new definition, the final proof in G3s will be prehistoric-cycle-free by any definition if the original proof G3s + $(\Box\text{Cut})$ was prehistoric-cycle-free by the new definition.

It is important to note, that all the following corollaries are not restricted to the annotations an_T of the proofs $\mathcal{T} = (T, R)$ given by the premise of the lemma but still hold for arbitrary annotations an. That means there is no implicit assumption that the families have only a single occurrence in the root sequents used in the lemma or theorem and the results can also be used in subtrees $T{\upharpoonright}S$ together with an annotation an_T for the complete tree.

Lemma 15. *[Weakening for G3s]* $\text{G3s} \vdash \Gamma \supset \Delta \Rightarrow \text{G3s} \vdash \Gamma, \Gamma' \supset \Delta, \Delta'$

Proof. By structural induction on the proof tree:

1. case: $\Gamma \supset \Delta$ is an axiom. Then also $\Gamma, \Gamma' \supset \Delta, \Delta'$ is an axiom.

2. case: $\Gamma \supset \Delta$ is proven by a rule different from $(\supset \Box)$. Use the induction hypothesis on the premises and append the same rule for a proof of $\Gamma, \Gamma' \supset \Delta, \Delta'$.

3. case: $\Gamma \supset \Delta$ is proven by a $(\supset \Box)$ rule. Add Γ' and Δ' as weakening formulas to the rule for a proof of $\Gamma, \Gamma' \supset \Delta, \Delta'$. ■

Corollary 15.1. *For any annotation* an *the proof for* G3s $\vdash \Gamma, \Gamma' \supset \Delta, \Delta'$ *as constructed in the main proof has the exact same prehistoric relations as the original proof for* G3s $\vdash \Gamma \supset \Delta$.

Proof. $(\supset \Box)$ rules are handled by the 3. case by new $(\supset \Box)$ rules that use the exact same premise and only in the history add the new weakening formulas. So all prehistoric paths are unchanged and all prehistoric relations remain the same. ■

The last corollary also follows from the fact 2.8 in Yu (2017, 787). In that paper Yu looks at prehistoric relations locally, i.e. taking only correspondence up to the current sequent in consideration. That means the graph of prehistoric relations has to be updated going up the proof tree as new rules add new correspondences and therefore unify vertices in the prehistoric relations graph which were still separate in the premise. To work with such changing graphs, Yu introduces the notion of isolated families. He shows that all $\Box$ occurrences introduced by weakening are isolated. That means they have no prehistoric relations themselves, which globally means that they can not add any prehistoric relations from adding correspondences later in the proof. This is exactly what the last corollary asserts.

Lemma 16. *[Inversion for G3s]*

- G3s $\vdash \Gamma \supset \Delta, A \to B \Rightarrow$ G3s $\vdash A, \Gamma \supset \Delta, B$
- G3s $\vdash A \to B, \Gamma \supset \Delta \Rightarrow$ G3s $\vdash \Gamma \supset \Delta, A$ *and* G3s $\vdash B, \Gamma \supset \Delta$
- G3s $\vdash \Box A, \Gamma \supset \Delta \Rightarrow$ G3s $\vdash A, \Box A, \Gamma \supset \Delta$
- G3s $\vdash \Gamma \supset \Delta, \Box A \Rightarrow$ G3s $\vdash \Gamma \supset \Delta, A$

Proof. For any formula C to inverse, we proof the lemma by structural induction on the proof tree:

1. case: C is weakening formula of the last rule. Just weaken in the necessary deconstructed formulas instead.

2. case: C is a side formula of the last rule. By induction hypothesis we can replace C by the necessary deconstructed formulas in the premises and append the same rule to get the necessary proof(s).

3. case: C is the principal formula of the last rule. Then proof(s) of the premise(s) without the last rule is/are already the necessary proof(s). ■

Corollary 16.1. *For any annotation* an *the constructed proofs do not introduce any new prehistoric relations.*

Proof. In the 1st case and 2nd case we only remove occurrences of $\Box$ so no new prehistoric relations are introduced. In the 3rd case, a rule is removed entirely, which again can not introduce new prehistoric relations. ■

Lemma 17. *[Contraction for G3s]*

- $\mathrm{G3s} \vdash A, A, \Gamma \supset \Delta \Rightarrow \mathrm{G3s} \vdash A, \Gamma \supset \Delta$
- $\mathrm{G3s} \vdash \Gamma \supset \Delta, A, A \Rightarrow \mathrm{G3s} \vdash \Gamma \supset \Delta, A$

Proof. By simultaneous induction over the proof tree and the build up of A:

1. case: At least one occurrence of A is a weakening formula of the last rule. Just remove it. Note that this case also covers all axioms.

2. case: Both occurrences of A are side formulas of the last rule. By induction hypothesis the premises of the rule are provable with the two A contracted. Append the same rule for the necessary proof.

3. case: One of the occurrences of A is the principal formula of the last rule, the other is a side formula. Use the inversion lemma (16) on the side formula A in the premises to deconstruct it. Use the induction hypothesis to contract the deconstructed parts of A. Append the same last rule without A as side formula to get the necessary proof. ∎

Corollary 17.1. *For any annotation* an *the constructed proofs do not introduce any new prehistoric relations.*

Proof. In the 1st case and 2nd case we only remove occurrences of $\Box$ so no new prehistoric relations are introduced. In the 3rd case, by corollary 16.1 no new prehistoric relations are introduced for the new proof where both occurrences of A are deconstructed. Moreover, in the case of appending a $(\supset \Box)$ rule, all occurrences in the new premise are also in the old premise and therefore no new prehistoric relations get introduced. ∎

Lemma 18. $\mathrm{G3s} \vdash B, \Box B, \Gamma \supset \Delta \Leftrightarrow \mathrm{G3s} \vdash B, \Gamma \supset \Delta$

Proof. The $(\Leftarrow)$ direction is just a weakening. The $(\Rightarrow)$ direction is shown by a structural induction over the proof tree for $B, \Box B, \Gamma \supset \Delta$:

1. case: $\Box B$ is a weakening formula of the last rule. Then removing it keeps the proof intact.

2. case: $\Box B$ is a side formula of the last rule. By induction hypothesis the premises of the rules are provable without $\Box B$. Append the same rule to get a proof of $B, \Gamma \supset \Delta$.

3. case: $\Box B$ is the principal formula of the last rule, then the premise is $B, B, \Box B, \Gamma \supset \Delta$. By induction hypothesis we get a proof for $B, B, \Gamma \supset \Delta$ and by contraction we get $B, \Gamma \supset \Delta$. ∎

Corollary 18.1. *For any annotation* an *the constructed proof does not introduce any new prehistoric relations.*

Proof. The new proof is the old proof with $\Box B$ removed and $\Box \supset$ rules with $\Box B$ as principal formula replaced by contractions, which do not introduce new prehistoric relations by corollary 17.1. So the new proof can not introduce any new prehistoric relations. ∎

Theorem 8. *[Cut Elimination for G3s] If* $\mathrm{G3s} \vdash \Gamma \supset \Delta, A$ *and* $\mathrm{G3s} \vdash A, \Gamma \supset \Delta$ *then* $\mathrm{G3s} \vdash \Gamma \supset \Delta$.

Proof. By a simultaneous induction over the depths of the proof trees $\mathcal{T}_L$ for $\Gamma \supset \Delta, A$ and $\mathcal{T}_R$ for $A, \Gamma \supset \Delta$ as well as the rank of A (i.e we will use the induction hypothesis to cut with the same formulas but shorter proof trees as well as to cut proof trees with lower rank formulas):

1. case: A is a weakening formula in the last rule of one of the proofs. We get the required proof for $\Gamma \supset \Delta$ by leaving out A from that proof.

2. case: A is a side formula in the last rule of one of the two proofs. We distinguish the following subcases:

2.1 case: A is a side formula in the last rule of $\mathcal{T}_R$, which is not a $(\subset \Box)$ rule. By induction hypothesis we can cut the weakened premises $A, \Gamma_i, \Gamma \supset \Delta_i, \Delta$ of that rule with a weakened $\Gamma_i, \Gamma \supset \Delta_i, \Delta, A$ proven by $\mathcal{T}_L$ to get $\Gamma_i, \Gamma \supset \Delta_i, \Delta$. Applying the same rule we get the a proof for $\Gamma, \Gamma \supset \Delta, \Delta$. By contraction we get a proof for $\Gamma \supset \Delta$.

