# LECTURES ON JUSTIFICATION LOGIC

THOMAS STUDER
*Institut für Informatik und angewandte Mathematik*
*Universität Bern*
*Neubrückstrasse 10*
*3012 Bern, Switzerland*
`tstuder@iam.unibe.ch`

November 28, 2012

# CONTENTS


Preface ........ 5

Chapter 1. Modal Logic Preliminaries ........ 7
- 1A. Syntax ........ 7
- 1B. Semantics ........ 8

Chapter 2. Defining Justification Logic ........ 15
- 2A. Syntax ........ 15
- 2B. Basic Properties ........ 16
- 2C. Basic Modular Models ........ 19
- 2D. Epistemic Models ........ 21
- 2E. Justifications and Belief ........ 25

Chapter 3. Relations with S4 ........ 27
- 3A. Forgetful Projection ........ 27
- 3B. Realization ........ 28

Chapter 4. Self-Referentiality ........ 39
- 4A. Generated Models ........ 39
- 4B. Self-Referential Knowledge ........ 43
- 4C. Notes ........ 46

Chapter 5. Decidability ........ 47
- 5A. Post's theorem ........ 47
- 5B. Finitary models ........ 48
- 5C. Establishing decidability ........ 52
- 5D. A decidable constant specification is not enough ........ 54

Chapter 6. Logical Omniscience ........ 55
- 6A. Logical Omniscience Tests ........ 55
- 6B. The case for S4 ........ 57
- 6C. Reflected fragments for $\mathsf{LP}_{\mathsf{CS}}$ ........ 58
- 6D. The general case ........ 60
- 6E. Notes ........ 61

References ........................................................ 63

# PREFACE

These are the lectures notes for a course on justification logic given at the University of Bern in the autumn semester 2012. They provide an introduction the Logic of Proofs with a focus on the mathematical methods that are important in this area.

These notes are far from being complete. For instance, the following topics are not covered: constructive realization, arithmetical semantics, and complexity results. Moreover, these notes only treat the Logic of Proofs but no other justification logics. They also contain only the mathematical material but almost no additional explanations or references.

**Acknowledgements.** I am grateful to Kentaro Sato for carefully proof reading this manuscript. I also would like to thank Roman Kuznets for many very helpful explanations and comments on justification logics.

CHAPTER 1

# MODAL LOGIC PRELIMINARIES

## 1A. Syntax

We start with a countable set of atomic propositions $\mathsf{Prop}$. *Formulas of the language $\mathcal{L}$ of modal logic* are given inductively as follows:

1. each atomic proposition $P \in \mathsf{Prop}$ is a formula of $\mathcal{L}$;
2. if $A$ is a formula of $\mathcal{L}$, then so is $\neg A$;
3. if $A$ and $B$ are formulas of $\mathcal{L}$, then so is $(A \to B)$;
4. if $A$ is a formula of $\mathcal{L}$, then so is $\Box A$.

If there is no danger of confusion, we will omit parentheses in formulas. As usual, we define $A \vee B := \neg A \to B$, $A \wedge B := \neg(\neg A \vee \neg B)$, and $\Diamond A := \neg\Box\neg A$. Further we let $\bot := P \wedge \neg P$ and $\top := P \vee \neg P$ for a fixed atomic proposition $P$.

For a set $\Gamma$ of formulas we define

$$\Box\Gamma := \{\Box A \mid A \in \Gamma\} \ .$$

If $\Gamma := \{A_1, \dots, A_n\}$ is a finite non-empty set of formulas, we set

$$\bigwedge\Gamma := A_1 \wedge \cdots \wedge A_n \ .$$

We can use this sloppy notation since in the logics we are going to consider conjunction is provably associative and commutative. For $\Gamma = \emptyset$, we define $\bigwedge\Gamma := \top$.

The *modal logic* $\mathsf{S4}$ is given by all instances of the following axiom schemes:

(P): finitely many schemes axiomatizing classical propositional logic
(K): $\Box(A \to B) \to (\Box A \to \Box B)$
(T): $\Box A \to A$
(4): $\Box A \to \Box\Box A$

and by the rules modus ponens and necessitation

$$\frac{A \qquad A \to B}{B}\,(\mathsf{MP}) \qquad\qquad \frac{A}{\Box A}\,(\mathsf{NEC}) \ .$$

We write $\mathsf{S4} \vdash A$ if a formula $A$ is derivable in $\mathsf{S4}$.

LEMMA 1. *Let $\Sigma$ be a finite set of formulas. We have that*

$$\mathsf{S4} \vdash (\bigwedge \Sigma) \to A \quad \textit{implies} \quad \mathsf{S4} \vdash (\bigwedge \Box\Sigma) \to \Box A \ . \tag{1}$$

PROOF. We first show

$$\mathsf{S4} \vdash A \to \Box A \tag{2}$$

for any propositional tautology $A$. We have $\mathsf{S4} \vdash A$ and thus by ($\mathsf{NEC}$) also $\mathsf{S4} \vdash \Box A$. Further we know $\mathsf{S4} \vdash \Box A \to (A \to \Box A)$ since this is a propositional tautology, too. Hence applying ($\mathsf{MP}$) yields (2).

Now we show (1) by induction on the cardinality $n$ of $\Sigma$. If $n = 0$, the premise is $\mathsf{S4} \vdash \top \to A$. By ($\mathsf{NEC}$), we infer $\mathsf{S4} \vdash \Box(\top \to A)$. By ($\mathsf{K}$) we have $\mathsf{S4} \vdash \Box(\top \to A) \to (\Box\top \to \Box A)$. Thus by ($\mathsf{MP}$), we infer $\mathsf{S4} \vdash \Box\top \to \Box A$. From (2) we have $\mathsf{S4} \vdash \top \to \Box\top$. Therefore, by propositional reasoning we conclude $\mathsf{S4} \vdash \top \to \Box A$.

For $n > 0$, let $B \in \Sigma$ and $\Sigma' := \Sigma \setminus \{B\}$. From the premise we get by propositional reasoning $\mathsf{S4} \vdash \bigwedge \Sigma' \to (B \to A)$. By I.H. we obtain $\mathsf{S4} \vdash \bigwedge \Box\Sigma' \to \Box(B \to A)$. By ($\mathsf{K}$) and propositional reasoning we find $\mathsf{S4} \vdash \bigwedge \Box\Sigma' \to (\Box B \to \Box A)$. Again by propositional reasoning this is $\mathsf{S4} \vdash \bigwedge \Box\Sigma' \wedge \Box B \to \Box A$, which is $\mathsf{S4} \vdash \bigwedge \Box\Sigma \to \Box A$, ⊣

## 1B. Semantics

DEFINITION 2. A binary relation $R \subseteq W \times W$ is called

1. *reflexive* if $R(u, u)$ for each $u \in W$
2. *transitive* if $R(u, v)$ and $R(v, w)$ imply $R(u, w)$ for any $u, v, w \in W$.

DEFINITION 3 (Kripke structure). A *Kripke structure* $\mathcal{M} = (W, R, \mathsf{val})$ is a triple where

1. $W$ is a non-empty set,
2. $R$ is a binary relation on $W$ that is transitive and reflexive,
3. $\mathsf{val}$ is a valuation function $\mathsf{val} : \mathsf{Prop} \to \mathcal{P}(W)$.

We define what it means for a formula to hold at a world in a Kripke structure inductively on the structure of the formula.

DEFINITION 4 (Truth). Let $\mathcal{M} = (W, R, \mathsf{val})$ be a Kripke structure. For $w \in W$, we define $\mathcal{M}, w \Vdash A$ by

1. $\mathcal{M}, w \Vdash P$ iff $w \in \mathsf{val}(P)$ for $P \in \mathsf{Prop}$
2. $\mathcal{M}, w \Vdash \neg A$ iff $\mathcal{M}, w \nVdash A$
3. $\mathcal{M}, w \Vdash A \to B$ iff $\mathcal{M}, w \nVdash A$ or $\mathcal{M}, w \Vdash B$
4. $\mathcal{M}, w \Vdash \Box A$ iff for all $v$ with $R(w, v)$ we have $\mathcal{M}, v \Vdash A$.

We write $\mathcal{M} \Vdash A$ if for all $w \in W$ we have $\mathcal{M}, w \Vdash A$. A formula $A$ is called *valid* if for all Kripke structures $\mathcal{M}$ we have $\mathcal{M} \Vdash A$.

THEOREM 5 (Soundness). *Let $A$ be a formula of $\mathcal{L}$. We have that*

$$\mathsf{S4} \vdash A \quad \textit{implies} \quad A \textit{ is valid} .$$

PROOF. By induction on the depth of the derivation of $A$ we show that $A$ is valid, i.e. for all Kripke structures $\mathcal{M} = (W, R, \mathsf{val})$ and all $w \in W$ we have $\mathcal{M}, w \Vdash A$. We only show some cases.

1. $A = \Box(B \to C) \to (\Box B \to \Box C)$. Assume (a) $\mathcal{M}, w \Vdash \Box(B \to C)$ and (b) $\mathcal{M}, w \Vdash \Box B$. We have to show $\mathcal{M}, w \Vdash \Box C$, that is (c) $\mathcal{M}, v \Vdash C$ for all $v$ with $R(w, v)$. From (b) we get $\mathcal{M}, v \Vdash B$ and from (a) we get $\mathcal{M}, v \Vdash B \to C$ for any $v$ with $R(w, v)$. Thus we conclude (c).
2. $A = \Box B \to B$. Assume $\mathcal{M}, w \Vdash \Box B$, that is $\mathcal{M}, v \Vdash B$ for all $v$ with $R(w, v)$. Since $R$ is reflexive, we have $R(w, w)$ and we conclude $\mathcal{M}, w \Vdash B$.
3. $A = \Box B \to \Box\Box B$. Assume (a) $\mathcal{M}, w \Vdash \Box B$. For all $u, v \in W$ with (b) $R(w, v)$ and (c) $R(v, u)$ we have $R(w, u)$ since $R$ is transitive. By (a) we find $\mathcal{M}, u \Vdash B$. Thus we have $\mathcal{M}, v \Vdash \Box B$ by (c) and we conclude $\mathcal{M}, w \Vdash \Box\Box B$ by (b).
4. $A = \Box B$ is the conclusion of a necessitation rule. By the induction hypothesis we find that $B$ is valid. In particular, that means $\mathcal{M}, v \Vdash B$ for all $v$ with $R(w, v)$. Thus we conclude $\mathcal{M}, w \Vdash \Box B$.

$\dashv$

To show completeness of S4 we make use of a so-called canonical model, which is based on maximal consistent sets. This construction is very general and will also be employed later to show completeness for justification logics.

The following definitions are not given for S4 only, but more general, for any logic L with with classical Boolean logic in the background.

DEFINITION 6 (Consistency). Let L be a logic. A set $\Gamma$ of formulas is L-*consistent* if

$$\mathsf{L} \nvdash \bigwedge \Sigma \to \bot$$

for each finite subset $\Sigma \subseteq \Gamma$.

A set $\Gamma$ is called *maximal* L-*consistent* if it is consistent whereas no proper superset is.

LEMMA 7 (Lindenbaum). *For each* L-*consistent set* $\Delta$ *there exists a maximal* L-*consistent set* $\Gamma$ *with* $\Delta \subseteq \Gamma$.

PROOF. Let $\Delta$ be an L-consistent set. Assume we are given a fixed enumeration of all formulas and let $A_i$ be the $i$-th element of this enumeration.

We set

$$\Delta_0 := \Delta$$

$$\Delta_{n+1} := \begin{cases} \Delta_n \cup \{A_n\} & \text{if } \Delta_n \cup \{A_n\} \text{ is } \mathsf{L}\text{-consistent} \\ \Delta_n & \text{otherwise} \end{cases}$$

$$\Gamma := \bigcup_{n \in \omega} \Delta_n$$

$\Gamma$ is $\mathsf{L}$-consistent for otherwise there exists a natural number $n$ such that already $\Delta_n$ is not $\mathsf{L}$-consistent, which contradicts the definition of $\Delta_n$.

$\Gamma$ is maximal $\mathsf{L}$-consistent. Otherwise there exists an $A_j \notin \Gamma$ such that $\Delta_n \cup \{A_j\}$ is $\mathsf{L}$-consistent for all $n$. Then, in particular, $\Delta_j \cup \{A_j\}$ is $\mathsf{L}$-consistent. Therefore $\Delta_j \cup \{A_j\} \subseteq \Gamma$ , which contradicts $A_j \notin \Gamma$. ⊣

Lemma 8 (Maximal consistent sets). *Let $\Gamma$ be a maximal $\mathsf{L}$-consistent set of formulas. We have:*

1. $A \in \Gamma$ *for all formulas* $A$ *with* $\mathsf{L} \vdash A$
2. $A \in \Gamma$ *if and only if* $\neg A \notin \Gamma$
3. $A \to B \in \Gamma$ *if and only if* $A \notin \Gamma$ *or* $B \in \Gamma$
4. $A \in \Gamma$ *and* $A \to B \in \Gamma$ *imply* $B \in \Gamma$

Proof. 1. Let $A$ be such that $\mathsf{L} \vdash A$ and suppose $A \notin \Gamma$. By maximal consistency of $\Gamma$ this implies that $\Gamma \cup \{A\}$ is not $\mathsf{L}$-consistent. Then there exists a finite $\Sigma \subseteq \Gamma$ such that

$$\mathsf{L} \vdash \bigwedge \Sigma \wedge A \to \bot \ .$$

Since $\mathsf{L} \vdash A$, this implies

$$\mathsf{L} \vdash \bigwedge \Sigma \to \bot \ ,$$

which means that $\Gamma$ is not $\mathsf{L}$-consistent. Contradiction.

2. First we show that $A \in \Gamma$ or $\neg A \in \Gamma$. We suppose to the contrary that $A \notin \Gamma$ and $\neg A \notin \Gamma$. Then both $\Gamma \cup \{A\}$ and $\Gamma \cup \{\neg A\}$ are not $\mathsf{L}$-consistent by maximal consistency of $\Gamma$. Thus there exist finite $\Sigma_1, \Sigma_2 \subseteq \Gamma$ such that

$$\mathsf{L} \vdash \bigwedge \Sigma_1 \wedge A \to \bot$$

and

$$\mathsf{L} \vdash \bigwedge \Sigma_2 \wedge \neg A \to \bot \ .$$

Thus we find for $\Sigma = \Sigma_1 \cup \Sigma_2$

$$\mathsf{L} \vdash \bigwedge \Sigma \wedge (A \vee \neg A) \to \bot \ .$$

Hence $\Gamma \cup \{A \vee \neg A\}$ is not $\mathsf{L}$-consistent. However, since $\mathsf{L} \vdash A \vee \neg A$, we know by the first claim of this lemma that $A \vee \neg A \in \Gamma$. Thus $\Gamma$ is not $\mathsf{L}$-consistent. Contradiction.

It remains to show that not both $A \in \Gamma$ and $\neg A \in \Gamma$. Suppose that $\{A, \neg A\} \subseteq \Gamma$. By $\mathsf{L} \vdash A \wedge \neg A \to \bot$ we immediately conclude that $\Gamma$ is not $\mathsf{L}$-consistent. Contradiction.

3. Assume

$$A \notin \Gamma \quad \text{or} \quad B \in \Gamma \ . \tag{3}$$

Suppose towards a contradiction that $A \to B \notin \Gamma$. Since $\Gamma$ is maximal consistent, there exists a finite $\Sigma \subseteq \Gamma$ such that

$$\mathsf{L} \vdash \bigwedge \Sigma \wedge (A \to B) \to \bot \ .$$

This implies

$$\mathsf{L} \vdash \bigwedge \Sigma \wedge \neg A \to \bot$$

and

$$\mathsf{L} \vdash \bigwedge \Sigma \wedge B \to \bot \ .$$

Thus $\Gamma \cup \{\neg A\}$ and $\Gamma \cup \{B\}$ are not $\mathsf{L}$-consistent. Thus $\neg A \notin \Gamma$ and $B \notin \Gamma$, which contradicts (3). We conclude $A \to B \in \Gamma$.

Assume $A \to B \in \Gamma$. Suppose $A \in \Gamma$ and $B \notin \Gamma$. Thus we have $\neg A \notin \Gamma$ by the previous claim of this lemma. There exist finite $\Sigma_1, \Sigma_2 \subseteq \Gamma$ such that

$$\mathsf{L} \vdash \bigwedge \Sigma_1 \wedge \neg A \to \bot$$

and

$$\mathsf{L} \vdash \bigwedge \Sigma_2 \wedge B \to \bot \ ,$$

which implies for $\Sigma = \Sigma_1 \cup \Sigma_2$

$$\mathsf{L} \vdash \bigwedge \Sigma \wedge (A \to B) \to \bot \ .$$

That means

$$\Gamma \cup \{A \to B\} \text{ is not } \mathsf{L}\text{-consistent,}$$

which contradicts $A \to B \in \Gamma$ and $\Gamma$ $\mathsf{L}$-consistent. We conclude $A \notin \Gamma$ or $B \in \Gamma$.

4. This is a reformulation of the direction from left to right from the previous claim.

$\dashv$

Definition 9 (Canonical model). For a set $\Gamma$ of formulas, we define

$$\Gamma/\Box := \{A \mid \Box A \in \Gamma\} \ .$$

The *canonical model* $\mathcal{M} = (W, R, \mathsf{val})$ is given by:

1. $W := \{\Gamma \mid \Gamma \text{ is a maximal } \mathsf{S4}\text{-consistent set of } \mathcal{L} \text{ formulas}\}$
2. $R := \{(\Gamma, \Delta) \mid \Gamma/\Box \subseteq \Delta\}$
3. $\Gamma \in \mathsf{val}(P)$ iff $P \in \Gamma$

Lemma 10. *The canonical model $\mathcal{M} = (W, R, \mathsf{val})$ is a Kripke structure.*

Proof. First we show that $W$ is non-empty. We show that the empty set $\emptyset$ is S4-consistent. Then by the Lindenbaum lemma, there exists a maximal consistent superset of $\emptyset$, which is an element of $W$. Suppose $\emptyset$ is not S4-consistent. That means

$$\mathsf{S4} \vdash \bigwedge \emptyset \to \bot \ ,$$

which is

$$\mathsf{S4} \vdash \top \to \bot \ .$$

Take now an S4-proof of $\top \to \bot$ and erase all $\Box$-connectives in each of its formulas. What remains is a chain of propositional formulas that are provable in classical propositional logic (an easy induction on the length of the original proof) and that ends with $\top \to \bot$, a contradiction.

Now we establish that $R$ is transitive and reflexive.

**Transitivity.** Assume $(\Gamma, \Delta) \in R$ and $(\Delta, \Phi) \in R$. Assume $A \in \Gamma/\Box$, that is $\Box A \in \Gamma$. Since $\Gamma$ is a maximal consistent set, it includes the axiom $\Box A \to \Box\Box A$, and we infer $\Box\Box A \in \Gamma$. We obtain $\Box A \in \Gamma/\Box \subseteq \Delta$. Therefore $A \in \Delta/\Box \subseteq \Phi$. We conclude $\Gamma/\Box \subseteq \Phi$, which gives us $(\Gamma, \Phi) \in R$.

**Reflexivity.** Assume $A \in \Gamma/\Box$, that is $\Box A \in \Gamma$. Since $\Gamma$ is maximal consistent, it includes the axiom $\Box A \to A$ and we infer $A \in \Gamma$. We have $\Gamma/\Box \subseteq \Gamma$, which gives us $(\Gamma, \Gamma) \in R$. ⊣

Lemma 11 (Truth lemma). *Let $\mathcal{M} = (W, R, \mathsf{val})$ be the canoncial model. For each world $\Gamma \in W$ and each formula $A$ we have*

$$\mathcal{M}, \Gamma \Vdash A \quad \textit{if and only if} \quad A \in \Gamma \ .$$

Proof. By induction on the structure of $A$.

1. $A$ is an atomic proposition $P$. We have $\mathcal{M}, \Gamma \Vdash P$ iff $\Gamma \in \mathsf{val}(P)$ iff $P \in \Gamma$.
2. $A = \neg B$. We have $\mathcal{M}, \Gamma \Vdash \neg B$ iff $\mathcal{M}, \Gamma \nVdash B$ iff (by I.H.) $B \notin \Gamma$ iff (by maximal consistency of $\Gamma$) $\neg B \in \Gamma$.
3. $A = B \to C$. We have $\mathcal{M}, \Gamma \Vdash B \to C$ iff $\mathcal{M}, \Gamma \nVdash B$ or $\mathcal{M}, \Gamma \Vdash C$ iff (by I.H.) $B \notin \Gamma$ or $C \in \Gamma$ iff (by maximal consistency of $\Gamma$) $B \to C \in \Gamma$.

4. $A = \Box B$. Assume $\Box B \in \Gamma$. Hence $B \in \Gamma/\Box$. If $(\Gamma, \Delta) \in R$, then $B \in \Delta$ by the definition of $R$ in the canonical model. Thus using I.H. we get $\mathcal{M}, \Delta \Vdash B$ for all $\Delta$ with $(\Gamma, \Delta) \in R$. That means $\mathcal{M}, \Gamma \Vdash \Box B$.

   Assume $\mathcal{M}, \Gamma \Vdash \Box B$. Then

$$\Gamma/\Box \cup \{\neg B\} \text{ is not } \mathsf{S4}\text{-consistent.} \tag{4}$$

   Suppose (4) does not hold. Then there exists a maximal consistent set $\Delta \supseteq \Gamma/\Box \cup \{\neg B\}$ and by the definition of $R$ we find $(\Gamma, \Delta) \in R$. Using I.H. we find $\mathcal{M}, \Delta \nVdash B$, which contradicts $\mathcal{M}, \Gamma \Vdash \Box B$. Hence (4) is established.

   Now (4) implies that there exists a finite subset $\Sigma \subseteq \Gamma/\Box$ such that

$$\mathsf{S4} \vdash \bigwedge \Sigma \wedge \neg B \to \bot \ .$$

   By propositional reasoning this implies

$$\mathsf{S4} \vdash \bigwedge \Sigma \to B \ .$$

   By Lemma 1 we get

$$\mathsf{S4} \vdash \bigwedge \Box\Sigma \to \Box B \ .$$

   Again by propositional reasoning we find

$$\mathsf{S4} \vdash \bigwedge \Box\Sigma \wedge \neg\Box B \to \bot \ .$$

   By $\Box\Sigma \subseteq \Gamma$, this implies that $\Gamma \cup \{\neg\Box B\}$ is not S4-consistent. Therefore, $\neg\Box B \notin \Gamma$ and thus, by maximal consistency, $\Box B \in \Gamma$.

$\dashv$

THEOREM 12 (Completeness). *Let $A$ be a formula of $\mathcal{L}$. We have that*

$$A \textit{ is valid} \quad \textit{implies} \quad \mathsf{S4} \vdash A \ .$$

PROOF. We show the contrapositive. Assume $\mathsf{S4} \nvdash A$. Then $\mathsf{S4} \nvdash \neg A \to \bot$. That is $\{\neg A\}$ is S4-consistent and hence contained in a maximal consistent set $\Gamma$. Let $\mathcal{M}$ be the canonical model. By the Truth lemma we find $\mathcal{M}, \Gamma \Vdash \neg A$, which is $\mathcal{M}, \Gamma \nVdash A$. Therefore $A$ is not valid. $\dashv$

CHAPTER 2

# DEFINING JUSTIFICATION LOGIC

## 2A. Syntax

*Justification terms* are built from countably many *constants* and countably many *variables* as follows:

1. each constant $c$ is a term;
2. each variable $x$ is a term;
3. if $s$ and $t$ are terms, then so is $(s \cdot t)$;
4. if $s$ and $t$ are terms, then so is $(s + t)$;
5. if $s$ is a term, then so is $!s$.

The set of *atomic terms* consists of all constants and variables. We denote the *set of all terms* by $\mathsf{Tm}$. A term is called *ground* if it does not contain variables. We use $t(x_1, \ldots, x_n)$ to express that the term $t$ contains at most the variables $x_1, \ldots, x_n$. The operations of $\cdot$ and $+$ are assumed to be left-associative in order to omit unnecessary parentheses. *Formulas* of the language $\mathcal{L}_{\mathsf{J}}$ are given inductively as follows:

1. each atomic proposition $P \in \mathsf{Prop}$ is a formula of $\mathcal{L}_{\mathsf{J}}$;
2. if $A$ is a formula of $\mathcal{L}_{\mathsf{J}}$, then so is $\neg A$;
3. if $A$ and $B$ are formulas of $\mathcal{L}_{\mathsf{J}}$, then so is $(A \to B)$;
4. if $t$ is a term and $A$ is a formula of $\mathcal{L}_{\mathsf{J}}$, then $t{:}A$ is a formula of $\mathcal{L}_{\mathsf{J}}$.

The axioms of the *Logic of Proofs* $\mathsf{LP}$ consist of all instances of the following schemes:

($\mathsf{P}$)**:** finitely many schemes axiomatizing classical propositional logic
($\mathsf{J}$)**:** $t{:}(A \to B) \to (s{:}A \to (t \cdot s){:}B)$,
($+$)**:** $t{:}A \to (t + s){:}A$ and $s{:}A \to (t + s){:}A$,
($\mathsf{jt}$)**:** $t{:}A \to A$,
($\mathsf{j4}$)**:** $t{:}A \to\ !t{:}t{:}A$.

A *constant specification* $\mathsf{CS}$ for $\mathsf{LP}$ is any subset

$$\mathsf{CS} \subseteq \{(c, F) \mid c \text{ is a constant and } A \text{ is an axiom of } \mathsf{LP}\}.$$

Constant specifications determine axiom instances for which the logic provides justifications. A constant specification $\mathsf{CS}$ for $\mathsf{LP}$ is called *axiomatically appropriate* if for each axiom $F \in \mathsf{LP}$, there is a constant $c$ such that $(c, F) \in \mathsf{CS}$.

The *deductive system* $\mathsf{LP}_{\mathsf{CS}}$ is the Hilbert system given by the axioms $\mathsf{LP}$ and by the rules modus ponens

$$\frac{A \quad A \to B}{B}\ (\mathrm{MP})$$

and axiom necessitation

$$\frac{}{c{:}F}\ (\mathrm{AN}), \quad \text{where } (c, F) \in \mathsf{CS}\ .$$

We write $\mathsf{LP}_{\mathsf{CS}} \vdash A$ to mean that the *formula* $A$ *is derivable in* $\mathsf{LP}_{\mathsf{CS}}$ and $\Delta \vdash_{\mathsf{LP}_{\mathsf{CS}}} A$ to mean that the *formula* $A$ *is derivable in* $\mathsf{LP}_{\mathsf{CS}}$ *from the set of formulas* $\Delta$. When the logic $\mathsf{LP}_{\mathsf{CS}}$ is clear from the context, the subscript $\mathsf{LP}_{\mathsf{CS}}$ is omitted. We write $\Delta, A$ for $\Delta \cup \{A\}$.

EXAMPLE 13. Assume we are given $\mathsf{LP}_{\mathsf{CS}}$ with

$$(a, A \to (A \vee B)) \in \mathsf{CS} \quad \text{and} \quad (b, B \to (A \vee B)) \in \mathsf{CS}\ .$$

Then the following is a theorem of $\mathsf{LP}_{\mathsf{CS}}$

$$(x{:}A \vee y{:}B) \to (a \cdot x + b \cdot y){:}(A \vee B)\ .$$

PROOF. From axiom necessitation we get

$$\mathsf{LP}_{\mathsf{CS}} \vdash a{:}(A \to (A \vee B)) \quad \text{and} \quad \mathsf{LP}_{\mathsf{CS}} \vdash b{:}(B \to (A \vee B))\ .$$

Using (J) and (MP) we obtain

$$\mathsf{LP}_{\mathsf{CS}} \vdash x{:}A \to (a \cdot x){:}(A \vee B) \quad \text{and} \quad \mathsf{LP}_{\mathsf{CS}} \vdash y{:}B \to (b \cdot y){:}(A \vee B)\ .$$

Finally, from (+) we have

$$\mathsf{LP}_{\mathsf{CS}} \vdash (a \cdot x){:}(A \vee B) \to (a \cdot x + b \cdot y){:}(A \vee B)$$

and

$$\mathsf{LP}_{\mathsf{CS}} \vdash (b \cdot y){:}(A \vee B) \to (a \cdot x + b \cdot y){:}(A \vee B)\ .$$

Using propositional reasoning, we obtain the desired result. $\dashv$

## 2B. Basic Properties

The *deduction theorem* is standard for justification logics.

THEOREM 14 (Deduction Theorem). *For any constant specification* $\mathsf{CS}$*, any* $\Delta \subseteq \mathcal{L}_{\mathsf{J}}$*, and arbitrary* $A, B \in \mathcal{L}_{\mathsf{J}}$*,*

$$\Delta, A \vdash_{\mathsf{LP}_{\mathsf{CS}}} B \quad \Longleftrightarrow \quad \Delta \vdash_{\mathsf{LP}_{\mathsf{CS}}} A \to B\ .$$

Proof. The direction from right to left is trivial. Assume that we have $\Delta \vdash_{\mathsf{LP}_{\mathsf{CS}}} A \rightarrow B$. Then also $\Delta, A \vdash_{\mathsf{LP}_{\mathsf{CS}}} A \rightarrow B$. Clearly, we also have $\Delta, A \vdash_{\mathsf{LP}_{\mathsf{CS}}} A$. Thus we conclude by modus ponens $\Delta, A \vdash_{\mathsf{LP}_{\mathsf{CS}}} B$.

The direction from left to right is shown by induction on the depth of the derivation $\Delta, A \vdash_{\mathsf{LP}_{\mathsf{CS}}} B$. We distinguish the following cases.

1. $B$ is an axiom of LP or $B \in \Delta$. We have $B \rightarrow (A \rightarrow B)$ is a propositional tautology, hence $\Delta \vdash_{\mathsf{LP}_{\mathsf{CS}}} B \rightarrow (A \rightarrow B)$. Moreover, $\Delta \vdash_{\mathsf{LP}_{\mathsf{CS}}} B$. By modus ponens we conclude $\Delta \vdash_{\mathsf{LP}_{\mathsf{CS}}} A \rightarrow B$.
2. $B = A$. Then $A \rightarrow B$ is a propositional tautology. Therefore, we conclude $\Delta \vdash_{\mathsf{LP}_{\mathsf{CS}}} A \rightarrow B$.
3. $B$ is the conclusion of an application of (MP) with the premises
$$\Delta, A \vdash_{\mathsf{LP}_{\mathsf{CS}}} C \quad \text{and} \quad \Delta, A \vdash_{\mathsf{LP}_{\mathsf{CS}}} C \rightarrow B \ .$$
By the induction hypothesis we find
$$\Delta \vdash_{\mathsf{LP}_{\mathsf{CS}}} A \rightarrow C \quad \text{and} \quad \Delta \vdash_{\mathsf{LP}_{\mathsf{CS}}} A \rightarrow (C \rightarrow B) \ .$$
Since $(A \rightarrow (C \rightarrow B)) \rightarrow ((A \rightarrow C) \rightarrow (A \rightarrow B))$ is a tautology, we find by modus ponens
$$\Delta \vdash_{\mathsf{LP}_{\mathsf{CS}}} (A \rightarrow C) \rightarrow (A \rightarrow B) \ .$$
Another application of modus ponens yields $\Delta \vdash_{\mathsf{LP}_{\mathsf{CS}}} A \rightarrow B$.
4. $B$ is the conclusion of an application of axiom necessiation. This is similar to the first case.

$\dashv$

An important property of justification logics is their ability to internalize their own notion of proof, as stated in the following lemma.

Lemma 15 (Internalization for Variables). *Let* CS *be an axiomatically appropriate constant specification. For arbitrary formulas $A, B_1, \dots, B_n$ of $\mathcal{L}_{\mathsf{J}}$, if*
$$B_1, \dots, B_n \vdash_{\mathsf{LP}_{\mathsf{CS}}} A \ ,$$
*then there is a term $t(x_1, \dots, x_n) \in \mathsf{Tm}$ such that*
$$x_1{:}B_1, \dots, x_n{:}B_n \vdash_{\mathsf{LP}_{\mathsf{CS}}} t(x_1, \dots, x_n){:}A$$
*for fresh variables $x_1, \dots, x_n$.*

Proof. By induction on the depth of the derivation $B_1, \dots, B_n \vdash_{\mathsf{LP}_{\mathsf{CS}}} A$. We distinguish the following cases.

1. $A$ is an axiom of LP. Since CS is axiomatically appropriate, there exists a constant $c$ with $(c, A) \in \mathsf{CS}$. We find by axiom necessitation that $x_1{:}B_1, \dots, x_n{:}B_n \vdash_{\mathsf{LP}_{\mathsf{CS}}} c{:}A$.
2. $A = B_i$. We set $t := x_i$ and $x_1{:}B_1, \dots, x_n{:}B_n \vdash_{\mathsf{LP}_{\mathsf{CS}}} t{:}A$ immediately follows.

3. $A$ is the conclusion of an application of (MP) with the premises

$$B_1, \dots, B_n \vdash_{\mathsf{LP_{CS}}} C \quad \text{and} \quad B_1, \dots, B_n \vdash_{\mathsf{LP_{CS}}} C \to A \ .$$

By the induction hypothesis there are terms $r, s$ with

$$x_1{:}B_1, \dots, x_n{:}B_n \vdash_{\mathsf{LP_{CS}}} r{:}C \quad \text{and} \quad x_1{:}B_1, \dots, x_n{:}B_n \vdash_{\mathsf{LP_{CS}}} s{:}(C \to A) \ .$$

Thus we conclude using (J) and (MP)

$$x_1{:}B_1, \dots, x_n{:}B_n \vdash_{\mathsf{LP_{CS}}} s \cdot r{:}A \ .$$

4. $A$ is the conclusion of an instance of axiom necessitation. Then $A$ has the form $c{:}F$ with $(c, F) \in \mathsf{CS}$. Thus we also have

$$x_1{:}B_1, \dots, x_n{:}B_n \vdash_{\mathsf{LP_{CS}}} c{:}F$$

and we conclude by (j4) and (MP)

$$x_1{:}B_1, \dots, x_n{:}B_n \vdash_{\mathsf{LP_{CS}}} !c{:}c{:}F \ .$$

⊣

If $n = 0$, the resulting statement is called *constructive necessitation*, which essentially is a justification counterpart of the modal necessitation rule.

COROLLARY 16 (Constructive Necessitation). *Let* $\mathsf{CS}$ *be an axiomatically appropriate constant specification. For any formula* $A \in \mathcal{L}_{\mathsf{J}}$*, if* $\mathsf{LP_{CS}} \vdash A$*, then* $\mathsf{LP_{CS}} \vdash t{:}A$ *for some ground term* $t \in \mathsf{Tm}$.

Combining the previous results, we also obtain internalization for the case when the assumptions are justified by arbitrary terms.