2.2 case: A is a side formula in the last rule of $\mathcal{T}_L$. This case is handled symmetrical to the previous one. Notice that the last rule can not be a $(\supset \Box)$ rule in this case, as that rule does not have any side formulas on the right.

2.3 case: A is a side formula in the last rule of $\mathcal{T}_R$, which is a $(\supset \Box)$ rule and a principal formula in the last rule of $\mathcal{T}_L$. Then A has the form $\Box A_0$ as it is a side formula of a $(\supset \Box)$ on the right. So the last rule of $\mathcal{T}_L$ is also a $(\supset \Box)$ rule and the proof has the following form:

$$\dfrac{\dfrac{\overset{\mathcal{T}_L}{\Box\Gamma_L \supset A_0}}{\Gamma'_L, \Box\Gamma_L \supset \Delta', \Box B, \Box A_0}(\supset \Box) \qquad \dfrac{\overset{\mathcal{T}_R}{\Box A_0, \Box\Gamma_R \supset B}}{\Gamma'_R, \Box A_0, \Box\Gamma_R \supset \Delta', \Box B}(\supset \Box)}{\Gamma \supset \Delta', \Box B}(\text{Cut})$$

where $\Delta = \Delta', \Box B$ and $\Gamma = \Gamma'_L, \Box\Gamma_L = \Gamma'_R, \Box\Gamma_R$.

The cut can be moved up on the right using weakening as follows:

$$\dfrac{\dfrac{\dfrac{\overset{\mathcal{T}_L}{\Box\Gamma_L \supset A_0}}{\Box\Gamma_R, \Box\Gamma_L \supset B, \Box A_0}(\supset \Box) \qquad \overset{\mathcal{T}'_R}{\Box A_0, \Box\Gamma_R, \Box\Gamma_L \supset B}}{\Box\Gamma_R, \Box\Gamma_L \supset B}(\text{Cut})}{\Gamma, \Box\Gamma_R, \Box\Gamma_L \supset \Delta', \Box B}(\supset \Box)$$

By the induction hypothesis and a contraction we get the required proof for $\Gamma \supset \Delta$ as $\Box\Gamma_L \subseteq \Gamma$ and $\Box\Gamma_R \subseteq \Gamma$.

3. case: A is the principal formula in the last rules of $\mathcal{T}_L$ and $\mathcal{T}_R$. Then we have the following subcases:

3.1: The last rules are axioms. Then A is atomic and $A \in \Delta$ and $A \in \Gamma$ as there is no axiom with a principal $\perp$ on the right. Therefore also $\Gamma \supset \Delta$ is an axiom.

3.2: A has the form $A_0 \to A_1$. Then the proof has the following form:

$$\dfrac{\dfrac{\overset{\mathcal{T}_L}{A_0, \Gamma \supset \Delta, A_1}}{\Gamma \supset \Delta, A_0 \to A_1}(\supset\to) \qquad \dfrac{\overset{\mathcal{T}_{R1}}{\Gamma \supset \Delta, A_0} \qquad \overset{\mathcal{T}_{R2}}{A_1, \Gamma \supset \Delta}}{A_0 \to A_1, \Gamma \supset \Delta}(\to\supset)}{\Gamma \supset \Delta}(\text{Cut})$$

Using weakening and two cuts with the lower rank formulas A_0 and A_1 we can transform that into:

$$\dfrac{\dfrac{\begin{array}{c}\mathcal{T}'_{R1}\\ \Gamma \supset \Delta, A_1, A_0\end{array} \qquad \begin{array}{c}\mathcal{T}_L\\ A_0, \Gamma \supset \Delta, A_1\end{array}}{\Gamma \supset \Delta, A_1}\,(\text{Cut}) \qquad \begin{array}{c}\mathcal{T}_{R2}\\ A_1, \Gamma \supset \Delta\end{array}}{\Gamma \supset \Delta}\,(\text{Cut})$$

Using the induction hypothesis we get the required cut-free proof for $\Gamma \supset \Delta$.

3.3: A has the form $\Box A_0$. Then the proof has the following form:

$$\dfrac{\dfrac{\begin{array}{c}\mathcal{T}_L\\ \Box\Gamma_0 \supset A_0\end{array}}{\Gamma_1, \Box\Gamma_0 \supset \Delta, \Box A_0}\,(\supset \Box) \qquad \dfrac{\begin{array}{c}\mathcal{T}_R\\ A_0, \Box A_0, \Gamma \supset \Delta\end{array}}{\Box A_0, \Gamma \supset \Delta}\,(\Box \supset)}{\Gamma \supset \Delta}\,(\text{Cut})$$

From the lemma 18, we get a proof $\mathcal{T}'_R$ for $A_0, \Gamma \supset \Delta$ and by weakening we get a proof $\mathcal{T}'_L$ for $\Gamma \supset \Delta, A_0$. From this and using a cut with the lower rank formula A_0 we get the following proof:

$$\dfrac{\begin{array}{c}\mathcal{T}'_L\\ \Gamma \supset \Delta, A_0\end{array} \qquad \begin{array}{c}\mathcal{T}'_R\\ A_0, \Gamma \supset \Delta\end{array}}{\Gamma \supset \Delta}\,(\text{Cut})$$

Using the induction hypothesis we get the required cut-free proof for $\Gamma \supset \Delta$. ■

Corollary 8.1. *For any annotation* an *the constructed proof for* $\Gamma \supset \Delta$ *only introduces new prehistoric relations* $i \prec j$ *between families* $\Box_i$ *and* $\Box_j$ *occurring in* $\Gamma \supset \Delta$ *where there exists a family* $\Box_k$ *in* A *such that* $i \prec k \prec j$ *in the original proof.*

Proof. The used weakenings and contractions do not introduce any new prehistoric relations by the corollaries 15.1 and 17.1. Also leaving out formulas as in case 1 and 3.1, as well as rearranging rules which are not $(\supset \Box)$ rules as in case 3.2 do not introduce any new prehistoric relations. Finally the use of lemma 18 in case 3.3 does not introduce any new prehistoric relations by corollary 18.1 and leaving out the $(\supset \Box)$ rule also can not introduce any new prehistoric relations.

So the only place where new prehistoric relations get introduced is by the new $(\supset \Box)$ in case 2.3. All prehistoric relations from $\Box\Gamma_R$ are already present from the $(\supset \Box)$ rule on the right in the original proof. So only prehistoric relations from $\Box\Gamma_L$ are new. For all families $\Box_i$ in $\Box\Gamma_L$ we have $i \prec k$ for the family $\Box_k$ in the cut formula introduced by the $(\supset \Box)$ rule on the left. Moreover $k \prec j$ for the same family because of the occurrence of $\Box A_0$ on the right. ■

Corollary 8.2. *For any annotation* an *the constructed proof for* $\Gamma \supset \Delta$ *does not introduce prehistoric cycles.*

Proof. Assume for contradiction that there exists a prehistoric cycle $i_0 \prec \ldots \prec i_{n-1} \prec i_0$ in the new proof. By the previous lemma for any prehistoric relation $i_k \prec i_{k+1 \bmod n}$ in the cycle either $i_k \prec i_{k+1 \bmod n}$ in the original proof or there is a family i'_k in the cut formula such that $i_k \prec i'_k \prec i_{k+1 \bmod n}$ in the original proof. Therefore we also have a prehistoric cycle in the original proof. ■

Theorem 9. *[($\Box$Cut) Elimination] If* G3s $\vdash \Gamma \supset \Delta, \Box A, \Box B$ *and* G3s $\vdash \Gamma \supset \Delta, \Box(A \to B), \Box B$ *then* G3s $\vdash \Gamma \supset \Delta, \Box B$

Proof. By a structural induction over the proof trees $\mathcal{T}_L$ for $\Gamma \supset \Delta, \Box A, \Box B$ and $\mathcal{T}_R$ for $\Gamma \supset \Delta, \Box(A \to B), \Box B$.

1. case: $\Box(A \to B)$ or $\Box A$ is a weakening formula of the last rule. Then removing them from that proof gives the required proof. This includes the case when $\Box B$ is the principal formula of the last rule of either proof, as then the last rule is $(\supset \Box)$ which has no side formulas on the right.

2. case: $\Box(A \to B)$ or $\Box A$ is a side formula of the last rule. Then also $\Box B$ is a side formula of that rule. Use the induction hypothesis on the premises of that rule with the other proof and append the same rule.