COROLLARY 17 (Internalization for Arbitrary Terms). *Let* $\mathsf{CS}$ *be an axiomatically appropriate constant specification. For arbitrary* $\mathcal{L}_{\mathsf{J}}$*-formulas* $A, B_1, \dots, B_n$ *and arbitrary terms* $s_1, \dots, s_n \in \mathsf{Tm}$*, if*

$$B_1, \dots, B_n \vdash_{\mathsf{LP_{CS}}} A \ .$$

*then there is a term* $t \in \mathsf{Tm}$ *such that*

$$s_1{:}B_1, \dots, s_n{:}B_n \vdash_{\mathsf{LP_{CS}}} t{:}A \ .$$

PROOF. Assume $B_1, \dots, B_n \vdash_{\mathsf{LP_{CS}}} A$. By Deduction Theorem 14,

$$\mathsf{LP_{CS}} \vdash B_1 \to (B_2 \to \cdots \to (B_n \to A) \cdots) \ .$$

By Constructive Necessitation, there is a ground term $s'$ such that

$$\mathsf{LP_{CS}} \vdash s'{:}\Big(B_1 \to (B_2 \to \cdots \to (B_n \to A) \cdots)\Big) \ .$$

By repeated applications of axiom (J) and modus ponens, we finally get for $t := s' \cdot s_1 \cdots s_n$,

$$s_1{:}B_1, \dots, s_n{:}B_n \vdash_{\mathsf{LP_{CS}}} t{:}A \ .$$

⊣

Definition 18 (Substitution). A *substitution* is a mapping $\sigma$ from variables to terms and from atomic propositions to formulas. Substitutions are then extended to terms and formulas in the obvious way: for an $\mathcal{L}_{\mathsf{J}}$-formula $B$, the formula $B\sigma$ is obtained from $B$ by simultaneously replacing all occurrences of $x$ with $\sigma(x)$ and all occurrences of $P$ with $\sigma(P)$ in $B$. For a set of formulas $\Phi$, we define $\Phi\sigma := \{B\sigma \mid B \in \Phi\}$ and for a constant specification $\mathsf{CS}$, we define $\mathsf{CS}\sigma := \{(c, A\sigma) \mid (c, A) \in \mathsf{CS}\}$.

Lemma 19 (Substitution). *For any constant specification* $\mathsf{CS}$*, any set of formulas* $\Delta \subseteq \mathcal{L}_{\mathsf{J}}$*, any formula* $A \in \mathcal{L}_{\mathsf{J}}$*, and any substitution* $\sigma$*,*

$$\Delta \vdash_{\mathsf{LP}_{\mathsf{CS}}} A \quad \textit{implies} \quad \Delta\sigma \vdash_{\mathsf{LP}_{\mathsf{CS}\sigma}} A\sigma \ .$$

Proof. Left as an exercise. ⊣

## 2C. Basic Modular Models

Definition 20. Let $X, Y \subseteq \mathcal{L}_{\mathsf{J}}$ and $t \in \mathsf{Tm}$. We define

1. $X{\cdot}Y := \{F \in \mathcal{L}_{\mathsf{J}} \mid G \to F \in X$ and $G \in Y$ for some formula $G \in \mathcal{L}_{\mathsf{J}}\}$;
2. $t{:}X := \{t{:}F \mid F \in X\}$.

Definition 21 (Basic evaluation). A *basic evaluation for* $\mathsf{LP}_{\mathsf{CS}}$, or a basic $\mathsf{LP}_{\mathsf{CS}}$-evaluation, is a function $*$ that maps atomic propositions to truth values $0, 1$ and maps justification terms to sets of formulas, $*\colon \mathsf{Prop} \to \{0,1\}$ and $*\colon \mathsf{Tm} \to \mathcal{P}(\mathcal{L}_{\mathsf{J}})$, such that for arbitrary $s, t \in \mathsf{Tm}$ and any $F \in \mathcal{L}_{\mathsf{J}}$,

1. $s^* \cdot t^* \subseteq (s \cdot t)^*$;
2. $s^* \cup t^* \subseteq (s + t)^*$;
3. $F \in t^*$ if $(t, F) \in \mathsf{CS}$;
4. $s{:}(s^*) \subseteq (!s)^*$.

Here $P^*$ for $P \in \mathsf{Prop}$ and $t^*$ for $t \in \mathsf{Tm}$ denote $*(P)$ and $*(t)$ respectively.

Definition 22 (Truth under a basic evaluation). We define what it means for a formula to *hold under a basic evaluation* $*$ inductively as follows:

$* \Vdash P$ if and only if $P^* = 1$ for $P \in \mathsf{Prop}$;
$* \Vdash F \to G$ if and only if $* \nVdash F$ or $* \Vdash G$;
$* \Vdash \neg F$ if and only if $* \nVdash F$;
$* \Vdash t{:}F$ if and only if $F \in t^*$.

Definition 23 (Factive evaluation). A basic $\mathsf{LP}_{\mathsf{CS}}$-evaluation $*$ is called *factive* if $F \in t^*$ implies $* \Vdash F$ for all $t \in \mathsf{Tm}$ and $F \in \mathcal{L}_{\mathsf{J}}$.

Definition 24 (Basic modular model). A *basic modular model for* $\mathsf{LP}_{\mathsf{CS}}$, or a basic modular $\mathsf{LP}_{\mathsf{CS}}$-model, is a basic $\mathsf{LP}_{\mathsf{CS}}$-evaluation $*$ that is factive.

THEOREM 25 (Soundness). *For any constant specification* $\mathsf{CS}$ *and any* $F \in \mathcal{L}_{\mathsf{J}}$ *we have*

$$\mathsf{LP}_{\mathsf{CS}} \vdash F \qquad \Longrightarrow \qquad * \Vdash F \text{ for all basic modular } \mathsf{LP}_{\mathsf{CS}}\text{-models } * .$$

PROOF. As usual, the proof is by induction on the length of the derivation of $F$. Let $*$ be a basic modular $\mathsf{LP}_{\mathsf{CS}}$-model. It is obvious that all instances of propositional axioms hold under $*$ and the rule (MP) is respected by the semantics. Soundness of the axioms (J), (+), and (j4), as well as that of the rule (AN), immediately follows from the definition of a basic evaluation.

It is also easy to see that all instances of (jt) hold under all factive basic evaluations. The argument for is as follows: if $* \Vdash t{:}F$, then $F \in t^*$, so $* \Vdash F$ by factivity of $*$. ⊣

Completeness is established by a maximal consistent set construction.

THEOREM 26 (Completeness). *For any constant specification* $\mathsf{CS}$ *and any* $F \in \mathcal{L}_{\mathsf{J}}$ *we have*

$$* \Vdash F \text{ for all basic modular } \mathsf{LP}_{\mathsf{CS}}\text{-models } * \qquad \Longrightarrow \qquad \mathsf{LP}_{\mathsf{CS}} \vdash F \ .$$

PROOF. Assume that $\mathsf{LP}_{\mathsf{CS}} \nvdash F$. Then $\{\neg F\}$ is $\mathsf{LP}_{\mathsf{CS}}$-consistent and, hence, is contained in some maximal $\mathsf{LP}_{\mathsf{CS}}$-consistent set $\Phi$. For this $\Phi$, any $P \in \mathsf{Prop}$, and any $t \in \mathsf{Tm}$, we define

$$P^* := \begin{cases} 1 & P \in \Phi \\ 0 & P \notin \Phi \end{cases} \qquad \text{and} \qquad t^* := \{F \in \mathcal{L}_{\mathsf{J}} \mid t{:}F \in \Phi\} \ . \tag{5}$$

It is easy to show that $*$ is a basic $\mathsf{LP}_{\mathsf{CS}}$-evaluation. By way of example, we will show Condition (1); the rest is similar. Suppose $A \in s^* \cdot t^*$. Then there is $B \in \mathcal{L}_{\mathsf{J}}$ such that $B \to A \in s^*$ and $B \in t^*$. By (5), $s{:}(B \to A) \in \Phi$ and $t{:}B \in \Phi$. By the maximal $\mathsf{LP}_{\mathsf{CS}}$-consistency of $\Phi$, also $(s \cdot t){:}A \in \Phi$. Thus, $A \in (s \cdot t)^*$ by (5).

We now show the so-called Truth Lemma: for all $D \in \mathcal{L}_{\mathsf{J}}$,

$$D \in \Phi \qquad \Longleftrightarrow \qquad * \Vdash D \ . \tag{6}$$

We establish (6) by induction on the structure of $D$:

1. $D = P \in \mathsf{Prop}$. Then $P \in \Phi \quad \Leftrightarrow \quad P^* = 1 \quad \Leftrightarrow \quad * \Vdash P$.
2. The cases when $D = \neg A$ and $D = A \to B$ are standard.
3. $D = t{:}A$. Then $t{:}A \in \Phi \quad \Leftrightarrow \quad A \in t^* \quad \Leftrightarrow \quad * \Vdash t{:}A$.

To show that $*$ is a basic modular $\mathsf{LP}_{\mathsf{CS}}$-model, we need to check that $*$ is factive. Suppose $F \in t^*$. Then $t{:}F \in \Phi$ by (5). Since (jt) is an axiom of $\mathsf{LP}_{\mathsf{CS}}$, we have $F \in \Phi$. Now $* \Vdash F$ follows by (6).

Since $\neg F \in \Phi$ by the construction of $\Phi$, we find $* \Vdash \neg F$ by (6). Thus, $* \nVdash F$ for the constructed basic modular $\mathsf{LP}_{\mathsf{CS}}$-model $*$. Completeness of $\mathsf{LP}_{\mathsf{CS}}$ follows by contraposition. ⊣

REMARK 27. Note that the case for $D = t{:}A$ in the proof of (6) is much simpler than the corresponding case for $\Box$ in the proof of Lemma 11.

LEMMA 28 (Consistency). *For any constant specification* $\mathsf{CS}$*, the logic* $\mathsf{LP_{CS}}$ *is consistent, that is the empty set* $\emptyset$ *is* $\mathsf{LP_{CS}}$*-consistent.*

PROOF. Suppose $\emptyset$ is not $\mathsf{LP_{CS}}$-consistent. That means

$$\mathsf{LP_{CS}} \vdash \bigwedge \emptyset \to \bot \ ,$$

which is

$$\mathsf{LP_{CS}} \vdash \top \to \bot \ .$$

Take now an $\mathsf{LP_{CS}}$-proof of $\top \to \bot$ and erase all justification terms (with ':'s) in each of its formulas. What remains is a chain of propositional formulas that are provable in classical propositional logic (an easy induction on the length of the original proof) and that ends with $\top \to \bot$, a contradiction. ⊣

An immediate consequence of the consistency of $\mathsf{LP_{CS}}$ is that $\mathsf{LP_{CS}}$ has a model. Indeed, since $\emptyset$ is $\mathsf{LP_{CS}}$-consistent, we have $\mathsf{LP_{CS}} \nvdash \top \to \bot$. By the completeness theorem, there exists a basic modular $\mathsf{CS}$-model that falsifies $\top \to \bot$.

## 2D. Epistemic Models

DEFINITION 29 (Quasimodel). We define a *quasimodel for* $\mathsf{LP_{CS}}$, or an $\mathsf{LP_{CS}}$-quasimodel, to be a triple $\mathcal{M} = (W, R, *)$, where $W \neq \varnothing$, $R \subseteq W \times W$, and the *evaluation* $*$ maps each world $w \in W$ to a basic $\mathsf{LP_{CS}}$-evaluation $*_w$. We will write $P^*_w$ instead of $*_w(P)$ and $t^*_w$ instead of $*_w(t)$.

DEFINITION 30 (Truth in quasimodels). We define what it means for a formula to *hold at a world* $w \in W$ *of a quasimodel* $\mathcal{M} = (W, R, *)$ inductively as follows:

$\mathcal{M}, w \Vdash P$ if and only if $P^*_w = 1$ for $P \in \mathsf{Prop}$;
$\mathcal{M}, w \Vdash F \to G$ if and only if $\mathcal{M}, w \nVdash F$ or $\mathcal{M}, w \Vdash G$;
$\mathcal{M}, w \Vdash \neg F$ if and only if $\mathcal{M}, w \nVdash F$;
$\mathcal{M}, w \Vdash t{:}F$ if and only if $F \in t^*_w$.

We write $\mathcal{M} \Vdash F$ if $\mathcal{M}, w \Vdash F$ for all $w \in W$.

For a given quasimodel $\mathcal{M} = (W, R, *)$ and a world $w \in W$, we define

$$\Box_w := \{F \in \mathcal{L}_\mathsf{J} \mid \mathcal{M}, v \Vdash F \text{ whenever } R(w, v)\} \ . \tag{7}$$

By analogy with basic modular models, we define the following notion:

DEFINITION 31 (Factive quasimodel). A quasimodel $\mathcal{M} = (W, R, *)$ is called *factive* if $F \in t^*_w$ implies $\mathcal{M}, w \Vdash F$ for all $w \in W$, $t \in \mathsf{Tm}$, and $F \in \mathcal{L}_\mathsf{J}$.

DEFINITION 32 (Modular model). A *modular model for* $\mathsf{LP}_{\mathsf{CS}}$, or a modular $\mathsf{LP}_{\mathsf{CS}}$-model, is an $\mathsf{LP}_{\mathsf{CS}}$-quasimodel $\mathcal{M} = (W, R, *)$ that meets the following conditions:

1. $t^*_w \subseteq \Box_w$ for all $t \in \mathsf{Tm}$ and $w \in W$; (JYB)
2. $R$ is reflexive;
3. $R$ is transitive.

These conditions may seem superfluous since $R$ plays no role in determining the truth of formulas. But Conditions (2) and (3) are well known for the corresponding modal axioms in modal logic and, hence, are needed, so to say, for backward compatibility: they ensure that the same semantics can be used for justification logics, logics of justifications and belief/knowledge, and modal logics. Condition (1) plays, in this respect, the role of a catalyzer allowing for a transition between these three formalisms. This condition essentially says that *justification yields belief*, abbreviated JYB. Indeed, whenever $F \in t^*_w$, we have $\mathcal{M}, w \Vdash t{:}F$ so that $F$ has a justification at $w$. The requirement that $F$ belong to $\Box_w$ says that $F$ must be believed at $w$ in the sense of Kripke models, i.e., hold at all worlds considered possible at $w$.

Note that, unlike for the case of basic modular models, we do not require that modular models be factive. Instead, this property is derived from JYB and the reflexicity restriction on $R$.

LEMMA 33 (Modular models are factive). *Let $\mathcal{M} = (W, R, *)$ be a modular $\mathsf{LP}_{\mathsf{CS}}$-model. Then $\mathcal{M}$ is factive.*

PROOF. Suppose $F \in t^*_w$. Then $F \in \Box_w$ by JYB. Since $R(w, w)$ by reflexivity of $R$, we obtain $\mathcal{M}, w \Vdash F$ from (7). $\dashv$

There is an additional property that follows from JYB but is peculiar to the possible-worlds scenario.

LEMMA 34 (Monotonicity). *Let $\mathcal{M} = (W, R, *)$ be a modular $\mathsf{LP}_{\mathsf{CS}}$-model. Then for any $t \in \mathsf{Tm}$ and for arbitrary $a, b \in W$, $R(a, b)$ implies $t^*_a \subseteq t^*_b$.*

PROOF. Assume $R(a, b)$ and $F \in t^*_a$. Then $t{:}F \in (!t)^*_a$ because $*_a$ is a basic evaluation for $\mathsf{LP}_{\mathsf{CS}}$. So $t{:}F \in \Box_a$ by JYB and $\mathcal{M}, b \Vdash t{:}F$ by (7), which means that $F \in t^*_b$. $\dashv$

The soundness and completeness of justification logics with respect to modular models are almost obvious:

THEOREM 35 (Soundness and Completeness, Modular Models). *For any constant specificaiton $\mathsf{CS}$ and any $F \in \mathcal{L}_{\mathsf{J}}$ we have*

$$\mathsf{LP}_{\mathsf{CS}} \vdash F \qquad \Longleftrightarrow \qquad \mathcal{M} \Vdash F \text{ for all modular } \mathsf{LP}_{\mathsf{CS}}\text{-models } \mathcal{M}\ . \tag{8}$$

Proof. It is sufficient to prove that any formula refutable by a basic modular model can be refuted at a world in a modular model and vice versa.

*Soundness.* Since $R$ plays no role in the definition of truth in modular models, for any modular model $\mathcal{M} = (W, R, *)$ and for any world $w \in W$, the basic $\mathsf{LP}_{\mathsf{CS}}$-evaluation $*_w$ satisfies exactly the same formulas as the world $w$ of $\mathcal{M}$ does, i.e., for all formulas $G$

$$\mathcal{M}, w \Vdash G \qquad \Longleftrightarrow \qquad *_w \Vdash G \ . \tag{9}$$

In particular, $*_w$ is factive since $\mathcal{M}$ is by Lemma 33. Hence $*_w$ is a basic modular $\mathsf{LP}_{\mathsf{CS}}$-model. By soundness of $\mathsf{LP}_{\mathsf{CS}}$ with respect to basic modular models, we get for any theorem $F$ of $\mathsf{LP}_{\mathsf{CS}}$ that $*_w \Vdash F$. Hence by (9) we get for any model $\mathcal{M} = (W, R, *)$, any world $w \in W$, and and theorem $F$ of $\mathsf{LP}_{\mathsf{CS}}$

$$\mathcal{M}, w \Vdash F \ .$$

*Completeness.* For the opposite direction, suppose $\mathsf{LP}_{\mathsf{CS}} \nvdash F$. By completeness of $\mathsf{LP}_{\mathsf{CS}}$ with respect to basic modular models, there exists a basic modular model $*$ such that $* \nVdash F$. We define an $\mathsf{LP}_{\mathsf{CS}}$-quasimodel $\mathcal{M} := (\{1\}, R, \star)$ with $\star_1 := *$. Hence by (9) we have $* \Vdash G$ iff $\mathcal{M}, 1 \Vdash G$ for all formulas $G$. To show that $\mathcal{M}$ is a modular $\mathsf{LP}_{\mathsf{CS}}$-model, which refutes all formulas refuted by $*$, it remains to make sure all the restrictions on $R$ and the condition JYB are met. We set $R := \{(1, 1)\}$, which is reflexive and transitive, so all restrictions on $R$ are met. Further, if $G \in t_1^\star$, then $G \in t^*$ by the definition of $\star$. Since $*$ is factive, $* \Vdash F$. Thus, $\mathcal{M}, 1 \Vdash G$ by (9) and, consequently, $G \in \Box_1$. Thus, $\mathcal{M}$ meets JYB and is a modular $\mathsf{LP}_{\mathsf{CS}}$-model such that $\mathcal{M} \nVdash F$. ⊣

It is reasonable to ask whether our semantics supports that every belief is evidenced by some justification. That is whether for all $F$

if $F$ holds in all accessible worlds, then there is a justification for $F$.