3. case: $\Box(A \to B)$ and $\Box A$ are the principal formula of the last rule. Then the last rules have the following form:

$$\dfrac{\dfrac{\begin{matrix}\mathcal{T}_L \\ \Box\Gamma_L \supset A\end{matrix}}{\Gamma'_L, \Box\Gamma_L \supset \Delta, \Box A, \Box B}(\supset \Box) \qquad \dfrac{\begin{matrix}\mathcal{T}_R \\ \Box\Gamma_R \supset A \to B\end{matrix}}{\Gamma'_R, \Box\Gamma_R \supset \Delta, \Box(A \to B), \Box B}(\supset \Box)}{\Gamma \supset \Delta, \Box B}(\Box\text{Cut})$$

where $\Delta = \Delta', \Box B$ and $\Gamma = \Gamma'_L, \Box\Gamma_L = \Gamma'_R, \Box\Gamma_R$.

By inversion for $(\supset\to)$ we get a proof $\mathcal{T}'_R$ for $A, \Box\Gamma_R \supset B$ from the first premise $\Box\Gamma_R \supset A \to B$. Using weakening and a normal cut on the formula A we get the following proof:

$$\dfrac{\dfrac{\begin{matrix}\mathcal{T}'_L \\ \Box\Gamma_L, \Box\Gamma_R \supset A\end{matrix} \qquad \begin{matrix}\mathcal{T}''_R \\ A, \Box\Gamma_L, \Box\Gamma_R \supset B\end{matrix}}{\Box\Gamma_L, \Box\Gamma_R \supset B}(\text{Cut})}{\Gamma, \Box\Gamma_L, \Box\Gamma_R \supset \Delta, \Box B}(\supset \Box)$$

By contraction and a cut elimination we get the required G3s proof for $\Gamma \supset \Delta, \Box B$ as $\Box\Gamma_L \subseteq \Gamma$ and $\Box\Gamma_R \subseteq \Gamma$. ∎

Corollary 9.1. *For any annotation* an *the constructed proof for* $\Gamma \supset \Delta, \Box B$ *does not introduce prehistoric cycles.*

Proof. Removing weakening or side formulas $\Box(A \to B)$ or $\Box A$ as in case 1 and 2 does not introduce new prehistoric relations.

Any prehistoric relation because of the new $(\supset \Box)$ rule in case 3 already exists in the original proof, as every $\Box$ occurrence in $\Box\Gamma_L$ or $\Box\Gamma_R$ also occurs in one of the two $(\supset \Box)$ rules in the original proof, which both introduce a $\Box$ of the same family as $\Box B$ by the definition of correspondence for ($\Box$Cut) (def. 23).

So the new proof with ($\Box$Cut) rules replaced by (Cut) rules does not introduce new prehistoric relations and therefore also no new prehistoric cycles. By corollary 8.2, the cut elimination to get a G3s proof does not introduce prehistoric cycles. ∎

Definition 24. The cycle-free fragment of a system Y, denoted by $Y^{\otimes}$, is the collection of all sequents that each have a prehistoric-cycle-free Y-proof (Yu 2017, 787).

Theorem 10. *The cycle-free fragments of G3s + ($\Box$Cut), G3s + (Cut) and G3s are identical.*

Proof. A prehistoric-cycle-free proof in G3s by the original definition (def. 17) is also prehistoric-cycle-free by the new definition (def. 21) as a negative family can not have any prehistoric families itself in a G3s-proof . So any sequent $\Gamma \subset \Delta \in G3s^{\otimes}$ is trivially also provable prehistoric-cycle-free in G3s + (Cut) and G3s + ($\Box$Cut) and we have $\mathrm{G3s}^{\otimes} \subseteq (\mathrm{G3s}+(\Box\,\mathrm{Cut}))^{\otimes}$ and $\mathrm{G3s}^{\otimes} \subseteq (\mathrm{G3s}+(\mathrm{Cut}))^{\otimes}$. Moreover $(\mathrm{G3s}+(\mathrm{Cut}))^{\otimes} \subseteq \mathrm{G3s}^{\otimes}$ by corollary 8.2 and $(\mathrm{G3s}+(\Box\,\mathrm{Cut}))^{\otimes} \subseteq (\mathrm{G3s}+(\mathrm{Cut}))^{\otimes} \subseteq \mathrm{G3s}^{\otimes}$ by corollary 9.1. All together we get:

$\mathrm{G3s}^{\otimes} = (\mathrm{G3s}+(\mathrm{Cut}))^{\otimes} = (\mathrm{G3s}+(\Box\,\mathrm{Cut}))^{\otimes}$ ■

In the theorem 2.21 Yu (2017, 793) shows that non-self-referentiality is not normal in T, K4 and S4. The results in this chapter hint at an explanation for this fact for S4 and at the possibility to still use modus ponens with further restrictions in the non-self-referential subset of S4. Namely, to consider the global aspects of self-referentiality coming from correspondence of occurrences, it is necessary when combining two proofs, that the two proofs together with the correct correspondences added are prehistoric-cycle-free. So we can only use modus ponens on two non-self-referential S4 theorems A and $A \to B$ if there are proofs of A and $A \to B$ such that the prehistoric relations of these proofs combined, together with identifying the occurrences of A in both proofs, are prehistoric-cycle-free. In that case we get a prehistoric-cycle-free G3s proof for B using cut elimination and corollary 8.1, which shows that B is also non-self-referential.

4.2 G3lp

Cornelia Pulver (2010, 62) introduces the system LPG3 by expanding G3c with rules for the build up of justification terms as well as the new axioms (Axc) and (Axt). To ensure that the contraction lemma holds, all rules have to be invertible (Pulver 2010, 61) which is the reason why contracting variants of all the justification rules are used for LPG3. Our variant G3lp will use the same rules to build up terms, but replace the axioms with rules $(\supset:)_c$ and $(\supset:)_t$ to keep the prehistoric relations of the proof intact. As there is a proof for $\supset A$ for any axiom A and also for $A \supset A$ for any formula A, these two rules are equivalent to the two axioms and invertible.

As we already did with G3s, we will use the full system with all classical operators for examples, but only the minimal subset with $\to$ and $\perp$ for proofs. So these two systems use the classical rules from G3s in figures 2.1 and 2.2 as well as the new LP rules in figure 4.1.

This system is adequate for the logic of proofs LP as shown in corollary 4.37 in Pulver (2010, 73). It also allows for weakening, contraction and inversion (Pulver 2010, 65ff). By corollary 4.36 in the same paper, G3lp without the $(\supset:)_c$ rule is equivalent to LP_0. Neither Pulver (2010) nor Artemov (2001) define Gentzen systems for a restricted logic of proofs LP(CS), perhaps because

Figure 4.1: G3lp

$$\frac{\supset A}{\Gamma \supset \Delta, c{:}A}\ (\supset:)_c\ (A \text{ an axiom of LP}) \qquad \frac{t{:}A \supset A}{t{:}A, \Gamma \supset \Delta, t{:}A}\ (\supset:)_t$$

$$\frac{A, t{:}A, \Gamma \supset \Delta}{t{:}A, \Gamma \supset \Delta}\ (:\supset) \qquad \frac{\Gamma \supset \Delta, t{:}A, !t{:}t{:}A}{\Gamma \supset \Delta, !t{:}t{:}A}\ (\supset!)$$

$$\frac{\Gamma \supset \Delta, s{:}A, t{:}A, (s+t){:}A}{\Gamma \supset \Delta, (s+t){:}A}\ (\supset +) \qquad \frac{\Gamma \supset \Delta, s{:}(A \to B), s \cdot t{:}B \qquad \Gamma \supset \Delta, t{:}A, s \cdot t{:}B}{\Gamma \supset \Delta, s \cdot t{:}B}\ (\supset \cdot)$$

it seems obvious that restricting whatever rule is used for introducing proof constants to CS gives a Gentzen system for LP(CS).