Definition 36 (Fully explanatory modular models). A modular $\mathsf{LP}_{\mathsf{CS}}$-model $\mathcal{M} = (W, R, *)$ is *fully explanatory* if for any $w \in W$ and any $F \in \mathcal{L}_{\mathsf{J}}$, $F \in \Box_w$ implies $F \in t_w^*$ for some $t \in \mathsf{Tm}$.

This notion can be seen as the converse of JYB and, taking the latter into account, can be reformulated as $\Box_w = \bigcup_{t \in \mathsf{Tm}} t_w^*$. Let us state the following theorem without proof.

Theorem 37. *Let* $\mathsf{CS}$ *be an axiomatically appriopriate constant specification. Then* $\mathsf{LP}_{\mathsf{CS}}$ *is sound and complete with respect to fully explanatory modular* $\mathsf{LP}_{\mathsf{CS}}$*-models.*

Proof. Soundness immediately follows from soundness with respect to modular $\mathsf{LP}_{\mathsf{CS}}$-models.

The hard task is to show completeness. We reuse the construction of a basic modular model $*_\Phi$ based on an $\mathsf{LP}_{\mathsf{CS}}$-consistent set $\Phi$ from the proof of Theorem 26. From all such $*_\Phi$ we create the canonical modular $\mathsf{LP}_{\mathsf{CS}}$-model and use (9) to transfer the properties of $*_\Phi$ proved for Theorem 26. Thus, we define $\mathcal{M}_c := (W, R, *)$, where

$$W := \{\Phi \subseteq \mathcal{L}_{\mathsf{J}} \mid \Phi \text{ is a maximal } \mathsf{LP}_{\mathsf{CS}}\text{-consistent set}\}$$

and $*_\Phi$ for each $\Phi \in W$ is defined by (5). Finally, we set $R(\Phi, \Psi)$ iff $\Phi/\sharp \subseteq \Psi$, where

$$\Phi/\sharp := \{F \in \mathcal{L}_{\mathsf{J}} \mid t{:}F \in \Phi \text{ for some term } t\} \ .$$

To show that $\mathcal{M}_c$ is a modular $\mathsf{LP}_{\mathsf{CS}}$-model, it remains to establish that the set $W$ is non-empty, reflexivity and transitivity of $R$, and the condition JYB.

We start with showing $W \neq \emptyset$. By Lemma 28 we know that the empty set is $\mathsf{LP}_{\mathsf{CS}}$-consistent. Thus by the Lindenbaum lemma, there exists a maximal $\mathsf{LP}_{\mathsf{CS}}$-consistent set, which is an element of $W$.

To show JYB, let $F \in t^*_\Phi$. Then $t{:}F \in \Phi$ by the definition of $*_\Phi$ and $F \in \Psi$ whenever $R(\Phi, \Psi)$ by the definition of $R$. By (6), $*_\Psi \Vdash F$, and $\mathcal{M}_c, \Psi \Vdash F$ by (9). Since $\Psi$ is chosen arbitrarily, $F \in \Box_\Phi$.

*R* is reflexive. Assume $t{:}F \in \Phi$. By maximal $\mathsf{LP}_{\mathsf{CS}}$-consistency of $\Phi$ and since $t{:}F \to F$ is an axiom of $\mathsf{LP}_{\mathsf{CS}}$ we conclude $F \in \Phi$. Therefore, $\Phi/\sharp \subseteq \Phi$, which means $R(\Phi, \Phi)$.

*R* is transitive. Assume $R(\Phi, \Psi)$ and $R(\Psi, \Delta)$ and let $t{:}F \in \Phi$. By maximal $\mathsf{LP}_{\mathsf{CS}}$-consistency of $\Phi$ and since $t{:}F \to\, !t{:}t{:}F$ is an axiom of $\mathsf{LP}_{\mathsf{CS}}$ we conclude $!t{:}t{:}F \in \Phi$. Hence $t{:}F \in \Psi$ and $F \in \Delta$. Therefore, $\Phi/\sharp \subseteq \Delta$, which means $R(\Phi, \Delta)$.

Finally, we show that $\mathcal{M}_c$ is fully explanatory. Assume towards a contradiction that $F \in \Box_\Phi$ for some $F \in \mathcal{L}_{\mathsf{J}}$ and $\Phi \in W$ but $F \notin t^*_\Phi$ for any $t \in \mathsf{Tm}$. Then $\Phi/\sharp \cup \{\neg F\}$ would be $\mathsf{LP}_{\mathsf{CS}}$-consistent.

Indeed, if $\Phi/\sharp \cup \{\neg F\}$ were $\mathsf{LP}_{\mathsf{CS}}$-inconsistent, then $G_1, \dots, G_n, \neg F \vdash_{\mathsf{LP}_{\mathsf{CS}}} \bot$ for some $G_1, \dots, G_n \in \Phi/\sharp$. Equivalently, there would be terms $s_1, \dots, s_n$ such that $s_i{:}G_i \in \Phi$ for $i = 1, \dots, n$ and $G_1, \dots, G_n \vdash_{\mathsf{LP}_{\mathsf{CS}}} F$. By Corollary 17, given the axiomatic appropriateness of $\mathsf{CS}$, there would be a term $t$ such that $s_1{:}G_1, \dots, s_n{:}G_n \vdash_{\mathsf{LP}_{\mathsf{CS}}} t{:}F$. By Deduction Theorem 14,

$$\mathsf{LP}_{\mathsf{CS}} \vdash s_1{:}G_1 \to (s_2{:}G_2 \to \cdots \to (s_n{:}G_n \to t{:}F) \cdots)$$

so that $t{:}F \in \Phi$ by the maximal $\mathsf{LP}_{\mathsf{CS}}$-consistency of $\Phi$ and $F \in t^*_\Phi$ by (5), contradicting our assumption.

Hence, the set $\Phi/\sharp \cup \{\neg F\}$ would be $\mathsf{LP}_{\mathsf{CS}}$-consistent and could be extended to a maximal $\mathsf{LP}_{\mathsf{CS}}$-consistent set $\Psi$. Clearly, $R(\Phi, \Psi)$ by the definition of $R$ and $\mathcal{M}_c, \Psi \Vdash \neg F$ by (6) and (9). Thus, $\mathcal{M}_c, \Psi \nVdash F$, which contradicts our assumption that $F \in \Box_\Phi$. $\dashv$

## 2E. Justifications and Belief

It should not be surprising that modular models can also be used for the joint language of justifications and belief. And while the condition JYB does not look out of place in justification logics, its real origins are, of course, modal, which is clearly seen in the following soundness proof. Many notions and conventions introduced in Sect. 2A are now generalized to the extended *language* $\mathcal{L}_\Box$ defined by the grammar:

$$F ::= P \mid \neg F \mid (F \to F) \mid t{:}F \mid \Box F$$

where $P \in \mathsf{Prop}$. The *set of axioms* $\mathsf{LP}^\Box$ consists of

- all the axioms of $\mathsf{LP}$ in the extended language $\mathcal{L}_\Box$;
- axiom $\Box(A \to B) \to (\Box A \to \Box B)$;
- axiom $\Box F \to F$;
- axiom $\Box F \to \Box\Box F$;
- axiom $t{:}F \to \Box F$.

The axiom $t{:}F \to \Box F$ is called the *connection axiom.* It formally states that justification yields belief. *Constant specifications for* $\mathsf{LP}^\Box$ are defined in the obvious way. Given a constant specification $\mathsf{CS}$ for $\mathsf{LP}^\Box$, the *deductive system* $\mathsf{LP}^\Box_{\mathsf{CS}}$ is the Hilbert system given by the axioms $\mathsf{LP}^\Box$ and by the rules (MP) and (AN), as in $\mathsf{LP}_{\mathsf{CS}}$, as well as by the usual necessitation rule from modal logic:

$$\frac{F}{\Box F}\ (\Box)\ .$$

A *basic evaluation for* $\mathsf{LP}^\Box_{\mathsf{CS}}$ and many other notions are defined in the same way as for $\mathsf{LP}_{\mathsf{CS}}$ except that each instance of the language $\mathcal{L}_{\mathsf{J}}$ should be replaced with $\mathcal{L}_\Box$ and those of the logic $\mathsf{LP}_{\mathsf{CS}}$ with $\mathsf{LP}^\Box_{\mathsf{CS}}$. In particular, the definition of a *modular model for* $\mathsf{LP}^\Box_{\mathsf{CS}}$ repeats that for $\mathsf{LP}_{\mathsf{CS}}$ with one extra clause for the truth of formulas $\Box F$:

$$\mathcal{M}, w \Vdash \Box F \qquad \Longleftrightarrow \qquad F \in \Box_w\ , \tag{10}$$

which is a standard definition recast in our notation.

THEOREM 38 (Soundness of Modular Models for $\mathsf{LP}^\Box_{\mathsf{CS}}$). *For any constant specification* $\mathsf{CS}$ *and any* $F \in \mathcal{L}_\Box$ *we have*

$$\mathsf{LP}^\Box_{\mathsf{CS}} \vdash F \qquad \Longrightarrow \qquad \mathcal{M} \Vdash F \textit{ for all modular } \mathsf{LP}^\Box_{\mathsf{CS}}\textit{-models } \mathcal{M}\ .$$

PROOF. Most of the proof repeats that of Theorem 25 or the standard argument for the modal axioms. The connection axiom $t{:}F \to \Box F$ is valid because of JYB and (10). ⊣

THEOREM 39 (Completeness of Modular Models for $\mathsf{LP}^{\Box}_{\mathsf{CS}}$). *For any constant specification* $\mathsf{CS}$ *and any* $F \in \mathcal{L}_{\Box}$ *we have*

$$\mathcal{M} \Vdash F \text{ for all modular } \mathsf{LP}^{\Box}_{\mathsf{CS}}\text{-models } \mathcal{M} \qquad \Longrightarrow \qquad \mathsf{LP}^{\Box}_{\mathsf{CS}} \vdash F \ .$$

PROOF. We define the canonical model $\mathcal{M}^{\Box}_{c} = (W, R, *)$ as follows:

$$W := \{\Phi \subseteq \mathcal{L}_{\Box} \mid \Phi \text{ is a maximal } \mathsf{LP}^{\Box}_{\mathsf{CS}}\text{-consistent set}\}$$

and $*_{\Phi}$ for each $\Phi \in W$ is defined by (5) except that $t^{*}_{\Phi}$ consists of $\mathcal{L}_{\Box}$-formulas instead of $\mathcal{L}_{\mathsf{J}}$-formulas. Finally, we set $R(\Phi, \Psi)$ iff $\Phi/\Box \subseteq \Psi$.

As usual, the Truth Lemma is established by induction on $D$: for all formulas $D \in \mathcal{L}_{\Box}$ and all maximal $\mathsf{LP}^{\Box}_{\mathsf{CS}}$-consistent sets $\Phi$,

$$D \in \Phi \qquad \Longleftrightarrow \qquad \mathcal{M}^{\Box}_{c}, \Phi \Vdash D \ . \tag{11}$$

The cases for propositions and Boolean connectives are straightforward. The case for $D = t{:}F$ does not involve $R$ and is essentially the same as in the proof of Theorem 26. The case for $D = \Box F$ is proved by the standard modal argument because $R$ is defined as in the modal canonical model.

It remains to show that $\mathcal{M}^{\Box}_{c}$ is a modular $\mathsf{LP}^{\Box}_{\mathsf{CS}}$-model. The proof that $*_{\Phi}$ is a basic $\mathsf{LP}^{\Box}_{\mathsf{CS}}$-evaluation is almost literally the same as in Theorem 26. The conditions on $R$ are established by the standard modal argument.

Thus, the proof of JYB is the only thing that needs to be done. Suppose $F \in t^{*}_{\Phi}$. Then $t{:}F \in \Phi$ by (5). Using the axiom instance $t{:}F \to \Box F$ and the maximal $\mathsf{LP}^{\Box}_{\mathsf{CS}}$-consistency of $\Phi$, we conclude that $\Box F \in \Phi$. Hence, $F \in \Psi$ whenever $R(\Phi, \Psi)$ by the definition of $R$. Thus, $\mathcal{M}, \Psi \Vdash F$ whenever $R(\Phi, \Psi)$ by (11), i.e., $F \in \Box_{\Phi}$. ⊣

COROLLARY 40 (Conservativity). *For any constant specification* $\mathsf{CS}$ *and any* $F \in \mathcal{L}_{\mathsf{J}}$ *we have*

$$\mathsf{LP}_{\mathsf{CS}} \vdash F \quad \Longleftrightarrow \quad \mathsf{LP}^{\Box}_{\mathsf{CS}} \vdash F \ .$$

PROOF. The direction form left to right is immediate since the axioms and rules of $\mathsf{LP}^{\Box}_{\mathsf{CS}}$ contain those of $\mathsf{LP}_{\mathsf{CS}}$. Hence every $\mathsf{LP}_{\mathsf{CS}}$-proof is also an $\mathsf{LP}^{\Box}_{\mathsf{CS}}$-proof.

For the direction from right to left suppose $\mathsf{LP}_{\mathsf{CS}} \nvdash F$. By Completeness Theorem 35 there exists a modular $\mathsf{LP}_{\mathsf{CS}}$-model $\mathcal{M}$ such that $\mathcal{M} \nVdash F$. Since every modular $\mathsf{LP}_{\mathsf{CS}}$-model also is a modular $\mathsf{LP}^{\Box}_{\mathsf{CS}}$-model, we get that exists a modular $\mathsf{LP}^{\Box}_{\mathsf{CS}}$-model $\mathcal{M}$ such that $\mathcal{M} \nVdash F$. By soundness of $\mathsf{LP}^{\Box}_{\mathsf{CS}}$ (Theorem 38) we conclude $\mathsf{LP}^{\Box}_{\mathsf{CS}} \nvdash F$. ⊣

$\mathsf{LP}^{\Box}_{\mathsf{CS}}$ is also conservative over $\mathsf{S4}$ with respect to $\mathcal{L}$-formulas. Instead of giving a semantic proof as above, we will establish this result by syntactic means, see Lemma 43 and its proof in the next chapter.

CHAPTER 3

# RELATIONS WITH $\mathsf{S4}$

## 3A. Forgetful Projection

DEFINITION 41 (Forgetful projection). The mapping $\circ : \mathcal{L}_{\mathsf{J}} \to \mathcal{L}$ is defined as follows

1. $P^\circ := P$ for $P \in \mathsf{Prop}$;
2. $(\neg A)^\circ := \neg A^\circ$;
3. $(A \to B)^\circ := A^\circ \to B^\circ$;
4. $(t{:}A)^\circ := \Box A^\circ$.

LEMMA 42 (Forgetful projection). *For any constant specification* $\mathsf{CS}$ *and any formula* $F$ *of* $\mathcal{L}_{\mathsf{J}}$ *we have*

$$\mathsf{LP}_{\mathsf{CS}} \vdash F \quad \textit{implies} \quad \mathsf{S4} \vdash F^\circ \ .$$

PROOF. By induction on the depth of the $\mathsf{LP}_{\mathsf{CS}}$ derivation.

1. $F$ is an instance of $(\mathsf{P})$. Then $F^\circ$ is a propositional tautology in the language $\mathcal{L}$ and, therefore, $\mathsf{S4} \vdash F^\circ$.
2. $F = t{:}(A \to B) \to (s{:}A \to (t \cdot s){:}B)$. Then
$$F^\circ = \Box(A \to B) \to (\Box A \to \Box{:}B) \ ,$$
which is an instance of $(\mathsf{K})$.
3. $F = r{:}A \to (t+s){:}A$ where $r = t$ or $r = s$. Then $F^\circ = \Box A \to \Box A$, which is a tautology and, therefore, $\mathsf{S4} \vdash F^\circ$.
4. $F = t{:}A \to A$. Then $F^\circ = \Box A \to A$, which is an instance of $(\mathsf{T})$.
5. $F = t{:}A \to\ !t{:}t{:}A$. Then $F^\circ = \Box A \to \Box\Box A$, which is an instance of $(\mathsf{4})$.
6. $F$ is the conclusion of an instance of $(\mathsf{MP})$ with the premises $A$ and $A \to F$. By I.H. we have $\mathsf{S4} \vdash A^\circ$ and $\mathsf{S4} \vdash (A \to F)^\circ$. Since by definition $(A \to F)^\circ = A^\circ \to F^\circ$, we can apply $(\mathsf{MP})$ and conclude $\mathsf{S4} \vdash F^\circ$.
7. $F$ is the conclusion of an instance of $(\mathsf{AN})$. Then $F$ has the form $c{:}A$ where $A$ is an axiom if $\mathsf{LP}_{\mathsf{CS}}$. Thus, we know $\mathsf{S4} \vdash A^\circ$. By $(\mathsf{NEC})$ we get $\mathsf{S4} \vdash \Box A^\circ$, which is $\mathsf{S4} \vdash F^\circ$.

⊣

We can employ a forgetful projection also to show that $\mathsf{LP}^{\Box}_{\mathsf{CS}}$ is a conservative extension of $\mathsf{S4}$ with respect to $\mathcal{L}$-formulas.

First, we extend the mapping $\circ$ such that $\circ : \mathcal{L}_{\Box} \to \mathcal{L}$ by adding

$$(\Box A)^{\circ} := \Box A^{\circ}$$

to its definition. Now we easily get the following lemma.