To work with prehistoric relations in G3lp proofs we need the following new or adapted definitions:

Definition 25. [Subformula] The set of subformulas sub(A) of a LP formula A is inductively defined as follows:

1. $\mathrm{sub}(P) = \{P\}$ for any atomic formula P
2. $\mathrm{sub}(\bot) = \{\bot\}$
3. $\mathrm{sub}(A_0 \to A_1) = \mathrm{sub}(A_0) \cup \mathrm{sub}(A_1) \cup \{A_0 \to A_1\}$
4. $\mathrm{sub}(s + t{:}A_0) = \mathrm{sub}(A_0) \cup \{s{:}A_0, t{:}A_0, s + t{:}A_0\}$
5. $\mathrm{sub}(t{:}A_0) = \mathrm{sub}(A_0) \cup \{t{:}A_0\}$

Definition 26. [Subterm] The set of subterms sub(t) of a LP justification term t is inductively defined as follows:

1. $\mathrm{sub}(x) = \{x\}$ for any variable x
2. $\mathrm{sub}(c) = \{c\}$ for any constant c
3. $\mathrm{sub}(!t) = \mathrm{sub}(t) \cup \{!t\}$
4. $\mathrm{sub}(s + t) = \mathrm{sub}(s) \cup \mathrm{sub}(t) \cup \{s + t\}$
5. $\mathrm{sub}(s \cdot t) = \mathrm{sub}(s) \cup \mathrm{sub}(t) \cup \{s \cdot t\}$

The set of subterms sub(A) of a LP formula A is the union of all sets of subterms for all terms occurring in A.

We use the symbol sub for all definitions of subterms and subformulas, as it will be clear from context which of the definitions is meant. Notice that by this definition $s{:}A$ is a subformula of $s + t{:}A$.

We expand the definition of correspondence (def. 5) to G3lp proofs as follows:

Definition 27. [Correspondence in G3lp] All topmost terms in active or principal formulas in the rules $(\supset \cdot)$, $(\supset +)$ $(\supset!)$ and $(:\supset)$ correspond to each other.

Notice that in the $(\supset!)$ rule, the topmost term t in the contraction formula therefore corresponds to the topmost proof term $!t$ in the principal formula.

The term t of the other active formula $!t{:}t{:}A$ on the other hand corresponds to the same term t in the principal formula.

By this definition, families of terms in G3lp consist not of occurrences of a single term t but of occurrences of subterms s of a top level term t. We will use $\bar{t}$ for the family of occurrences corresponding to the *top level* term t, i.e. seen as a set of terms instead of term occurrences we have $\bar{t} \subseteq \mathrm{sub}(t)$. So for any term occurrence s, $\bar{s}$ is not necessarily the full family of s in the complete proof tree as s could be a subterm of the top level term t of the family. For any occurrence s in a sequent S of the proof tree though, $\bar{s}$ is the family of s relative to the subtree $T{\upharpoonright}S$ as all related terms in the premises of G3lp rules are subterms of the related term in the conclusion.

We also see that most rules of G3lp only relate terms to each other used for the same subformula A. The two exceptions are the $(\supset \cdot)$ rule and the $(\supset !)$ rule. Similar to the cut rules from the previous chapter, $(\supset \cdot)$ relates subformulas and symbols of different polarities as well as terms used for different formulas. So we will use the same approach to define prehistoric relations of term families for any polarity:

Definition 28. [Prehistoric Relation in G3lp] A family $\bar{t_i}$ has a *prehistoric relation* to another family $\bar{t_j}$, in notation $i \prec j$, if there is a $(\supset :)$ rule introducing an occurrence belonging to $\bar{t_j}$ with premise S, such that there is an occurrence belonging to $\bar{t_i}$ in S.

Given that we now have defined families of terms and prehistoric relations between them in G3lp, it is interesting to see what happens with this relations if we look at the forgetful projection of a G3lp proof. That is, what happens on the G3s side if we construct a proof tree with the forgetful projections of the original sequents. Of course we do not get a pure G3s proof as most of the G3lp rules have no direct equivalent in G3s. We will therefore define new rules, which are the forgetful projection of a G3lp rule denoted for example by $(\supset !)^{\circ}$ for the forgetful projection of a $(\supset !)$ rule. The following two lemmas show that all this new rules are admissible in G3s + $(\Box\,\mathrm{Cut})$.

Lemma 19. $\mathrm{G3lp} \vdash \Gamma \supset \Delta, \Box A$ *iff* $\mathrm{G3lp} \vdash \Gamma \supset \Delta, \Box\Box A$.

Proof. The $(\Leftarrow)$ direction is just inversion for $(\supset \Box)$. The $(\Rightarrow)$ direction is proven by the following structural induction:

1. case: $\Box A$ is a weakening formula of the last rule. Just weaken in $\Box\Box A$.

2. case: $\Box A$ is a side formula of the last rule. Use the induction hypothesis on the premises and append the same last rule.

3. case: $\Box A$ is the principal formula of the last rule. Then the last rule is a $(\supset \Box)$ rule and has the following form:

$$\frac{\Box\Gamma \supset A}{\Gamma', \Box\Gamma \supset \Delta, \Box A}\,(\supset \Box)$$

Use an additional $(\supset \Box)$ rule to get the necessary proof as follows:

$$\frac{\dfrac{\Box\Gamma \supset A}{\Box\Gamma \supset \Box A}\,(\supset \Box)}{\Gamma', \Box\Gamma \supset \Delta, \Box\Box A}\,(\supset \Box)$$

■

Lemma 20. *The forgetful projection of all rules in G3lp are admissible in G3s + ($\Box$Cut).*

Proof. The subset G3c is shared by G3lp and G3s and is therefore trivially admissible. The forgetful projection of the rule ($\supset +$) is just a contraction and therefore also admissible. The forgetful projection of the rules $(\supset:)_t$ and $(\supset:)_c$ are $(\supset \Box)$ rules in G3s. The forgetful projection of $(\supset \cdot)$ is a ($\Box$ Cut). Finally the forgetful projection of a $(\supset!)$ rule has the following form:

$$\frac{\Gamma \supset \Delta, \Box A, \Box\Box A}{\Gamma \supset \Delta, \Box\Box A} (\supset!)^\circ$$

That rule is admissible by lemma 19 and a contraction. ■

Instead of working with a G3s system with all this extra rules included, we will define a forgetful projection from a G3lp proof to a G3s + ($\Box$ Cut) proof by eliminating all the contractions using the algorithm implicitly defined in the proof of lemma 17 (contraction) and eliminating the $(\supset!)^\circ$ rules by the algorithm implicitly described in the proof for lemma 19.

For the following lemmas and proofs we fix an arbitrary G3lp proof $\mathcal{T} = (T, R)$ and its forgetful projection $\mathcal{T}^\circ = (T', R')$ as defined below.

Definition 29. [Forgetful Projection of a G3lp Proof] The forgetful projection of a G3lp proof $\mathcal{T} = (T, R)$ for a LP sequent $\Gamma \supset \Delta$ is the G3s + ($\Box$ Cut) proof $\mathcal{T}^\circ = (T', R')$ for $\Gamma^\circ \supset \Delta^\circ$ inductively defined as follows:

1. case: The last rule of $\mathcal{T}$ is an axiom. Then $\mathcal{T}^\circ$ is just $\Gamma^\circ \supset \Delta^\circ$ which is an axiom of G3s.

2. case: The last rule of $\mathcal{T}$ is a $(\supset\rightarrow)$ or a $(\rightarrow\supset)$ rule with premises S_i. Then $\mathcal{T}^\circ$ has the same last rule with $(\mathcal{T}{\upharpoonright}S_i)^\circ$ as proofs for the premises S_i°.

3. case: The last rule of $\mathcal{T}$ is a $(\supset:)_t$ or $(\supset:)_t$ rule with premise S. Then $\mathcal{T}^\circ$ has a $(\supset \Box)$ as last rule with $(\mathcal{T}{\upharpoonright}S)^\circ$ as proof for the premise S°.

4. case: The last rule of $\mathcal{T}$ is a $(\supset +)$ rule with premise S. Then $\mathcal{T}^\circ$ is $(\mathcal{T}{\upharpoonright}S)^\circ$ with the necessary contraction applied.

5. case: The last rule of $\mathcal{T}$ is a $(\supset \cdot)$ rule with premises S_0 and S_1. Then $\mathcal{T}^\circ$ has a ($\Box$ Cut) as last rule with $(\mathcal{T}{\upharpoonright}S_i)^\circ$ as proofs for the premises S_i°.