Lemma 43 (Conservativity). *For any constant specification* $\mathsf{CS}$ *and any* $F \in \mathcal{L}$ *we have*

$$\mathsf{S4} \vdash F \quad \Longleftrightarrow \quad \mathsf{LP}^{\Box}_{\mathsf{CS}} \vdash F \enspace .$$

Proof. The direction form left to right is immediate since the axioms and rules of $\mathsf{LP}^{\Box}_{\mathsf{CS}}$ contain those of $\mathsf{S4}$. Hence every $\mathsf{S4}$-proof is also an $\mathsf{LP}^{\Box}_{\mathsf{CS}}$-proof.

For the direction from right to left suppose we have an $\mathsf{LP}^{\Box}_{\mathsf{CS}}$-proof of $F$. If we apply the forgetful projection to all formulas in that proof, we obtain a chain of formulas that are provable in $\mathsf{S4}$ and that ends in $F$. ⊣

## 3B. Realization

Definition 44 (Realization). A *realization* is a mapping $r : \mathcal{L} \to \mathcal{L}_{\mathsf{J}}$ such that $(r(A))^{\circ} = A$.

We can think of a realization as a function that replaces occurrences of $\Box$ in an $\mathcal{L}$ formula with evidence terms, thus creating a formula of $\mathcal{L}_{\mathsf{J}}$.

Definition 45. We say a justification logic $\mathsf{LP}_{\mathsf{CS}}$ *realizes* $\mathsf{S4}$ if there is a realization $r$ such that for any $\mathcal{L}$-formula $A$ we have

$$\mathsf{S4} \vdash A \quad \text{implies} \quad \mathsf{LP}_{\mathsf{CS}} \vdash r(A) \enspace .$$

Of course, not every $\mathsf{LP}_{\mathsf{CS}}$ realizes $\mathsf{S4}$. In particular, consider the empty constant specification $\emptyset$. In $\mathsf{LP}_{\emptyset}$ we cannot apply the axiom necessitation rule and thus we cannot establish constructive necessiation (Corollary 16) for $\mathsf{LP}_{\emptyset}$. That means we do not have an explicit analogue of $\mathsf{S4}$'s necessitation rule in $\mathsf{LP}_{\emptyset}$ and thus $\mathsf{LP}_{\emptyset}$ will not realize $\mathsf{S4}$.

The aim of this section is to establish a realization theorem. That is we are going to identify certain properties for constant specifications such that if a constant specification $\mathsf{CS}$ satisfies them, then $\mathsf{LP}_{\mathsf{CS}}$ will realize $\mathsf{S4}$.

For the rest of this section, we fix a modal formula $A$ of $\mathcal{L}$. Moreover, we assume that $\mathsf{CS}$ is an axiomatically appropriate constant specification and we let $\mathcal{M}_c = (W, R, *)$ be the canonical model from the proof of Theorem 37. Let us mention again two main properties of $\mathcal{M}_c$, which will be important for establishing a realization theorem. First, we have the Truth Lemma for $\mathcal{M}_c$

$$D \in \Phi \quad \Longleftrightarrow \quad \mathcal{M}_c, \Phi \Vdash D \ . \tag{12}$$

Second, we need the following property, which can be read off from the proof of Theorem 37.

$$\text{If } \Phi/\sharp \cup \{\neg C\} \text{ is not } \mathsf{LP}_{\mathsf{CS}}\text{-consistent, then there is a term } t \text{ such that } t{:}C \in \Phi. \tag{13}$$

Definition 46 (Polarity). We assign *polarities* to subformula occurrences of $A$ in the usual way.

1. $A$ itself is a positive subformula occurrence of $A$.
2. If $\neg B$ is a positive (negative) subformula occurrence of $A$, then $B$ is a negative (positive) subformula occurrence of $A$.
3. If $B \to C$ is a positive (negative) subformula occurrence of $A$, then $B$ is a negative (positive) subformula occurrence of $A$ and $C$ is a positive (negative) subformula occurrence of $A$.
4. If $\Box B$ is a positive (negative) subformula occurrence of $A$, then $B$ is a positive (negative) subformula occurrence of $A$.

Definition 47 (Annotation). An *annotation* for $A$ is a mapping $\mathcal{A}$ assigning variables to negatively occurring subformulas of the form $\Box B$ such that different occurrences are assigned different variables.

For the rest of this section, we also fix an annotation $\mathcal{A}$ for $A$. We will now define a mapping that assigns a set of *potential pre-realizations* to subformula occurrences of $A$ with respect to the annotation $\mathcal{A}$.

Definition 48 (Potential pre-realizations). The mapping

$$|\cdot| : \{B \text{ is a subformula occurrence of } A\} \to \mathcal{P}(\mathcal{L}_{\mathsf{J}})$$

from subformula occurrences of $A$ to sets of $\mathcal{L}_{\mathsf{J}}$-formulas is inductively defined as follows

1. $|P| := \{P\}$ for $P \in \mathsf{Prop}$;
2. $|\neg B| := \{\neg B' \mid B' \in |B|\}$;
3. $|B \to C| := \{B' \to C' \mid B' \in |B| \text{ and } C' \in |C|\}$;
4. if $\Box B$ is a positive subformula occurrence, then

   $$|\Box B| := \{t{:}(B'_1 \vee \cdots \vee B'_n) \mid t \in \mathsf{Tm} \text{ and } B'_1, \ldots B'_n \in |B|\} \ ;$$

5. if $\Box B$ is a negative subformula occurrence, then

   $$|\Box B| := \{x{:}B' \mid \mathcal{A}(\Box B) = x \text{ and } B' \in |B|\} \ .$$

We use $\neg|B|$ to denote the set $\{\neg B' \mid B' \in |B|\}$.

Since $B$ is a *subformula occurrence* of $A$, the set $\neg|B|$ is different from $|\neg B|$. In particular, $\neg|A|$ is defined while $|\neg A|$ is not.

We can consider $\mathcal{M}_c$ as a Kripke structure by simply ignoring the evaluation function for terms. Even though it is always clear from the context whether $\Vdash$ is used with respect to modular models or Kripke semantics (i.e. with respect to $\mathcal{L}_{\mathsf{J}}$-formulas or $\mathcal{L}$-formulas), we will sometimes write $\Vdash_{\mathsf{LP_{CS}}}$ and $\Vdash_{\mathsf{S4}}$, respectively, to emphasize this point.

For a set $\Gamma$ of formulas, $\mathcal{M}_c, \Phi \Vdash \Gamma$ means $\mathcal{M}_c, \Phi \Vdash B$ for all $B \in \Gamma$.

LEMMA 49. *Let* CS *be axiomatically appropriate.*

1. *If $B$ is a positive subformula occurrence of $A$ and $\mathcal{M}_c, \Phi \Vdash_{\mathsf{LP_{CS}}} \neg|B|$, then $\mathcal{M}_c, \Phi \Vdash_{\mathsf{S4}} \neg B$.*
2. *If $B$ is a negative subformula occurrence of $A$ and $\mathcal{M}_c, \Phi \Vdash_{\mathsf{LP_{CS}}} |B|$, then $\mathcal{M}_c, \Phi \Vdash_{\mathsf{S4}} B$.*

PROOF. The two claims are shown simultaneously by induction on the structure of $B$.

1. $B = P$ with $P \in \mathsf{Prop}$. This case is trivial since $|P| = \{P\}$.
2. $B = \neg C$ occurring positively. Assume $\mathcal{M}_c, \Phi \Vdash_{\mathsf{LP_{CS}}} \neg|\neg C|$. That is for each $C' \in |C|$ we have $\mathcal{M}_c, \Phi \Vdash_{\mathsf{LP_{CS}}} \neg(\neg C')$, which yields
$$\mathcal{M}_c, \Phi \Vdash_{\mathsf{LP_{CS}}} |C| \ .$$
Since $C$ occurs negatively, we get by I.H. $\mathcal{M}_c, \Phi \Vdash_{\mathsf{S4}} C$ and thus $\mathcal{M}_c, \Phi \Vdash_{\mathsf{S4}} \neg B$.
3. $B = \neg C$ occurring negatively. Assume $\mathcal{M}_c, \Phi \Vdash_{\mathsf{LP_{CS}}} |\neg C|$. That is for each $C' \in |C|$ we have $\mathcal{M}_c, \Phi \Vdash_{\mathsf{LP_{CS}}} \neg C'$, which yields
$$\mathcal{M}_c, \Phi \Vdash_{\mathsf{LP_{CS}}} \neg|C| \ .$$
Since $C$ occurs positively, we get by I.H. $\mathcal{M}_c, \Phi \Vdash_{\mathsf{S4}} \neg C$ and thus $\mathcal{M}_c, \Phi \Vdash_{\mathsf{S4}} B$.
4. $B = D \to C$ occurring positively. Assume $\mathcal{M}_c, \Phi \Vdash_{\mathsf{LP_{CS}}} \neg|D \to C|$. That is for each $D' \in |D|$ and each $C' \in |C|$ we have
$$\mathcal{M}_c, \Phi \Vdash_{\mathsf{LP_{CS}}} \neg(D' \to C') \ .$$
Thus we have $\mathcal{M}_c, \Phi \Vdash_{\mathsf{LP_{CS}}} D'$ for each $D' \in |D|$ and $\mathcal{M}_c, \Phi \Vdash_{\mathsf{LP_{CS}}} \neg C'$ for each $C' \in |C|$, which is
$$\mathcal{M}_c, \Phi \Vdash_{\mathsf{LP_{CS}}} |D| \quad \text{and} \quad \mathcal{M}_c, \Phi \Vdash_{\mathsf{LP_{CS}}} \neg|C| \ .$$
Since $D$ occurs negatively and $C$ occurs positively, we get by I.H.
$$\mathcal{M}_c, \Phi \Vdash_{\mathsf{S4}} D \quad \text{and} \quad \mathcal{M}_c, \Phi \Vdash_{\mathsf{S4}} \neg C \ .$$
Hence
$$\mathcal{M}_c, \Phi \Vdash_{\mathsf{S4}} \neg(D \to C) \ .$$

5. $B = D \to C$ occurring negatively. Assume $\mathcal{M}_c, \Phi \Vdash_{\mathsf{LP_{CS}}} |D \to C|$. We distinguish two cases. If $\mathcal{M}_c, \Phi \Vdash_{\mathsf{LP_{CS}}} \neg|D|$, then, as $D$ occurs positively, we get $\mathcal{M}_c, \Phi \Vdash_{\mathsf{S4}} \neg D$ by I.H. Thus $\mathcal{M}_c, \Phi \Vdash_{\mathsf{S4}} D \to C$ and we are done.

   Otherwise there is a $D' \in |D|$ with $\mathcal{M}_c, \Phi \Vdash_{\mathsf{LP_{CS}}} D'$. Since we have $\mathcal{M}_c, \Phi \Vdash_{\mathsf{LP_{CS}}} D' \to C'$ for any $C' \in |C|$ by assumption, we also get $\mathcal{M}_c, \Phi \Vdash_{\mathsf{LP_{CS}}} C'$ for any $C' \in |C|$, which is $\mathcal{M}_c, \Phi \Vdash_{\mathsf{LP_{CS}}} |C|$. Since $C$ occurs negatively, we get $\mathcal{M}_c, \Phi \Vdash_{\mathsf{S4}} C$ by I.H. and thus

$$\mathcal{M}_c, \Phi \Vdash_{\mathsf{S4}} D \to C \ .$$

6. $B = \Box C$ occurs positively. Assume $\mathcal{M}_c, \Phi \Vdash_{\mathsf{LP_{CS}}} \neg|\Box C|$. We are done if we can show

$$\Phi/\sharp \cup \neg|C| \text{ is } \mathsf{LP_{CS}}\text{-consistent.} \tag{14}$$

   For then there exists a maximal $\mathsf{LP_{CS}}$-consistent set $\Psi$ with

$$\Phi/\sharp \cup \neg|C| \subseteq \Psi \ .$$

   By definition of $R$, this means $R(\Phi, \Psi)$. Because of $\neg|C| \subseteq \Psi$ we get $\mathcal{M}_c, \Psi \Vdash_{\mathsf{LP_{CS}}} \neg|C|$ by the Truth Lemma (12) for $\mathcal{M}_c$. Since $C$ occurs positively, we get $\mathcal{M}_c, \Psi \Vdash_{\mathsf{S4}} \neg C$ by I.H. and thus $\mathcal{M}_c, \Phi \Vdash_{\mathsf{S4}} \neg\Box C$.

   Let us now show the open claim (14). Suppose towards a contradiction that $\Phi/\sharp \cup \neg|C|$ is not $\mathsf{LP_{CS}}$-consistent. That is there are

$$C_1, \dots, C_n \in |C|$$

   such that $\Phi/\sharp \cup \{\neg C_1, \dots, \neg C_n\}$ is not $\mathsf{LP_{CS}}$-consistent, which is

$$\Phi/\sharp \cup \{\bigwedge(\neg C_i)\} \text{ is not } \mathsf{LP_{CS}}\text{-consistent.}$$

   That is $\Phi/\sharp \cup \{\neg \bigvee C_i\}$ is not $\mathsf{LP_{CS}}$-consistent. By (13), this implies that there is a term $t$ such that $t{:}\bigvee C_i \in \Phi$. Hence by the Truth Lemma (12) for $\mathcal{M}_c$, we get $\mathcal{M}_c, \Phi \Vdash_{\mathsf{LP_{CS}}} t{:}\bigvee C_i$, which contradicts the assumption $\mathcal{M}_c, \Phi \Vdash_{\mathsf{LP_{CS}}} \neg|\Box C|$.
7. $B = \Box C$ occurs negatively. Assume $\mathcal{M}_c, \Phi \Vdash_{\mathsf{LP_{CS}}} |\Box C|$. Let $C' \in |C|$ and $\mathcal{A}(\Box C) = x$. We have $\mathcal{M}_c, \Phi \Vdash_{\mathsf{LP_{CS}}} x{:}C'$. By JYB we get for each $\Psi$ with $R(\Phi, \Psi)$ that $\mathcal{M}_c, \Psi \Vdash_{\mathsf{LP_{CS}}} C'$ holds. Since $C'$ is arbitrary, we have $\mathcal{M}_c, \Psi \Vdash_{\mathsf{LP_{CS}}} |C|$. By I.H. we obtain $\mathcal{M}_c, \Psi \Vdash_{\mathsf{S4}} C$ for each $\Psi$ with $R(\Phi, \Psi)$. Finally, this yields $\mathcal{M}_c, \Phi \Vdash_{\mathsf{S4}} \Box C$.

⊣

LEMMA 50. *Let* $\mathsf{CS}$ *be axiomatically appropriate. If* $\mathsf{S4} \vdash A$*, then there are* $A_1, \dots, A_n \in |A|$ *such that*

$$\mathsf{LP_{CS}} \vdash A_1 \vee \cdots \vee A_n \ .$$

Proof. We show the contrapositive. Assume for any $A_1, \ldots, A_n \in |A|$, we have $\mathsf{LP}_{\mathsf{CS}} \nvdash A_1 \vee \cdots \vee A_n$. That means $\neg|A|$ is consistent. So the canonical model $\mathcal{M}_c$ contains a world $\Phi$ with $\neg|A| \subseteq \Phi$. By the Truth Lemma (12) we get

$$\mathcal{M}_c, \Phi \Vdash_{\mathsf{LP}_{\mathsf{CS}}} \neg|A| \ .$$

As $A$ is a positive subformula of itself, we can use Lemma 49 to conclude

$$\mathcal{M}_c, \Phi \Vdash_{\mathsf{S4}} \neg A \ ,$$

which is $\mathcal{M}_c, \Phi \nVdash_{\mathsf{S4}} A$. Since $\mathcal{M}_c$ is not only a model of $\mathsf{LP}_{\mathsf{CS}}$ but also of $\mathsf{S4}$, we can use soundness of $\mathsf{S4}$ to conclude $\mathsf{S4} \nvdash A$. ⊣

The next step is to look at *potential realizations* and it remains to be shown that the previously defined potential pre-realizations can be turned into such potential realizations.

Definition 51 (Potential realizations). The mapping

$$\|\cdot\| : \{B \text{ is a subformula occurrence of } A\} \to \mathcal{P}(\mathcal{L}_{\mathsf{J}})$$

is defined like $|\cdot|$ with $\|\cdot\|$ in place of $|\cdot|$ everywhere except for the case of $\Box B$ being a positive subformula occurrence, which is defined by

$$\|\Box B\| := \{t{:}B' \mid B' \in \|B\| \text{ and } t \in \mathsf{Tm}\} \ .$$

Definition 52 (Negative occurrence of a variable). Let $B$ be a subformula of $A$ and $B' \in |B|$. We say an occurrence of a variable $x$ in $B'$ is *negative* if the corresponding occurrence of $\Box$ in $A$ is negative.

The previous definition means that an occurrence of $x$ is negative if it has been introduced by $|\cdot|$ because of a negatively occurring subformula $\Box C$ such that $\mathcal{A}(\Box C) = x$.

In this chapter, by *substitution* we mean *variable substitution*: a substitution that maps variables to terms but is the identity function on atomic propositions.

The *domain* $\mathsf{dom}(\sigma)$ of a substitution $\sigma$ is defined as

$$\mathsf{dom}(\sigma) := \{x \mid x \neq \sigma(x)\} \ .$$

A substitution $\sigma$ with finite domain $\mathsf{dom}(\sigma) = \{x_1, \ldots, x_n\}$ will also be denoted by $\{x_1/t_1, \ldots, x_n/t_n\}$ where $\sigma(x_i) = t_i$ for $1 \leq i \leq n$.

For two substitutions $\sigma_1$ and $\sigma_2$ with $\mathsf{dom}(\sigma_1) \cap \mathsf{dom}(\sigma_2) = \emptyset$ we define the substitution $\sigma_1 \cup \sigma_2$ by

$$(\sigma_1 \cup \sigma_2)(x) := \begin{cases} \sigma_1(x) & \text{if } x \in \mathsf{dom}(\sigma_1) \\ \sigma_2(x) & \text{if } x \in \mathsf{dom}(\sigma_2) \\ x & \text{otherwise.} \end{cases}$$

A substitution $\sigma$ is said to *live on a formula* $C' \in |C|$ for a subformula $C$ of $A$ if

$$\mathsf{dom}(\sigma) \subseteq \{x \mid x \text{ is a variable occurring negatively in } C'\} \ .$$

A substitution $\sigma$ satisfies the *no new variable condition* if for any variable $x$

$$\sigma(x) \text{ contains no variables other than } x.$$

Remark 53. We say a substitution $\sigma$ *lives away from* a formula $C' \in |C|$ for a subformula $C$ of $A$ if

$$\mathsf{dom}(\sigma) \cap \{x \mid x \text{ is a variable occurring negatively in } C'\} = \emptyset \ .$$

Assume we are given two substitution $\sigma_1$ and $\sigma_2$ such that

1. $\sigma_1$ lives on some formula $C'$,
2. $\sigma_2$ lives away from that $C'$, and
3. both substitutions meet the no new variable condition.

We then have $\sigma_1\sigma_2 = \sigma_2\sigma_1 = \sigma_1 \cup \sigma_2$.

However, the statement that $\sigma$ lives away from $C'$ does not imply that $C' = C'\sigma$ since variables in the domain of $\sigma$ still might occur positively in $C'$. This observation is important in the proof of Lemma 56 in the cases for implications and negatively occurring modalities.

Definition 54 (Schematic $\mathsf{CS}$). We say that a constant specification is *schematic* if it satisfies the following: for each constant c, the set of axioms $\{A \mid (c, A) \in \mathsf{CS}\}$ consists of all instances of one or several (possibly zero) axiom schemes of $\mathsf{LP}$.

Suppose for example that

$$A \to (B \to A) \tag{15}$$

and

$$(A \to B) \to ((B \to C) \to (A \to C)) \tag{16}$$

are two axiom schemes of $\mathsf{LP}$. Further suppose that the constant $c$ justifies all instances of (15) and that $d$ justifies all instances of (16). We then find $(c, P \to (Q \to P)) \in \mathsf{CS}$, but we may have $(d, P \to (Q \to P)) \notin \mathsf{CS}$ since $P \to (Q \to P)$ is not an instance of (16). Moreover, we observe that both

$$\big(c, (P \to P) \to ((P \to P) \to (P \to P))\big) \in \mathsf{CS}$$

and

$$\big(d, (P \to P) \to ((P \to P) \to (P \to P))\big) \in \mathsf{CS} \ .$$

Schematicness is important when we deal with substitutions for the following reason. Let $\sigma$ be a substitution. If $A$ is an instance of a given axiom scheme, then also $A\sigma$ is an instance of that scheme. Hence if $\mathsf{CS}$

is schematic and $(c, A) \in CS$, then we also have $(c, A\sigma) \in CS$. That is schematic constant specifications are closed under substitutions. Thus we get the following corollary of Lemma 19.

COROLLARY 55. *Let* $\mathsf{CS}$ *be a schematic constant specification. We have for any* $\mathcal{L}_{\mathsf{J}}$*-formula* $A$ *and any substitution* $\sigma$

$$\mathsf{LP}_{\mathsf{CS}} \vdash A \quad \textit{implies} \quad \mathsf{LP}_{\mathsf{CS}} \vdash A\sigma \ .$$

LEMMA 56. *Let* $\mathsf{CS}$ *be an axiomatically appropriate and schematic constant specification. For every subformula* $B$ *of* $A$ *and for all*

$$B_1, \dots, B_n \in |B| \ ,$$

*there is a formula* $B' \in \|B\|$ *and a substitution* $\sigma$ *that lives on* $B'$ *and meets the no new variable condition such that*

1. *if* $B$ *is a positive subformula occurrence of* $A$*, then*
$$\mathsf{LP}_{\mathsf{CS}} \vdash (B_1 \vee \cdots \vee B_n)\sigma \to B' \ ;$$
2. *if* $B$ *is a negative subformula occurrence of* $A$*, then*
$$\mathsf{LP}_{\mathsf{CS}} \vdash B' \to (B_1 \wedge \cdots \wedge B_n)\sigma \ .$$

PROOF. The two claims are shown simultaneously by induction on the structure of $B$.

1. $B = P$ where $P \in \mathsf{Prop}$. In this case both statement are trivial since $|B| = \|B\| = \{P\}$ and thus we can use the empty substitution (i.e. the identity function).
2. $B = \neg C$ occurring positively. Let $\neg C_1, \dots, \neg C_n \in |\neg C|$. Then $C$ occurs negatively and $C_1, \dots, C_n \in |C|$. By I.H. there is $C' \in \|C\|$ and a substitution $\sigma$ with the required properties such that
$$\mathsf{LP}_{\mathsf{CS}} \vdash C' \to (C_1 \wedge \cdots \wedge C_n)\sigma \ .$$
Using the contrapositive, we get
$$\mathsf{LP}_{\mathsf{CS}} \vdash (\neg C_1 \vee \cdots \vee \neg C_n)\sigma \to \neg C' \ .$$
Since $\neg C' \in \|\neg C\|$ and $\sigma$ satisfies the required properties, we are done.
3. $B = \neg C$ occurring negatively. This case is symmetric to the previous case and, therefore, omitted.
4. $B = D \to C$ occurring positively. Let
$$D_1 \to C_1, \dots, D_n \to C_n \in |D \to C| \ .$$
We find that $C$ occurs positively, $D$ occurs negatively, and
$$C_1, \dots, C_n \in |C| \quad \text{as well as} \quad D_1, \dots, D_n \in |D| \ .$$
Hence by I.H. there are $C' \in \|C\|$ and $\sigma_C$ as well as $D' \in \|D\|$ and $\sigma_D$ such that
$$\mathsf{LP}_{\mathsf{CS}} \vdash (C_1 \vee \cdots \vee C_n)\sigma_C \to C'$$

and

$$\mathsf{LP}_{\mathsf{CS}} \vdash D' \to (D_1 \wedge \cdots \wedge D_n)\sigma_D \ .$$

Since $\mathsf{CS}$ is schematic, we find by Corollary 55

$$\mathsf{LP}_{\mathsf{CS}} \vdash ((C_1 \vee \cdots \vee C_n)\sigma_C \to C')\sigma_D$$

and

$$\mathsf{LP}_{\mathsf{CS}} \vdash (D' \to (D_1 \wedge \cdots \wedge D_n)\sigma_D)\sigma_C \ .$$

Because $\sigma_C$ lives on $C'$, $\sigma_D$ lives on $D'$, $C$ and $D$ are different subformula occurrences of $A$, and $\mathcal{A}$ is a proper annotation, we find

$$\mathsf{dom}(\sigma_C) \cap \mathsf{dom}(\sigma_D) = \emptyset \ .$$

Hence we can define $\sigma = \sigma_C \cup \sigma_D$. We have (see Remark 53)

$$\mathsf{LP}_{\mathsf{CS}} \vdash (C_1 \vee \cdots \vee C_n)\sigma \to C'\sigma_D$$

and

$$\mathsf{LP}_{\mathsf{CS}} \vdash D'\sigma_C \to (D_1 \wedge \cdots \wedge D_n)\sigma \ .$$

Using propositional reasoning we obtain

$$\mathsf{LP}_{\mathsf{CS}} \vdash ((D_1 \to C_1) \vee \cdots \vee (D_n \to C_n))\sigma \to (D'\sigma_C \to C'\sigma_D) \ .$$

It is easy to see that $D'\sigma_C \to C'\sigma_D \in \|D \to C\|$ and that $\sigma$ satisfies the required properties. Thus we are done with this case.

5. $B = D \to C$ occurring negatively. Again, this case is symmetric to the previous case.
6. $B = \Box C$ occurring positively. Let

$$t_1{:}C_1, \ldots, t_n{:}C_n \in |\Box C| \ .$$

Each $C_i$ is a disjunction of elements $C_{i,1}, \ldots, C_{i,n_i} \in |C|$. We apply I.H. to the collection of all these $C_{i,j}$ and obtain a formula $C' \in \|C\|$ and a substitution $\sigma$ such that

$$\mathsf{LP}_{\mathsf{CS}} \vdash (C_1 \vee \cdots \vee C_n)\sigma \to C' \ .$$

Thus we immediately get for each $1 \le i \le n$

$$\mathsf{LP}_{\mathsf{CS}} \vdash C_i\sigma \to C' \ .$$

By Constructive Necessitation there are ground terms $u_i$ such that for each $1 \le i \le n$

$$\mathsf{LP}_{\mathsf{CS}} \vdash u_i{:}(C_i\sigma \to C') \ .$$

Using the application axiom ($\mathsf{J}$) we find for each $1 \le i \le n$

$$\mathsf{LP}_{\mathsf{CS}} \vdash (t_i{:}C_i)\sigma \to (u_i \cdot t_i\sigma){:}C' \ .$$

We set $s := (u_1 \cdot t_1\sigma) + \cdots + (u_n \cdot t_n\sigma)$ and obtain by the axiom (+) for each $1 \leq i \leq n$

$$\mathsf{LP}_{\mathsf{CS}} \vdash (t_i{:}C_i)\sigma \to s{:}C'$$

and thus

$$\mathsf{LP}_{\mathsf{CS}} \vdash \big((t_1{:}C_1)\sigma \vee \cdots \vee (t_n{:}C_n)\sigma\big) \to s{:}C' \enspace ,$$

which is

$$\mathsf{LP}_{\mathsf{CS}} \vdash \big((t_1{:}C_1) \vee \cdots \vee (t_n{:}C_n)\big)\sigma \to s{:}C' \enspace .$$

Obviously we have $s{:}C' \in \|\Box C\|$ and the substitution $\sigma$ satisfies the required properties by I.H.

7. $B = \Box C$ occurring negatively. Let

$$x{:}C_1, \ldots, x{:}C_n \in |\Box C|$$

where $\mathcal{A}(\Box C) = x$. By I.H. there is a $C' \in \|C\|$ and a substitution $\sigma$ such that

$$\mathsf{LP}_{\mathsf{CS}} \vdash C' \to (C_1 \wedge \cdots \wedge C_n)\sigma \enspace .$$

Because $\sigma$ lives on $C'$, $C$ is a subformula of $A$, and $\mathcal{A}$ is a proper annotation, we have $x \notin \mathsf{dom}(\sigma)$. Furthermore we have for each $1 \leq i \leq n$

$$\mathsf{LP}_{\mathsf{CS}} \vdash C' \to C_i\sigma \enspace .$$

Thus by Constructive Necessitation there are ground terms $t_i$ such that for each $1 \leq i \leq n$

$$\mathsf{LP}_{\mathsf{CS}} \vdash t_i{:}(C' \to C_i\sigma) \enspace .$$

We set $s := t_1 + \cdots t_n$ and obtain by the axiom (+) for each $1 \leq i \leq n$

$$\mathsf{LP}_{\mathsf{CS}} \vdash s{:}(C' \to C_i\sigma) \enspace .$$

Using the application axiom (J) we find for each $1 \leq i \leq n$

$$\mathsf{LP}_{\mathsf{CS}} \vdash x{:}C' \to (s \cdot x){:}(C_i\sigma) \enspace .$$

We let $\sigma' := \sigma \cup \{x/(s \cdot x)\}$ (remember that $x \notin \mathsf{dom}(\sigma)$) and obtain for each $1 \leq i \leq n$

$$\mathsf{LP}_{\mathsf{CS}} \vdash x{:}C' \to (x{:}C_i)\sigma' \enspace .$$

Hence

$$\mathsf{LP}_{\mathsf{CS}} \vdash x{:}C' \to (x{:}C_1 \wedge \cdots \wedge x{:}C_n)\sigma' \enspace .$$

It is easy to see that $\sigma'$ lives on $x{:}C'$ and satisfies the no new variable condition, which establishes the result in this case.

$\dashv$

THEOREM 57 (Semantic Realization). *Let* $\mathsf{CS}$ *be an axiomatically appropriate and schematic constant specification. If* $\mathsf{S4} \vdash A$*, then there exists* $A' \in \|A\|$ *with* $\mathsf{LP}_{\mathsf{CS}} \vdash A'$*.*

PROOF. Assume $\mathsf{S4} \vdash A$. By Lemma 50 there are $A_1, \dots, A_n \in |A|$ such that

$$\mathsf{LP}_{\mathsf{CS}} \vdash A_1 \vee \cdots \vee A_n \ . \tag{17}$$

By Lemma 56 there is a $A' \in \|A\|$ and a substitution $\sigma$ such that

$$\mathsf{LP}_{\mathsf{CS}} \vdash (A_1 \vee \cdots \vee A_n)\sigma \to A' \ . \tag{18}$$

Since $\mathsf{CS}$ is schematic, we obtain from (17) and Corollary 55

$$\mathsf{LP}_{\mathsf{CS}} \vdash (A_1 \vee \cdots \vee A_n)\sigma \ ,$$

which toghether with (18) yields $\mathsf{LP}_{\mathsf{CS}} \vdash A'$. ⊣

CHAPTER 4

# SELF-REFERENTIALITY

## 4A. Generated Models

Definition 58 (Evidence closure). Let $\mathcal{B} \subseteq \mathsf{Tm} \times \mathcal{L}_{\mathsf{J}}$. For a set $X \subseteq \mathsf{Tm} \times \mathcal{L}_{\mathsf{J}}$ we define $\mathsf{cl}_{\mathcal{B}}(X)$ by:

1. if $(t, A) \in \mathcal{B}$, then $(t, A) \in \mathsf{cl}_{\mathcal{B}}(X)$;
2. if $(s, A) \in X$ or $(t, A) \in X$, then $(s + t, A) \in \mathsf{cl}_{\mathcal{B}}(X)$;
3. if $(s, A) \in X$ and $(t, A \to B) \in X$, then $(t \cdot s, B) \in \mathsf{cl}_{\mathcal{B}}(X)$;
4. if $(t, A) \in X$, then $(!t, t{:}A) \in \mathsf{cl}_{\mathcal{B}}(X)$.

Note that $\mathsf{cl}_{\mathcal{B}}$ is a monotone operator on $\mathsf{Tm} \times \mathcal{L}_{\mathsf{J}}$, that is

$$X \subseteq Y \quad \text{implies} \quad \mathsf{cl}_{\mathcal{B}}(X) \subseteq \mathsf{cl}_{\mathcal{B}}(Y) \tag{19}$$

for all $X, Y \subseteq \mathsf{Tm} \times \mathcal{L}_{\mathsf{J}}$. Hence, $\mathsf{cl}_{\mathcal{B}}$ has a least fixed point, which is shown as usual.

Lemma 59 (Least fixed point). *There is a unique* $R \subseteq \mathsf{Tm} \times \mathcal{L}_{\mathsf{J}}$ *such that*

1. $\mathsf{cl}_{\mathcal{B}}(R) = R$,
2. *for any* $S \subseteq \mathsf{Tm} \times \mathcal{L}_{\mathsf{J}}$, *if* $\mathsf{cl}_{\mathcal{B}}(S) \subseteq S$, *then* $R \subseteq S$.

Proof. Let $C := \{S \subseteq \mathsf{Tm} \times \mathcal{L}_{\mathsf{J}} \mid \mathsf{cl}_{\mathcal{B}}(S) \subseteq S\}$. Since $\mathsf{Tm} \times \mathcal{L}_{\mathsf{J}} \in C$, we know that $C$ is non-empty. Let $R := \bigcap C$. The second claim now holds by definition. And the uniqueness of $R$ is an easy corollary of the second claim.

It remains to establish $\mathsf{cl}_{\mathcal{B}}(R) = R$ . Let $S \in C$. Since $R \subseteq S$ and $\mathsf{cl}_{\mathcal{B}}$ is monotone, we find $\mathsf{cl}_{\mathcal{B}}(R) \subseteq \mathsf{cl}_{\mathcal{B}}(S)$. We also have $\mathsf{cl}_{\mathcal{B}}(S) \subseteq S$, so $\mathsf{cl}_{\mathcal{B}}(R) \subseteq S$. Since $S$ is an arbitrary element of $C$ and $R = \bigcap C$, this implies $\mathsf{cl}_{\mathcal{B}}(R) \subseteq R$.

To show $R \subseteq \mathsf{cl}_{\mathcal{B}}(R)$, we first observe that since $\mathsf{cl}_{\mathcal{B}}(R) \subseteq R$, we have $\mathsf{cl}_{\mathcal{B}}(\mathsf{cl}_{\mathcal{B}}(R)) \subseteq \mathsf{cl}_{\mathcal{B}}(R)$ by monotonicity. Thus $\mathsf{cl}_{\mathcal{B}}(R) \in C$, which gives us $R \subseteq \mathsf{cl}_{\mathcal{B}}(R)$ because $R = \bigcap C$. ⊣

Definition 60 (Evidence relation). Let $\mathcal{B} \subseteq \mathsf{Tm} \times \mathcal{L}_{\mathsf{J}}$. We define *the minimal evidence relation* $\mathcal{E}(\mathcal{B})$ *over* $\mathcal{B}$ to be the least fixed point of $\mathsf{cl}_{\mathcal{B}}$.

We immediately get the following properties of $\mathcal{E}(\mathcal{B})$.

Lemma 61. *Let $\mathcal{B} \subseteq \mathsf{Tm} \times \mathcal{L}_\mathsf{J}$ and assume $(t, A) \in \mathcal{E}(\mathcal{B})$. Then we have $(t, A) \in \mathcal{B}$ or one of the following cases*

1. *$t = r + s$ with $(r, A) \in \mathcal{E}(\mathcal{B})$ or $(s, A) \in \mathcal{E}(\mathcal{B})$;*
2. *$t = s \cdot r$ and there exists a formula $B$ with*
$$(r, B) \in \mathcal{E}(\mathcal{B}) \text{ and } (s, B \to A) \in \mathcal{E}(\mathcal{B}) \ ;$$
3. *$t = !s$ and there exists a formula $B$ with $A = s{:}B$ and $(s, B) \in \mathcal{E}(\mathcal{B})$.*

Further, we get the following lemma.