6. case: The last rule of $\mathcal{T}$ is a $(\supset!)$ rule with premise S. Then we get a G3s + ($\Box$ Cut) proof for $\Gamma^\circ \supset \Delta^\circ, \Box\Box A$ from the proof $(\mathcal{T}{\upharpoonright}S)^\circ$ by lemma 19. $\mathcal{T}^\circ$ is that proof with the additional $\Box\Box A$ removed by contraction as $\Box\Box A \in \Delta^\circ$.

To reason about the relations between a G3lp proof $\mathcal{T}$ and its forgetful projection $\mathcal{T}^\circ$, the following algorithm to construct $\mathcal{T}^\circ$ is useful:

1. Replace all sequents by their forgetful projection.
2. Add the additional $(\supset \Box)$ rules and prepend additional $\Box$ where necessary, so that the forgetful projections of $(\supset!)$ reduce to simple contractions.
3. Eliminate all contractions to get a G3s + ($\Box$ Cut) proof.

It is not immediately clear that contracting formulas only removes occurrences as the proof uses inversion which in turn also adds weakening formulas. But all the deconstructed parts weakened in this way get contracted again in the next step of the contraction. In the end the contracted proof tree is always a subset of the original proof tree.

That means that also $\mathcal{T}^\circ$ is a subset of the tree constructed in step 2 of the algorithm. From this we see that all $\Box$ occurrences in $\mathcal{T}^\circ$ have a term occurrence in $\mathcal{T}$ mapped to them if we consider the extra $\Box$ occurrences introduced in step 2 (resp. in case 6 of the definition) as replacements of the same term as the $\Box$ occurrences they are contracted with and also consider the extra sequents $\Box\Gamma \supset \Box A$ introduced in step 2 as copies of the same formulas in the original sequent $\Gamma', \Box\Gamma \supset \Delta, \Box A$ derived by the original $(\supset \Box)$ rule.

Lemma 21. *For any family f_i of $\Box$ occurrences in $\mathcal{T}^\circ$ there is a unique proof term family $\bar{t}_{\tilde{i}}$ in $\mathcal{T}$ such that $s \in \bar{t}_{\tilde{i}}$ for all proof term occurrences s mapped to $\Box$ occurrences in f_i.*

Proof. For any two directly corresponding $\Box$ occurrences we show that the two mapped term occurrences correspond directly or by reflexive closure:

1. case: The two $\Box$ occurrences are added in step 2 of the algorithm. Then the mapped term occurrences are the same occurrence and correspond by reflexive closure.

2. case: The two $\Box$ occurrence correspond directly by a rule which is the forgetful projection of a rule in $\mathcal{T}$. Then the mapped term occurrences also correspond as all G3lp rules with a direct equivalent in G3s have the same correspondences. Notice that lemma 19 only removes weakening formulas from existing $(\supset \Box)$ rules. So this still holds for $(\supset \Box)$ rules and their corresponding $(\supset:)$ rules even after applying lemma 19.

3. case: The two $\Box$ occurrences correspond directly by a $(\supset \Box)$ rule added in step 2 of the algorithm. Then the rule together with the previous rule has the following form:

$$\dfrac{\dfrac{\Box\Gamma \supset A}{\Box\Gamma \supset \Box A}\,(\supset \Box)}{\Gamma', \Box\Gamma \supset \Delta, \Box\Box A}\,(\supset \Box)$$

As the formulas in $\Box\Gamma \supset \Box A$ are considered copies of the original sequent $\Gamma', \Box\Gamma \supset \Delta, \Box A$, and the sequent $\Gamma', \Box\Gamma \supset \Delta, \Box\Box A$ is considered the same sequent with an additional $\Box$ symbol, the mapped term occurrences are actually the same and therefore correspond by reflexive closure.

As direct correspondence in the G3s proof is a subset of correspondence in the G3lp proof, so is its transitive and reflexive closure. So for any two corresponding $\Box$ occurrences of a family f_i the mapped term occurrences also correspond and therefore belong to the same family $\bar{t}_{\tilde{i}}$. ■

Lemma 22. *If $i \prec j$ in $\mathcal{T}^\circ$ then either $\tilde{i} = \tilde{j}$ or $\tilde{i} \prec \tilde{j}$ in $\mathcal{T}$ for the term families $\bar{t}_{\tilde{i}}$ and $\bar{t}_{\tilde{j}}$ from the previous lemma.*

Proof. $i \prec j$ in $\mathcal{T}^\circ$, so there is a $(\supset \Box)$ rule in $\mathcal{T}^\circ$ introducing an occurrence $\Box_j$ of f_j with an occurrence $\Box_i$ of f_i in the premise. For the mapped term occurrences s_i and s_j in $\mathcal{T}$ we have $s_i \in \bar{t}_{\tilde{i}}$ and $s_j \in \bar{t}_{\tilde{j}}$ by the previous lemma. From this it follows that $\tilde{i} \prec \tilde{j}$ or $\tilde{i} = \tilde{j}$ by an induction on the proof height:

1. case: The ($\supset \Box$) rule is the forgetful projection of a ($\supset$:) rule. Then we have $\tilde{i} \prec \tilde{j}$ directly by the definition of prehistoric relations for G3lp proofs using the occurrences s_i in the premise of the rule ($\supset$:) introducing the occurrence s_j.

2. case: The ($\supset \Box$) rule is added in step 2 of the algorithm. Then the rule together with the previous rule has the following form:

$$\dfrac{\dfrac{\Box\Gamma \supset A}{\Box\Gamma \supset \Box_k A}\ (\supset \Box)}{\Gamma', \Box\Gamma \supset \Delta, \Box_j \Box_k A}\ (\supset \Box)$$

For the term occurrence s_k mapped to the occurrence $\Box_k$ we have $s_j = !s_k$ and $s_k \in \bar{t}_{\tilde{j}}$ as s_j is the top level term of the principal formula of a ($\supset$!) rule. If the occurrence $\Box_i$ is the occurrence $\Box_k$ then $\tilde{i} = \tilde{j}$ and we are finished. If the occurrence $\Box_i$ is not the occurrence $\Box_k$ then there is a corresponding occurrence $\Box'_i$ with a corresponding mapped term s'_i in the sequent $\Box\Gamma \supset A$ and we have $i \prec k$ from the previous ($\supset \Box$). As $\bar{t}_{\tilde{j}}$ is also the term family of s_k we get $\tilde{i} \prec \tilde{j}$ or $\tilde{i} = \tilde{j}$ by induction hypothesis on the shorter proof up to the that ($\supset \Box$) rule with the occurrences $\Box'_i$, s'_i, $\Box_k$ and s_k. ∎

Corollary 22.1. *If $\mathcal{T}$ is prehistoric-cycle-free then also $\mathcal{T}^\circ$ is prehistoric-cycle-free.*

Proof. The contraposition follows directly from the lemma as for any cycle $i_0 \prec ... \prec i_n \prec i_0$ in $\mathcal{T}^\circ$ we get a cycle in $\mathcal{T}$ by removing duplicates in the list $\tilde{i}_0, ..., \tilde{i}_n$ of mapped term families $\bar{t}_{\tilde{i}_0}, ...\bar{t}_{\tilde{i}_n}$. ∎

We will now come back to our example formula $\neg\Box(P \wedge \neg\Box P)$ from chapter 3.1. Figure 4.2 contains a proof of the same realization $\neg x{:}(P \wedge \neg t \cdot x{:}P)$ in G3lp as well as the forgetful projection of that proof in G3s + ($\Box$Cut). For simplicity we assumed that $(A \wedge B \rightarrow A)$ is an axiom A0 and therefore t is a constant.

This proofs display the logical dependencies making the formula self-referential in quite a different way than the original G3s proof in figure 3.1. There are 3 families of $\Box$ in the G3s + ($\Box$Cut) proof. Two are the same families as in the G3s proof, occur in the root sequent and have a consistent polarity throughout the proof. I therefore simply use the symbols $\boxplus$ and $\boxminus$ for this families. The third one is part of the cut formula and therefore does not occur in the final sequent and does not have consistent polarity throughout the proof. I use $\Box$ for occurrences of this family in the proof.

All left prehistoric relations of the proof are from left branch of the cut where we have $\boxminus \prec_L \boxplus$ and the cycle $\boxplus \prec_L \boxplus$. Other than in the G3s proof, the two $\boxplus$ occurrences are used for different formulas P and $P \wedge \neg\Box P$ and the connection between the two is established by the ($\Box$Cut) with $\Box(P \wedge \neg\Box P \rightarrow P)$. A similar situation is necessary for any prehistoric cycle in a G3lp proof as we will show formally.

Lemma 23. *All occurrences belonging to a term family $\bar{t}$ in a premise S of any ($\supset$:) rule are occurrences of the top level term t itself.*

Proof. All G3lp rules only relate different terms if they are top level terms on the right. All occurrences of $s \in \bar{t}$ in a premise S of a ($\supset$:) rule correspond either

Figure 4.2: G3lp proof

$$
\dfrac{
\dfrac{\dfrac{\dfrac{\dfrac{
P, \neg t\cdot x{:}P, x{:}(P\wedge\neg t\cdot x{:}P) \supset P
\qquad
\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{P, t\cdot x{:}P \supset P}{t\cdot x{:}P \supset P}\,(:\supset)}{P, t\cdot x{:}P, x{:}(P\wedge\neg t\cdot x{:}P) \supset t\cdot x{:}P}\,(\supset:)_t}{P, x{:}(P\wedge\neg t\cdot x{:}P) \supset t\cdot x{:}P, \neg t\cdot x{:}P}\,(\supset\neg)}{P, \neg t\cdot x{:}P, x{:}(P\wedge\neg t\cdot x{:}P) \supset \neg t\cdot x{:}P}\,(\neg\supset)}{}
}{P, \neg t\cdot x{:}P, x{:}(P\wedge\neg t\cdot x{:}P) \supset P\wedge\neg t\cdot x{:}P}\,(\supset\wedge)}{P\wedge\neg t\cdot x{:}P, x{:}(P\wedge\neg t\cdot x{:}P) \supset P\wedge\neg t\cdot x{:}P}\,(\wedge\supset)}{x{:}(P\wedge\neg t\cdot x{:}P) \supset P\wedge\neg t\cdot x{:}P}\,(:\supset)}{P, x{:}(P\wedge\neg t\cdot x{:}P) \supset x{:}(P\wedge\neg t\cdot x{:}P), t\cdot x{:}P}\,(\supset:)_t
\qquad
\dfrac{\dfrac{\dfrac{P, \neg t\cdot x{:}P \supset P}{P\wedge\neg t\cdot x{:}P \supset P}\,(\wedge\supset)}{\supset P\wedge\neg t\cdot x{:}P \to P}\,(\supset\to)}{P, x{:}(P\wedge\neg t\cdot x{:}P) \supset t{:}(P\wedge\neg t\cdot x{:}P \to P), t\cdot x{:}P}\,(\supset:)_c
}{
\dfrac{\dfrac{\dfrac{\dfrac{P, x{:}(P\wedge\neg t\cdot x{:}P) \supset t\cdot x{:}P}{P, \neg t\cdot x{:}P, x{:}(P\wedge\neg t\cdot x{:}P) \supset}\,(\neg\supset)}{P\wedge\neg t\cdot x{:}P, x{:}(P\wedge\neg t\cdot x{:}P) \supset}\,(\wedge\supset)}{x{:}(P\wedge\neg t\cdot x{:}P) \supset}\,(:\supset)}{\supset \neg x{:}(P\wedge\neg t\cdot x{:}P)}\,(\supset\neg)
}\,(\supset\cdot)
$$

$$
\dfrac{
\dfrac{\dfrac{\dfrac{\dfrac{
P, \neg\boxplus P, \boxminus(P\wedge\neg\boxplus P) \supset P
\qquad
\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{P, \Box P \supset P}{\Box P \supset P}\,(\Box\supset)}{P, \Box P, \boxminus(P\wedge\neg\boxplus P) \supset \boxplus P}\,(\supset\Box)}{P, \boxminus(P\wedge\neg\boxplus P) \supset \boxplus P, \neg\Box P}\,(\supset\neg)}{P, \neg\boxplus P, \boxminus(P\wedge\neg\boxplus P) \supset \neg\Box P}\,(\neg\supset)}{}
}{P, \neg\boxplus P, \boxminus(P\wedge\neg\boxplus P) \supset P\wedge\neg\Box P}\,(\supset\wedge)}{P\wedge\neg\boxplus P, \boxminus(P\wedge\neg\boxplus P) \supset P\wedge\neg\Box P}\,(\wedge\supset)}{\boxminus(P\wedge\neg\boxplus P) \supset P\wedge\neg\Box P}\,(\Box\supset)}{P, \boxminus(P\wedge\neg\boxplus P) \supset \boxplus(P\wedge\neg\Box P), \boxplus P}\,(\supset\Box)
\qquad
\dfrac{\dfrac{\dfrac{P, \neg\Box P \supset P}{P\wedge\neg\Box P \supset P}\,(\wedge\supset)}{\supset P\wedge\neg\Box P \to P}\,(\supset\to)}{P, \boxminus(P\wedge\neg\boxplus P) \supset \boxplus(P\wedge\neg\Box P \to P), \boxplus P}\,(\supset\Box)
}{
\dfrac{\dfrac{\dfrac{\dfrac{P, \boxminus(P\wedge\neg\boxplus P) \supset \boxplus P}{P, \neg\boxplus P, \boxminus(P\wedge\neg\boxplus P) \supset}\,(\neg\supset)}{P\wedge\neg\boxplus P, \boxminus(P\wedge\neg\boxplus P) \supset}\,(\wedge\supset)}{\boxminus(P\wedge\neg\boxplus P) \supset}\,(\Box\supset)}{\supset \neg\boxminus(P\wedge\neg\boxplus P)}\,(\supset\neg)
}\,(\Box\text{Cut})
$$

as part of a strict subformula on the right or as part of a subformula on the left of the conclusion. A formula on the left can only correspond to a subformula on the right as a strict subformula. Therefore all corresponding occurrences of s on the right in the remaining path up to the root are part of a strict subformula and so all corresponding occurrences of s, left or right, in the remaining path are occurrences of the same term s. As t itself is a corresponding occurrence of s in that path, we get $t = s$. ■

Corollary 23.1. *If $i \prec j$ for two term families $\bar{t_i}$ and $\bar{t_j}$ of a G3lp proof then there is $(\supset:)$ rule introducing an occurrence $s \in \bar{t_j}$ in a formula $s{:}A$ such that there is an occurrence of t_i in $s{:}A$ (as a term, not as a family, i.e. the occurrence of t_i is not necessary in $\bar{t_i}$).*

Proof. Follows directly from the lemma and the definition of prehistoric relations for G3lp. ■

The last corollary gives us a close relationship between prehistoric relations in G3lp and occurrences of terms in $(\supset:)$ rules. But it does not differentiate between the two variants $(\supset:)_c$ and $(\supset:)_t$ used for introducing elements from CS and input formulas $t{:}A$. It is therefore necessary to expand the definition of self-referentiality by considering all basic justifications and not only the justification constants:

Definition 30. [Inputs] The *inputs* IN of a G3lp proof are all LP formulas which are the principal formula of a $(\supset:)_t$ or $(\supset:)_c$ rule.

Notice that the used constant specifications CS is a subset of the inputs IN. The interpretation here is that $(\supset:)_t$ introduces arguments to Skolem style functions by proving the trivial identity function $t{:}A \to t{:}A$. So we have two different clearly marked sources of basic proofs in G3lp, on the one hand there are the constants justifying known axioms, on the other hand there are presupposed existing proofs or arguments to proof functions. Based on this expanded notion, we can also expand the definition of self-referentiality to input sets:

Definition 31. [Self-Referential Inputs]

- A input set IN is *directly self-referential* if there is a term t such that $t{:}A(t) \in$ IN.
- A input set IN is *self-referential* if there is a subset $A \subseteq$ IN such that $A := \{t_0{:}A(t_1), ..., t_{n-1}{:}A(t_0)\}$.