Lemma 62 (Monotonicity of $\mathcal{E}$). *Let $\mathcal{B}, \mathcal{C} \subseteq \mathsf{Tm} \times \mathcal{L}_\mathsf{J}$. We have that*
$$\mathcal{E}(\mathcal{B}) \subseteq \mathcal{E}(\mathcal{B} \cup \mathcal{C}) \ .$$

Proof. By induction on $t$ we show $(t, A) \in \mathcal{E}(\mathcal{B})$ implies $(t, A) \in \mathcal{E}(\mathcal{B} \cup \mathcal{C})$ for all formulas $A$. Assume $(t, A) \in \mathcal{E}(\mathcal{B})$.

If $(t, A) \in \mathcal{B}$, then $(t, A) \in \mathcal{B} \cup \mathcal{C}$, and it follows that $(t, A) \in \mathcal{E}(\mathcal{B} \cup \mathcal{C})$. If $(t, A) \notin \mathcal{B}$, then we have one of the following cases.

1. $t = r + s$. Then by Lemma 61 we have $(r, A) \in \mathcal{E}(\mathcal{B})$ or $(s, A) \in \mathcal{E}(\mathcal{B})$. By I.H. we find $(r, A) \in \mathcal{E}(\mathcal{B} \cup \mathcal{C})$ or $(s, A) \in \mathcal{E}(\mathcal{B} \cup \mathcal{C})$ and thus we get $(t, A) = (r + s, A) \in \mathcal{E}(\mathcal{B} \cup \mathcal{C})$.
2. $t = r \cdot s$. Then by Lemma 61 $\{(s, B), (r, B \to A)\} \subseteq \mathcal{E}(\mathcal{B})$ for some formula $B$. By I.H. we find $\{(s, B), (r, B \to A)\} \subseteq \mathcal{E}(\mathcal{B} \cup \mathcal{C})$. We get $(t, A) = (r \cdot s, A) \in \mathcal{E}(\mathcal{B} \cup \mathcal{C})$.
3. $t = !s$. Then $A = s{:}B$ and $(s, B) \in \mathcal{E}(\mathcal{B})$ by Lemma 61. Then by I.H. we find $(s, B) \in \mathcal{E}(\mathcal{B} \cup \mathcal{C})$ and thus $(t, A) = (!s, s{:}B) \in \mathcal{E}(\mathcal{B} \cup \mathcal{C})$.

⊣

Definition 63 (Generated Model). A *generated model* is a pair $\mathcal{M} = (\mathsf{val}, \mathcal{B})$ where $\mathsf{val} \subseteq \mathsf{Prop}$ and $\mathcal{B} \subseteq \mathsf{Tm} \times \mathcal{L}_\mathsf{J}$. For a constant specification $\mathsf{CS}$, the generated model $\mathcal{M}$ is called a *generated* $\mathsf{CS}$*-model* if $\mathsf{CS} \subseteq \mathcal{B}$.

Definition 64 (Truth). Let $\mathcal{M} = (\mathsf{val}, \mathcal{B})$ be a generated model and $D$ be a formula. We define the relation $\mathcal{M} \Vdash D$ by

1. $\mathcal{M} \Vdash P$ iff $P \in \mathsf{val}$
2. $\mathcal{M} \Vdash \neg A$ iff $\mathcal{M} \nVdash A$
3. $\mathcal{M} \Vdash A \to B$ iff $\mathcal{M} \nVdash A$ or $\mathcal{M} \Vdash B$
4. $\mathcal{M} \Vdash t{:}A$ iff $(t, A) \in \mathcal{E}(\mathcal{B})$ and $\mathcal{M} \Vdash A$.

A formula $D$ is *valid with respect to generated* $\mathsf{CS}$*-models* if $\mathcal{M} \Vdash D$ for all generated $\mathsf{CS}$-models $\mathcal{M}$.

Remark 65. The above truth definition of $t{:}A$ for generated models is different from that for basic modular models. Basic modular models provide a clear ontological separation of justifications and truth. In generated models, truth and justifications are intertwined in that $t$ is only evidence for $F$ if $F$ is true. The philosophical objections to such a paradigm also have practical roots. In court, evidence is used to determine the truth of the matter. However, if the acceptability of the evidence were to depend on this truth, it would create a vicious circle. However, although ontologically less transparent, generated models provide a very efficient means to establish many important properties of justification logic.

Theorem 66 (Soundness). *For all formulas $D$,*

$$\mathsf{LP}_{\mathsf{CS}} \vdash D \quad \textit{implies} \quad D \textit{ is valid with respect to generated } \mathsf{CS}\textit{-models.}$$

Proof. As usual the proof is by induction on the length of the derivation of $D$. Let $\mathcal{M} = (\mathsf{val}, \mathcal{B})$ be a generated $\mathsf{CS}$-model.

1. $D$ is an instance of $(\mathsf{P})$. All instances of propositional tautologies hold under $\mathcal{M}$.
2. $D$ is an instance of $(\mathsf{J})$, $(+)$ or $(\mathsf{j4})$. In these cases, $\mathcal{M} \Vdash D$ follows immediately from the fact that $\mathcal{E}(\mathcal{B})$ is closed under $\mathsf{cl}_{\mathcal{B}}$.
3. $D = t{:}A \to A$ is an instance of axiom $(\mathsf{jt})$. Suppose $\mathcal{M} \Vdash t{:}A$. By Definition 64 this implies $\mathcal{M} \Vdash A$ and we are done.
4. $D$ is the conclusion of an instance of $(\mathsf{MP})$. It is trivial to see that modus ponens preserves truth in a model.
5. $D = c{:}A$ is the conclusion of an instance of (AN). Hence $(c, A) \in \mathsf{CS}$. Since $\mathcal{M} = (\mathsf{val}, \mathcal{B})$ is a generated $\mathsf{CS}$ model, we find $(c, A) \in \mathcal{B}$. Since $\mathcal{E}(\mathcal{B})$ is closed under $\mathsf{cl}_{\mathcal{B}}$, this implies $(c, A) \in \mathcal{E}(\mathcal{B})$, which finally yields $\mathcal{M} \Vdash c{:}A$.

⊣

To establish completeness we again employ a maximal consistent set construction.

Definition 67 (Induced model). Let $\Phi$ be a maximal $\mathsf{LP}_{\mathsf{CS}}$-consistent set of formulas. The generated model $\mathcal{M}_\Phi = (\mathsf{val}_\Phi, \mathcal{B}_\Phi)$ that is *induced by* $\Phi$ is given by

1. $P \in \mathsf{val}_\Phi$ iff $P \in \Phi \cap \mathsf{Prop}$.
2. $(t, A) \in \mathcal{B}_\Phi$ iff $t{:}A \in \Phi$.

$\mathcal{M}_\Phi$ is a generated $\mathsf{CS}$-model. Indeed, suppose $(c, B) \in \mathsf{CS}$. Then we have $\mathsf{LP}_{\mathsf{CS}} \vdash c{:}B$, which by maximal consistency of $\Phi$ implies $c{:}B \in \Phi$. Since $c$ is atomic, we conclude $(c, B) \in \mathcal{B}_\Phi$.

Lemma 68 (Canonical evidence). *Let $\Phi$ be a maximal $\mathsf{LP}_{\mathsf{CS}}$-consistent set. Then*

$$t{:}A \in \Phi \quad \Longleftrightarrow \quad (t, A) \in \mathcal{E}(\mathcal{B}_\Phi) \ .$$

Proof. For the direction from left to right, suppose that $t{:}A \in \Phi$. Then $(t, A) \in \mathcal{B}_\Phi$, which gives us $(t, A) \in \mathcal{E}(\mathcal{B}_\Phi)$ since $\mathcal{E}(\mathcal{B}_\Phi)$ is closed under $\mathsf{cl}_{\mathcal{B}_\Phi}$.

The direction from right to left is shown by induction on $t$. Suppose $(t, A) \in \mathcal{E}(\mathcal{B}_\Phi)$. We distinguish the following cases for $t$.

1. $t$ atomic. By Lemma 61 we find $(t, A) \in \mathcal{B}_\Phi$ and thus $t{:}A \in \Phi$.
2. $t = r + s$. If $(t, A) \in \mathcal{B}_\Phi$ we are done as in the case before. Otherwise we find by Lemma 61 that $(r, A) \in \mathcal{E}(\mathcal{B}_\Phi)$ or $(a, A) \in \mathcal{E}(\mathcal{B}_\Phi)$. By the induction hypothesis we obtain $r{:}A \in \Phi$ or $s{:}A \in \Phi$. By axiom (+) and maximal consistency of $\Phi$ we conclude $t{:}A \in \Phi$.
3. $t = r \cdot s$. If $(t, A) \in \mathcal{B}_\Phi$ we are done as in the first case. Otherwise we find by Lemma 61 that there exists $B$ with $(s, B) \in \mathcal{E}(\mathcal{B}_\Phi)$ and $(r, B \to A) \in \mathcal{E}(\mathcal{B}_\Phi)$. By the induction hypothesis we obtain $s{:}B \in \Phi$ and $r{:}(B \to A) \in \Phi$. By axiom ($\mathsf{J}$) and maximal consistency of $\Phi$ we conclude $t{:}A \in \Phi$.
4. $t = !s$. If $(t, A) \in \mathcal{B}_\Phi$ we are done as in the first case. Otherwise we find by Lemma 61 that there exists $B$ with $A = s{:}B$ and $(s, B) \in \mathcal{E}(\mathcal{B}_\Phi)$. By the induction hypothesis we obtain $s{:}B \in \Phi$. By axiom ($\mathsf{j4}$) and maximal consistency of $\Phi$ we conclude $t{:}A \in \Phi$.

⊣

Lemma 69 (Truth lemma). *Let $\Phi$ be a maximal $\mathsf{LP}_{\mathsf{CS}}$-consistent set of formulas. Then*

$$A \in \Phi \iff \mathcal{M}_\Phi \Vdash A \ .$$

Proof. By induction on the structure of $A$.

1. $A \in \mathsf{Prop}$. We have $A \in \Phi$ iff (by definition) $A \in \mathsf{val}_\Phi$ iff (by definition) $\mathcal{M}_\Phi \Vdash A$.
2. $A = \neg B$. We have $\neg B \in \Phi$ iff (by maximal consistency of $\Phi$) $B \notin \Phi$ iff (by IH) $\mathcal{M}_\Phi \nVdash B$ iff $\mathcal{M}_\Phi \Vdash \neg B$.
3. $A = B \to C$. We have $B \to C \in \Phi$ iff (by maximal consistency of $\Phi$) $B \notin \Phi$ or $C \in \Phi$ iff (by IH) $\mathcal{M}_\Phi \nVdash B$ or $\mathcal{M}_\Phi \Vdash C$ iff $\mathcal{M}_\Phi \Vdash B \to C$.
4. $A = t{:}B$. We have $t{:}B \in \Phi$ iff (by Lemma 68) $(t, B) \in \mathcal{E}(\mathcal{B}_\Phi)$ iff (by definition) $\mathcal{M}_\Phi \Vdash t{:}B$.
5. $A = t{:}B$. We have that $t{:}B \in \Phi$ implies

$$(t, B) \in \mathcal{E}(\mathcal{B}_\Phi) \quad \text{and} \quad B \in \Phi \tag{20}$$

by Lemma 68 and by axiom ($jt$) together with maximal consistency of $\Phi$, respectively. By definition, (20) is eqivalent to $\mathcal{M}_\Phi \Vdash t{:}B$. The direction from right to left follows immediately from Lemma 68.

⊣

As usual, we now obtain completeness of $\mathsf{LP}_{\mathsf{CS}}$ with respect to generated $\mathsf{CS}$-models.

Theorem 70 (Completeness). *For all formulas $D$,*

$$D \text{ is valid with respect to generated } \mathsf{CS}\text{-models} \quad \Longrightarrow \quad \mathsf{LP}_{\mathsf{CS}} \vdash D \ .$$

## 4B. Self-Referential Knowledge

A justification logics exhibits self-referentiality when a term $t$ proves something about itself

$$\mathsf{LP}_{\mathsf{CS}} \vdash t{:}F(t) \ .$$

Constructions of this kind are, of course, perfectly legal in the language of justification logic. In fact, when the constant specification $\mathsf{CS}$ is schematic, there must be self-referential theorems. Assume, for example, that a constant $c$ justifies all instances of axiom ($\mathsf{jt}$), that is all instances of $t{:}F \to F$. Then, in particular, we have

$$\mathsf{LP}_{\mathsf{CS}} \vdash c{:}(c{:}F \to F) \ ,$$

which is a self-referential theorem.

In this section, we show that self-referentiality is inherent in the notion of knowledge axiomatized by $\mathsf{S4}$ and $\mathsf{LP}_{\mathsf{CS}}$. That means we necessarily need a self-referential constant specification $\mathsf{CS}$ in order to realize $\mathsf{S4}$ in $\mathsf{LP}_{\mathsf{CS}}$.

Definition 71 (Self-referential $\mathsf{CS}$). A constant specification $\mathsf{CS}$ is called *directly self-referential* if $(c, A) \in \mathsf{CS}$ for some axiom $A$ that contains at least one occurrence of the constant $c$.

In the following we show that $\mathsf{S4}$ and $\mathsf{LP}_{\mathsf{CS}}$ describe directly self-referential knowledge. That means if $\mathsf{LP}_{\mathsf{CS}}$ realizes $\mathsf{S4}$ for some constant specification $\mathsf{CS}$, then that constant specification must be directly self-referential.

Consider the following formula of $\mathcal{L}$

$$G := \neg\Box((P \to \Box P) \to \bot)$$

where $P$ is a fixed atomic proposition. Clearly, $G$ is valid and thus by completeness $\mathsf{S4} \vdash G$. Let $\mathsf{CS}$ be the maximal constant specification that is not directly self-referential, that is

$$\begin{aligned}\mathsf{CS} := \{(c, A) \mid \ & c \text{ is a constant and } A \text{ is an axiom of } \mathsf{LP} \text{ such that} \\ & c \text{ does not occur in } A\} \ .\end{aligned}$$

We show that for any realization $F$ of $G$, that is for any $\mathcal{L}_{\mathsf{J}}$ formula $F$ with $F^\circ = G$, there is a generated $\mathsf{CS}$-model $\mathcal{M}$ with $\mathcal{M} \not\Vdash F$. Hence by soundness $\mathsf{LP}_{\mathsf{CS}} \nvdash F$.

Let $t$ and $t'$ be arbitrary terms and set $F := \neg t{:}((P \to t'{:}P) \to \bot)$. We set

$$\mathcal{B}_0 := \mathsf{CS}$$
$$\mathcal{B} := \mathsf{CS} \cup \{(t, (P \to t'{:}P) \to \bot)\} \ .$$

LEMMA 72. *For any subterm $s$ of $t'$ we have:*

1. *If $(s, H) \in \mathcal{E}(\mathcal{B}_0)$, then $\mathsf{LP}_{\mathsf{CS}} \vdash H$ and $H$ does not contain occurrences of $t'$.*
2. *If $(s, H) \notin \mathcal{E}(\mathcal{B}_0)$ but $(s, H) \in \mathcal{E}(\mathcal{B})$, then $H$ has at least one occurrence of $t'$. Moreover, if $H$ is an implication, then*

$$H = (P \to t'{:}P) \to \bot \ .$$

PROOF. By induction on the structure of $s$. We distinguis the following cases.

$s = x$ is a variable.

1. By Lemma 61 we find $(x, H) \in \mathcal{B}_0 = \mathsf{CS}$. Thus Clause 1 is vacuously true.
2. By Lemma 61 we find $(x, H) \in \mathcal{B}$. Thus

$$t = x \text{ and } H = (P \to t'{:}P) \to \bot \ .$$

The latter does contain $t'$ and is the only allowed implication.

$s = c$ is a constant.

1. By Lemma 61 we find $(c, H) \in \mathcal{B}_0 = \mathsf{CS}$. Thus $H$ must be an axiom of $\mathsf{LP}$ and we have $\mathsf{LP}_{\mathsf{CS}} \vdash H$. Since $\mathsf{CS}$ is not directly self-referential, $H$ cannot contain occurrences of $c$, which is a subterm of $t'$. Thus $H$ cannot contain $t'$ either.
2. By Lemma 61 we find $(c, H) \in \mathcal{B}$. Since $(s, H) \notin \mathcal{E}(\mathcal{B}_0) \supseteq \mathsf{CS}$, we must have

$$t = c \text{ and } H = (P \to t'{:}P) \to \bot \ .$$

The latter does contain $t'$ and is the only allowed implication.

$s = s_1 + s_2$.

1. By Lemma 61 we find $(s_i, H) \in \mathcal{E}(\mathcal{B}_0)$ for $i = 1$ or $i = 2$. Thus by I.H. we have $\mathsf{LP}_{\mathsf{CS}} \vdash H$ and $H$ does not contain occurrences of $t'$.
2. By Lemma 61 there are two cases:
   (a) $t = s_1 + s_2$ and $H = (P \to t'{:}P) \to \bot$. The latter does contain $t'$ and is the only allowed implication.
   (b) $(s_i, H) \notin \mathcal{E}(\mathcal{B}_0)$ but $(s_i, H) \in \mathcal{E}(\mathcal{B})$ for $i = 1$ or $i = 2$. By I.H. the formula $H$ contains $t'$ and, if an impliciation, is $(P \to t'{:}P) \to \bot$.

$s = s_1 \cdot s_2$.