With this definitions we finally arrive at a counterpart to Yu's theorem:

Theorem 11. *If the input set IN of a G3lp proof is non-self-referential then the proof is prehistoric-cycle-free.*

Proof. We show the contraposition. Assume there is a prehistoric cycle $i_0 \prec i_1 \prec ... \prec i_{n-1} \prec i_0$. By the corollary 23.1 there exists formulas $s_k{:}A_k$ in IN such that $t_{i_k} \in \mathrm{sub}(A_k)$ and $s_k \in \mathrm{sub}(t_{i_{k'}})$ with $k' := k + 1 \mod n$. From the latter and $t_{i_{k'}} \in \mathrm{sub}(A_{k'})$ follows $s_k \in \mathrm{sub}(A_{k'})$. So $\{s_k{:}A_k \mid 0 \le k < n\} \subseteq$ IN is a self-referential subset of IN. ■

Corollary 11.1. *The forgetful projection A° of a LP formula provable with a non-self-referential input set IN is provable prehistoric-cycle-free in G3s.*

Proof. If $\mathcal{T}$ is a proof of A from non-self-referential inputs IN, then $\mathcal{T}$ is prehistoric-cycle-free as proven above. So by corollary 22.1 $\mathcal{T}^\circ$ is a prehistoric-cycle-free proof of A° in G3s$+(\Box\,\mathrm{Cut})$. Finally there is a prehistoric-cycle-free proof of A° in G3s by corollary 9.1. ■

4.3 Counterexample

The main result of the last chapter does not exactly match Yu's result. I have shown that prehistoric cycles in G3s are sufficient for self-referentiality but only for the expanded definition of self-referentiality considering the set of all inputs IN. The question arises if this expansion is actually necessary. The following counterexample shows that indeed, prehistoric cycles in G3s are not sufficient for needing a self-referential CS.

Lemma 24. *The S4 formula* $A \equiv \Box(P \wedge \neg\Box P \to P) \to \neg\Box(P \wedge \neg\Box P)$ *has a realization in* LPG_0.

Proof. Set $A^r \equiv y{:}(P \wedge \neg y \cdot x{:}P \to P) \to \neg x{:}(P \wedge \neg y \cdot x{:}P)$. We have $y{:}(P \wedge \neg y \cdot x{:}P \to P) \vdash_{\mathrm{LPG}_0} \neg x{:}(P \wedge \neg y \cdot x{:}P)$ by the same derivation as for $\mathrm{LP} \vdash \neg x{:}(P \wedge \neg t \cdot x{:}P)$ replacing the introduction of $t{:}(P \wedge \neg t \cdot x{:}P \to P)$ by the assumption $y{:}(P \wedge \neg y \cdot x{:}P \to P)$ and t by y. So by the deduction theorem $\mathrm{LPG}_0 \vdash y{:}(P \wedge \neg y \cdot x{:}P \to P) \to \neg x{:}(P \wedge \neg y \cdot x{:}P)$.[1] ■

Lemma 25. *The S4 formula* $\Box(P \wedge \neg\Box P \to P) \to \neg\Box(P \wedge \neg\Box P)$ *has no prehistoric-cycle-free proof.*

Proof. By inversion for G3s in one direction and an easy deduction in the other, we have $G3s \vdash\supset \Box(P \wedge \neg\Box P \to P) \to \neg\Box(P \wedge \neg\Box P)$ iff $G3s \vdash \Box(P \wedge \neg\Box P \to P), \Box(P \wedge \neg\Box P) \supset$. In both directions the proofs remain prehistoric-cycle-free if the other proof was prehistoric-cycle-free. For a proof of $\Box(P \wedge \neg\Box P \to P), \Box(P \wedge \neg\Box P) \supset$ we have two possibilities for the last rule:

1. case: The last rule is a $(\Box\supset)$ rule with $\Box(P \wedge \neg\Box P \to P)$ as the principal formula. Then the following proof tree shows that we need a proof for the sequent $P, \Box(P \wedge \neg\Box P \to P), \Box(P \wedge \neg\Box P) \supset$ which is just the original sequent weakened by P on the left:

$$\dfrac{\dfrac{\dfrac{P \wedge \neg\Box P, \Box(P \wedge \neg\Box P \to P), \Box(P \wedge \neg\Box P) \supset P \wedge \neg\Box P}{\Box(P \wedge \neg\Box P \to P), \Box(P \wedge \neg\Box P) \supset P \wedge \neg\Box P}\,(\Box\supset) \qquad P, \Box(P \wedge \neg\Box P \to P), \Box(P \wedge \neg\Box P) \supset}{P \wedge \neg\Box P \to P, \Box(P \wedge \neg\Box P \to P), \Box(P \wedge \neg\Box P) \supset}\,(\to\supset)}{\Box(P \wedge \neg\Box P \to P), \Box(P \wedge \neg\Box P) \supset}\,(\Box\supset)$$

So for the remaining of the proof we will have to check if weakening P on the left helps to construct a prehistoric-cycle-free proof.

2. case: The last rule is a $(\Box\supset)$ rule with $\Box(P \wedge \neg\Box P)$ as the principal formula. We get as premise the sequent $P \wedge \neg\Box P, \Box(P \wedge \neg\Box P \to P), \Box(P \wedge \neg\Box P) \supset$ which again by inversion and an easy deduction is provable prehistoric-cycle-free iff $P, \Box(P \wedge \neg\Box P \to P), \Box(P \wedge \neg\Box P) \supset \Box P$ is provable prehistoric-cycle-free.

[1] If we assume that $P \wedge \neg y \cdot x{:}P \to P$ is an axiom A0, this matches the more general result in corollary 7.2 in Artemov (2001, 14): $\mathrm{LP(CS)} \vdash F$ if and only if $\mathrm{LPG}_0 \vdash \mathrm{CS} \supset F$.

It is clear that using $(\Box \supset)$ rules on this sequent just adds additional copies of the existing formulas by the same arguments. So by contraction if there is a prehistoric-cycle-free proof for this sequent, then there is also one ending in a $(\supset \Box)$ rule. The premise of this rule has to have the form $\Box(P \wedge \neg\Box P \to P) \supset P$ to avoid a prehistoric cycle. But the following Kripke model shows that $\Box(P \wedge \neg\Box P \to P) \to P$ is not a theorem of S4 and therefore not provable at all: $W := w, val(P) := \emptyset, R := \{(w,w)\}$. We have $w \Vdash P \wedge \neg\Box P \to P$ because $w \Vdash \neg P$ and therefore also $w \Vdash \neg(P \wedge \neg\Box P)$. As w is the only world we get $w \Vdash \Box(P \wedge \neg\Box P \to P)$ which leads to the final $w \Vdash \neg(\Box(P \wedge \neg\Box P \to P) \to P)$ again because $w \Vdash \neg P$.

As all possibilities for a prehistoric-cycle-free proof of $\Box(P \wedge \neg\Box P \to P), \Box(P \wedge \neg\Box P) \supset$ are exhausted, there is no such proof and therefore also no prehistoric-cycle-free proof of $\supset \Box(P \wedge \neg\Box P \to P) \to \neg\Box(P \wedge \neg\Box P)$ ■

Theorem 12. *There exists a S4−theorem A and a LP-formula B such that A has no prehistoric-cycle-free G3s−proof,* $B^\circ = A$ *and* $\mathrm{LP}(\mathrm{CS}^{\circledast}) \vdash B$

Proof. $A := \Box(P \wedge \neg\Box P \to P) \to \neg\Box(P \wedge \neg\Box P)$ is a theorem of S4, as $\neg\Box(P \wedge \neg\Box P)$ already is a theorem of S4. By the previous lemma, there is no prehistoric-cycle-free proof for A and by the first lemma $B := y{:}(P \wedge \neg y \cdot x{:}P \to P) \to \neg x{:}(P \wedge \neg y \cdot x{:}P)$ is a realization of A provable in LP_0 and therefor also in $\mathrm{LP}(\mathrm{CS}^{\circledast})$. ■

Finally the question arises if prehistoric cycles are also a necessary condition on self-referential S4 theorems under the expanded definition. For this it is necessary to clarify the term inputs for Hilbert style proofs used in the original definition of LP and in the realization theorem (thm. 6) as there is no direct equivalent for $(\supset{:})_t$ rules in the Hilbert style LP calculus as there is for $(\supset{:})_c$ rules. Looking at the adequacy proof for G3lp, $(\supset{:})_t$ is used only for the base cases $A \supset A$ in proofing axioms of LP. In the other direction, a $(\supset{:})_t$ rule is translated first to the trivial proof for $t{:}A \vdash_{\mathrm{LP}} t{:}A$, but the usage of deduction theorem could change that to a different proof for example for $\vdash_{\mathrm{LP}} t{:}A \to t{:}A$.

So far, the situation seems pretty clear cut, and we have inputs as assumptions or as subformulas with negative polarity of formulas proven by the deduction theorem. This also matches the notion that $(\supset{:})_t$ rules introduce the arguments of Skolem functions used in the LP realization. Unfortunately the deductions as constructed in the deduction theorem sometimes use existing formulas with swapped polarities. That is, in a deduction constructed by the deduction theorem, subformulas can occur with negative polarity which only occurred with positive polarity in the original deduction. Moreover formulas can be necessary to derive the final formula without occurring in that formula. So there is no guarantee that all necessary inputs actually occur in the final formula or that a formula occurring with negative polarity somewhere in the proof is an input.

So we have no clear definition of inputs in the original definition of LP matching the definition of inputs in G3lp, and therefore also currently no way to expand Yu's result to all inputs. But we can stipulate that the inputs of a derivation d as constructed by the realization theorem (thm. 6) are exactly the realizations of formulas $\boxminus_i A$ with negative polarity in the original G3s proof. As G3s enjoys the subformula property, that means all inputs used in the proof thus constructed are actually also inputs in the final formula of the proof, a property which does not necessarily hold for all derivations as discussed above. We have to assume without proof that this definition of inputs somehow matches the exact definition given in the context of G3lp proofs. That is, there exists a G3lp proof for a G3s proof where only realizations of formulas with negative polarity are introduced by $(\supset:)_t$. Given this stipulations and assumptions, the following sketch of a proof tries to argue for the necessity of prehistoric cycles for the expanded definition of self-referentiality:

Conjecture. *If a S4−theorem A has a left-prehistoric-cycle-free G3s−proof, then there is a LP−formula B s.t.* $B^\circ = A$ *and* $\mathrm{LP}(\mathrm{IN}^\circledast) \vdash B$.

Proof. Given a left-prehistoric-cycle-free G3s−proof $\mathcal{T} = (T, R)$ for A, use lemma 13 and the realization theorem (thm. 6) to construct a realization function r_T^N and a constant specification CS^N such that $B := r_T^N(\mathrm{an}_T(A))$ is a realization of A and $\mathrm{LP} \vdash B$ by the constructed deduction d. To simplify the following, we do not enforce a injective constant specification here and allow multiple proof constants for the same formula. From this it follows that any constant $c_{i,j,k}$ is exclusively used when handling the $(\supset \Box)$ rule $R_{i,j}$.

Assume for contradiction that the set of inputs IN used for d is self-referential. That is there is a subset $\{t_0{:}A_0(t_1), ..., t_{n-1}{:}A_{n-1}(t_0)\}$ of IN. The occurrences of $t_{k+1 \bmod n}$ in $t_k{:}A_i$ have to be a subterm of realization term for a principal family i_k as the construction of such realization terms are the only place where the constants and variables of IN can get reused. For every consecutive pair of principal families i_k and $i_{k'}$ thus given, there is a constant or variable $t_{k'}$ such that $t_{k'}$ occurs in the realization term for i_k and there is a subterm of the realization term for $i_{k'}$ occurring in $t_{k'}{:}A_{k'} \in \mathrm{IN}$. We distinguish the following cases:

1. case: $t_{k'}$ is a variable x_j. Then the formula $t_{k'}{:}A_{k'}$ is the realization of an annotated S4 formula $\boxminus_j A(\boxplus_{i_{k'}})$. That formula occurs on the left of a $(\supset \Box)$ rule introducing an occurrence of $\boxplus_k$ as x_j is in the realization term of $\boxplus_k$. Therefore we have $i_{k'} \prec i_k$.

2. case: $t_{k'}$ is a constant $c_{j,l,m}$. Then the formula $t_{k'}{:}A_{k'}$ is added to the CS when handling a $(\supset \Box)$ rule $R_{j,l}$ introducing an occurrence of $\boxplus_j$. $c_{j,l,m}$ is in the realization term of $\boxplus_k$ so $R_{j,l}$ lies in a prehistory of $\boxplus_k$. At the same time, the term $t_{k'}$ occurs in the formula $c_{j,l,m}{:}A_{k'}$ as part of a term t used in the construction of the realization of $\boxplus_{k'}$. As $c_{j,l,m}{:}A_{k'}$ is introduced when realizing $R_{j,l}$, $A_{k'}$ occurs in the proof of the premise and there has to be an occurrence of $\boxplus_{k'}$ in the prehistory of $R_{j,l}$. Together we get that $\boxplus_{k'}$ occurs in a prehistory of $\boxplus_k$ and therefore $i_{k'} \prec i_k$ by lemma 8.

So for all $k < n$ we get $i_{k'} \prec i_k$ and the list of principal families $i_0, ..., i_{n-1}$ is therefore a prehistoric cycle in $\mathcal{T}$. ■

Chapter 5

Conclusion

In this paper, I reproduced the main result of Yu's paper "Prehistoric Phenomena and Self-referentiality" and then expanded on it defining prehistoric relations for Gentzen systems with cut rules and finally for a Gentzen system G3lp for the logic of proofs LP. This allows to study prehistoric relations directly in LP and leads to a negative answer on Yu's conjecture that prehistoric cycles are sufficient for self-referential S4 theorems. It also leads to an expanded definition of self-referentiality considering all inputs used to construct justification terms. With that expanded definition of self-referentiality prehistoric cycles are *sufficient* for self-referential theorems in S4, which is the main result of this paper.

Given this expansion, the question goes back to the other direction. That is, are prehistoric cycles also necessary for the expanded definition of self-referentiality? Unfortunately this question is not easy to answer, as already transferring the definitions of inputs to the original Hilbert style calculus poses problems. A more detailed discussion of Skolem style functions and their role in LP realizations will hopefully help to clear this up. It is possible that the definition of input variables relative to a subformula occurrence and the machinery used to work with input variables in Studer and Kuznets (n.d. ch 3.2) already provides a part of the answer.

Yu later expanded his result to modal logics T and K4 and their justification counterparts in Yu (2014). Another open question is whether the same generalization can be done with the results of this paper. That is, if there are Gentzen style systems without structural rules for T and K4 together with a consistent definition of term correspondence and prehistoric relations and a translation to some variant of G3s.

Acknowledgments

My sincere thanks go to Prof. Dr. Thomas Studer for his support and inspiration. Without him I would not be at this point.

Bibliography

Artemov, Sergei. 2008. "The Logic of Justification." *The Review of Symbolic Logic* 1 (04): 477. https://doi.org/10.1017/S1755020308090060.

Artemov, Sergei, and Melvin Fitting. 2016. "Justification Logic." In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta, Winter 2016. Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/win2016/entries/logic-justification/.

Artemov, Sergei N. 1995. *Operational Modal Logic*. Vol. 95. 29. Mathematical Sciences Institute, Cornell University.

———. 2001. "Explicit Provability and Constructive Semantics," March, 36.

Brezhnev, Vladimir. 2001. "On the Logic of Proofs." In *Proceedings of the Sixth ESSLLI Student Session, Helsinki*, 35–46.

Brezhnev, Vladimir, and Roman Kuznets. 2006. "Making Knowledge Explicit: How Hard It Is." *Theoretical Computer Science* 357 (1-3): 23–34. https://doi.org/10.1016/j.tcs.2006.03.010.

Fitting, Melvin. 2005. "The Logic of Proofs, Semantically." *Annals of Pure and Applied Logic* 132 (1): 1–26.

———. 2009. "Realizations and LP." *Annals of Pure and Applied Logic* 161 (3): 368–87. https://doi.org/10.1016/j.apal.2009.07.010.

Kuznets, Roman. 2010. "Self-Referential Justifications in Epistemic Logic." *Theory of Computing Systems* 46 (4): 636–61. https://doi.org/10.1007/s00224-009-9209-3.

Pulver, Cornelia. 2010. "Self-Referentiality in Contraction-Free Fragments of Modal Logic S4." Universität Bern.

Studer, Thomas, and Roman Kuznets. n.d. "Logics of Proofs and Justifications." College Publications, in preparation.

Troelstra, A. S., and Helmut Schwichtenberg. 2000. *Basic Proof Theory*. 2nd ed. Cambridge Tracts in Theoretical Computer Science 43. Cambridge ; New York: Cambridge University Press.

Yu, Junhua. 2010. "Prehistoric Phenomena and Self-Referentiality," 13.

———. 2014. "Prehistoric Graph in Modal Derivations and Self-Referentiality." *Theory of Computing Systems* 54 (2): 190–210. https://doi.org/10.1007/s00224-

013-9510-z.

———. 2017. "On Non-Self-Referential Fragments of Modal Logics." *Annals of Pure and Applied Logic* 168 (4): 776–803.