1. By Lemma 61 , there must exists a formula $C$ such that

$$(s_1, C \to H) \in \mathcal{E}(\mathcal{B}_0) \text{ and } (s_2, C) \in \mathcal{E}(\mathcal{B}_0) \ .$$

By I.H. both $\mathsf{LP}_{\mathsf{CS}} \vdash C \to H$ and $\mathsf{LP}_{\mathsf{CS}} \vdash C$. Hence also $\mathsf{LP}_{\mathsf{CS}} \vdash H$. By I.H. we know that $C \to H$ does not contain $t'$, hence neither does $H$.
2. By Lemma 61 there are three cases:
   (a) $t = s_1 \cdot s_2$ and $H = (P \to t'{:}P) \to \bot$. The latter does contain $t'$ and is the only allowed implication.
   (b) There is a formula $C$ with

$$\begin{aligned}(s_1, C \to H) &\in \mathcal{E}(\mathcal{B})\\ (s_2, C) &\in \mathcal{E}(\mathcal{B})\\ (s_1, C \to H) &\notin \mathcal{E}(\mathcal{B}_0) \ .\end{aligned}$$

   We show that these three statements are, in fact, inconsistent. By I.H. for $s_1$ we find that

$$C \to H \quad = \quad (P \to t'{:}P) \to \bot \ .$$

   So

$$C = P \to t'{:}P \ , \tag{21}$$

   which is an implication different from the one allowed in the conclusion of Clause 2. Now we have two cases, both giving a contradiction.
   (i) $(s_2, C) \notin \mathcal{E}(\mathcal{B}_0)$. We apply I.H., Clause 2 and obtain that $C$ must have a form different from (21). Contradiction.
   (ii) $(s_2, C) \in \mathcal{E}(\mathcal{B}_0)$. We apply I.H., Clause 1 and obtain that $C$ does not contain occurrences of $t'$. Contradiction.
   (c) There is a formula $C$ with

$$\begin{aligned}(s_1, C \to H) &\in \mathcal{E}(\mathcal{B})\\ (s_2, C) &\in \mathcal{E}(\mathcal{B})\\ (s_2, C) &\notin \mathcal{E}(\mathcal{B}_0) \ .\end{aligned}$$

   We show that these three statements are also inconsistent. By I.H. for $s_2$ we find that $C$ contains an occurrence of $t'$. Thus the formula $C \to H$ also contains an occurrence of $t'$. Now we have two cases, both giving a contradiction.
   (i) $(s_1, C \to H) \in \mathcal{E}(\mathcal{B}_0)$. We apply I.H., Clause 1 to $s_1$ and obtain that the formula $C \to H$ does not contain $t'$. Contradiction.
   (ii) $(s_1, C \to H) \notin \mathcal{E}(\mathcal{B}_0)$. This is the same situation as in Case 2b.

$s = !s_1$

1. By Lemma 61, $H = s_1{:}C$ for some formula $C$ and $(s_1, C) \in \mathcal{E}(\mathcal{B}_0)$. By I.H. $C$ does not contain $t'$. Since $\mathcal{B}_0 = \mathsf{CS}$ we know by Lemma 62 that $(s_1, C) \in \mathcal{E}(\mathcal{C})$ for any $\mathcal{C} \supseteq \mathsf{CS}$. Therefore, $\mathcal{M} \Vdash s_1{:}C$ for a generated $\mathsf{CS}$ model $\mathcal{M}$. By completeness for generated $\mathsf{CS}$ models (Theorem 70), we obtain $\mathsf{LP}_{\mathsf{CS}} \vdash s_1{:}C$, that is $\mathsf{LP}_{\mathsf{CS}} \vdash H$.
2. By Lemma 61 there are two cases:
   (a) $t = !s_1$ and $H = (P \to t'{:}P) \to \bot$. The latter does contain $t'$ and is the only allowed implication.
   (b) $H = s_1{:}C$ for some formula $C$ with
   $$(s_1, C) \in \mathcal{E}(\mathcal{B})$$
   $$(s_1, C) \notin \mathcal{E}(\mathcal{B}_0)\ .$$
   By I.H. the formula $C$ contains $t'$, thus so does $H = s_1{:}C$. Since $H$ is not an implication, there is no additional claim to show and we are done.

⊣

Theorem 73. *For* $\mathsf{CS}$ *and* $F$ *as given above, we have* $\mathsf{LP}_{\mathsf{CS}} \nvdash F$.

Proof. Consider $\mathcal{B}_0$ and $\mathcal{B}$ as given above. By Lemma 72, Clause 1 we find $(t', P) \notin \mathcal{E}(\mathcal{B}_0)$ for otherwise we would get $\mathsf{LP}_{\mathsf{CS}} \vdash P$, which of course does not hold. Then we also find

$$(t', P) \notin \mathcal{E}(\mathcal{B}) \tag{22}$$

for otherwise we would get by Lemma 72, Clause 2 that $t'$ occurs in $P$, which obviously is not the case.

Consider the generated $\mathsf{CS}$ model $\mathcal{M} = (\mathsf{val}, \mathcal{B})$ where $\mathsf{val} = \mathsf{Prop}$. By (22) we know $\mathcal{M} \nVdash t'{:}P$. Since $P \in \mathsf{val}$ we thus obtain $\mathcal{M} \Vdash (P \to t'{:}P) \to \bot$. By $(t, (P \to t'{:}P) \to \bot) \in \mathcal{B}$ we get $\mathcal{M} \Vdash t{:}((P \to t'{:}P) \to \bot)$, which is $\mathcal{M} \nVdash \neg t{:}((P \to t'{:}P) \to \bot)$. Now $\mathsf{LP}_{\mathsf{CS}} \nvdash F$ follows by soundness of $\mathsf{LP}_{\mathsf{CS}}$. ⊣

## 4C. Notes

This chapter is based on work by Brezhnev and Kuznets [3, 6, 7].

CHAPTER 5

# DECIDABILITY

## 5A. Post's theorem

In modal logic, decidability often is a consequence of the finite model property. As we have seen earlier, $\mathsf{LP}_{\mathsf{CS}}$ is sound and complete even with respect to single world models (a basic modular model is nothing else than a single world modular model). Unfortunately this does not settle the question of decidability for the Logic of Proofs at all. The main problem is that the evidence relation is necessarily not a finite object.

We need a detailed analysis of how decidability proofs that rely on the finite model property really work. This leads us to the definition of *finitary* models, which allows us to establish decidability.

Definition 74. We say a logic $\mathsf{L}$ is decidable, if the set $\{F \mid \mathsf{L} \vdash F\}$ is decidable.

Theorem 75 (Post's Theorem). *If both a set and its complement are recursively enumerable, then the set is decidable.*

Lemma 76. *Let a finitely axiomatizable logic* $\mathsf{L}$ *be sound and complete with respect to a class of models* $\mathbb{C}$*, such that*

1. *the class* $\mathbb{C}$ *is recursively enumerable, and*
2. *the binary relation* $\mathcal{M} \Vdash F$ *between formulae and models from* $\mathbb{C}$ *is decidable.*

*Then* $\mathsf{L}$ *is decidable.*

Proof. Of course, the set of theorems of a finitely axiomatizable logic is recursively enumerable.

Here is an algorithm to recursively enumerate its complement. Since both the set of formulas and the models $\mathcal{M}$ of $\mathbb{C}$ are recursively enumerable, there is an enumeration of all pairs $(\mathcal{M}, F)$. For each pair in this enumeration, the algorithm checks whether $\mathcal{M} \Vdash \neg F$. If it is, the algorithm outputs $F$, otherwise it skips to the next pair. In this way, the algorithm will enumerate all non-theorems of $\mathsf{L}$, so the complement of $\mathsf{L}$ is recursively enumerable, too.

By Post's Theorem, $\mathsf{L}$ is decidable. ⊣

## 5B. Finitary models

Now we introduce the class of *finitary models* that fulfills the conditions of Lemma 76.

DEFINITION 77 (Finitary model). Let $\mathsf{CS}$ be a decidable schematic constant specification. Let $\mathcal{C} \subseteq \mathsf{Tm} \times \mathcal{L}_\mathsf{J}$ be finite and set $\mathcal{B} = \mathsf{CS} \cup \mathcal{C}$. Further let $\mathsf{val}$ be a finite valuation, that is a finite subset of $\mathsf{Prop}$. Then we call the generated $\mathsf{CS}$-model $\mathcal{M} = (\mathsf{val}, \mathcal{B})$ a *finitary* $\mathsf{CS}$*-model*.

COROLLARY 78. *The class of finitary* $\mathsf{CS}$*-models is recursively enumerable.*

Given a finitary model $\mathcal{M} = (\mathsf{val}, \mathcal{B})$ we want to decide whether $(t, F) \in \mathcal{E}(\mathcal{B})$ holds. However, this cannot be achieved directly since there are terms that evidence infinitely many formulas. In particular, if $(c, A) \in \mathsf{CS}$, then $c$ does not only evidence the formula $A$ but also all other (infinitely many) instances of the same axiom scheme. What we need is something like

for any given term $t$, the set $\{F \mid (t, F) \in \mathcal{E}(\mathcal{B})\}$ is finite.

In order to achieve this, we extend $\mathcal{L}_\mathsf{J}$ by schematic variables for formulas and schematic variables for terms, which results in a language called $\mathcal{L}_\mathsf{J}^\mathsf{s}$. We use $U, V, \dots$ to denote schemes of formulas, i.e. elements of $\mathcal{L}_\mathsf{J}^\mathsf{s}$, as opposed to $F, G, \dots$ that are reserved for formulas themselves, i.e. elements of $\mathcal{L}_\mathsf{J}$. The letters $r, s, t, \dots$ may denote both schemes of terms and terms themselves.

In this chapter a substitution is a mapping from schematic term variables to terms and from schematic formula variables to formulas. As before, we extend this notion of substitution to $\mathcal{L}_\mathsf{J}^\mathsf{s}$-formulas.

DEFINITION 79 (Unifier).

1. Given two $\mathcal{L}_\mathsf{J}^\mathsf{s}$-schemes $U$ and $V$, a *unifier* for $U$ and $V$ is a substitution $\sigma$ such that $U\sigma = V\sigma$.
2. A unifier $\mu$ is called *most general unifier* (mgu) if for each unifier $\sigma$, there is a substitution $\tau$ such that $\sigma = \mu\tau$.
3. An $\mathcal{L}_\mathsf{J}$-formula $A$ is called an *instance* of an $\mathcal{L}_\mathsf{J}^\mathsf{s}$-scheme $U$, if there is a substitution $\sigma$ with $A = U\sigma$.

In the following we will talk of *the* most general unifier. We can do so, because if there are several mgus, they are all equivalent up to renaming of schematic variables.

REMARK 80. Let $\mathsf{CS}$ be a schematic constant specification. It is easy to see that there is a $\mathsf{CS}^\mathsf{s} \subseteq \mathsf{Tm} \times \mathcal{L}_\mathsf{J}^\mathsf{s}$ such that

1. for a each constant $c$, the set $\{U \mid (c, U) \in \mathsf{CS}^{\mathsf{s}}\}$ is finite;
2. if $(c, A) \in \mathsf{CS}$, then there is a schema $U$ such that $(c, U) \in \mathsf{CS}^{\mathsf{s}}$ and $A$ is an instance of $U$;
3. if $(c, U) \in \mathsf{CS}^{\mathsf{s}}$, then $(c, A) \in \mathsf{CS}$ for all instances $A$ of $U$.

Definition 81. We assume that for each term $t$ we are given a finite set $\|t\|^0 \subseteq \mathcal{L}^{\mathsf{s}}_{\mathsf{J}}$. We then define sets $\|t\|^i \subseteq \mathcal{L}^{\mathsf{s}}_{\mathsf{J}}$ inductively by

$$\|t\|^{i+1} := \|t\|^i \cup X$$

where $X$ is given by

1. if $t = r + s$, then $X = \|r\|^i \cup \|s\|^i$;
2. if $t = r \cdot s$, then $X$ is the set of all $V$ such that there exists $U_1, U_2, V_1$, and $\sigma$ with
   (a) $U_1 \to V_1 \in \|r\|^i$,
   (b) $U_2 \in \|s\|^i$,
   (c) $\sigma$ is the most general unifier of $U_1$ and $U_2$, and
   (d) $V = V_1\sigma$;
3. if $t = !s$, then $X$ is the set of all $V$ such that $V = s{:}U$ and $U \in \|s\|^i$.

Example 82. Let $T$ be a schematic variable for terms and $P, Q$ be schematic variables for formulas. Further let $c$ be a justification for all instances of the scheme $Q \to Q$ and let $d$ be a justification for all instances of $(\mathsf{jt})$, that is all instances of $T{:}P \to P$. Formally we set

$$Q \to Q \in \|c\|^0 \quad \text{and} \quad T{:}P \to P \in \|d\|^0 \ .$$

We find that

$$c{:}(Q \to Q) \in \|!c\|^1 \ .$$

Then we observe that $\sigma := \{T/c, P/(Q \to Q)\}$ is the most general unifier of $T{:}P$ and $c{:}(Q \to Q)$. Hence we get

$$P\sigma \in \|d \cdot (!c)\|^2 \ ,$$

which is

$$Q \to Q \in \|d \cdot (!c)\|^2 \ .$$

Lemma 83. *For each $i$ and each $t$, the set $\|t\|^i$ is finite and can be effectively constructed.*

Proof. By induction on $i$. For $i = 0$ the claim holds by assumption. To show that $\|t\|^{i+1}$ is finite, we first observe that by I.H. the claim holds for $\|t\|^i$. Thus it is enough to show that $X$ is finite and can effectively be constructed. We distinguish the cases for $t$.

1. $t = r + s$. The claim follows by I.H. for $\|r\|^i$ and $\|s\|^i$.

2. $t = r \cdot s$. By I.H. there are only finitely many $U_1, U_2, V_1$ such that $U_1 \to V_1 \in \|r\|^i$ and $U_2 \in \|s\|^i$. Moreover, they can be effectively constructed. Thus there are only finitely many $V$ with $V = V_1\sigma$ where $\sigma$ is the mgu of $U_1$ and $U_2$. Moreover these $V$ also can be effectively constructed.
3. $t = !s$. By I.H. the set $\|s\|^i$ is finite and they can be effectively constructed. Thus there are only finitely many $V$ such that $V = s{:}U$ with $U \in \|s\|^i$ and they also can be effectively constructed.

⊣

DEFINITION 84 (Rank). The *rank* of a term is inductively defined by:

1. $\mathsf{rk}(c) := \mathsf{rk}(x) := 0$
2. $\mathsf{rk}(s+t) := \max(\mathsf{rk}(s), \mathsf{rk}(t)) + 1$
3. $\mathsf{rk}(s \cdot t) := \max(\mathsf{rk}(s), \mathsf{rk}(t)) + 1$
4. $\mathsf{rk}(!s) := \mathsf{rk}(s) + 1$

LEMMA 85. *For each $i$ and each $t$ we have that for all $U$*

$$U \in \|t\|^i \quad \textit{implies} \quad U \in \|t\|^{\mathsf{rk}(t)} \ .$$

PROOF. First remember that we have

$$k \le l \quad \text{implies} \quad \|t\|^k \subseteq \|t\|^l \ . \tag{23}$$

We show the claim by induction on $i$. For $i = 0$ the claim follows immediately by (23) For the induction step suppose $U \in \|t\|^{i+1}$. If $U \in \|t\|^i$, then the claim follows immediately by I.H. Otherwise we have one of the following cases.

1. $t = r + s$ and $U \in \|r\|^i \cup \|s\|^i$. The claim follows from the I.H. and $\mathsf{rk}(r), \mathsf{rk}(s) < \mathsf{rk}(t)$.
2. $t = r \cdot s$ and there is are $V_1, V_2, U_1$ with $V_1 \to U_1 \in \|r\|^i$ and $V_2 \in \|s\|^i$ and $U = U_1\sigma$ where $\sigma$ is the most general unifier of $V_1$ and $V_2$. Let $j := \mathsf{rk}(t) - 1 \ge \mathsf{rk}(r), \mathsf{rk}(s)$. By I.H. and (23) we find $V_1 \to U_1 \in \|r\|^j$ and $V_2 \in \|s\|^j$. Hence we obtain $V \in \|t\|^{\mathsf{rk}(t)}$.
3. $t = !s$. We have $U = s{:}V$ and $V \in \|s\|^i$. By I.H. we get $V \in \|s\|^{\mathsf{rk}(s)}$ and thus $s{:}V \in \|!s\|^{\mathsf{rk}(s)+1}$, which is $U \in \|t\|^{\mathsf{rk}(t)}$

⊣

LEMMA 86. *Let $\mathsf{CS}$ be a decidable schematic constant specification. Let $\mathcal{X} \subseteq \mathsf{Tm} \times \mathcal{L}_{\mathsf{J}}$ be finite. Set $\mathcal{B} = \mathsf{CS} \cup \mathcal{X}$.*
*For each term $t$ and each formula $G$, it is decidable whether $(t, G) \in \mathcal{E}(\mathcal{B})$.*

PROOF. First we construct $\|s\|^0$ for each subterm $s$ of $t$ as follows:

1. First construct $\mathsf{CS}^{\mathsf{s}}$ as in Remark 80. Now we add the schema $U$ to $\|s\|^0$ if $(s, U) \in \mathsf{CS}^{\mathsf{s}}$.
2. If $(s, F) \in \mathcal{X}$ (again there are only finitely many), add $F$ to $\|s\|^0$.

Now we show for each subterm $s$ of $t$ that

$$(s, G) \in \mathcal{E}(\mathcal{B}) \quad \text{if and only if}$$
$$\text{there exists } i \text{ such that } G \text{ is an instance of some } U \in \|s\|^i. \quad (24)$$

We show the direction from left to right by induction on $s$.

1. $s$ atomic. $(s, G) \in \mathcal{E}(\mathcal{B})$ implies $(s, G) \in \mathcal{B}$. If $(s, G) \in \mathcal{X}$, then $G \in \|s\|^0$ and we are done. Otherwise, $(s, G) \in \mathsf{CS}$, which means that $G$ is an instance of an axiom scheme $U$ of $\mathsf{LP}$ such that $U \in \|s\|^0$.
2. $s = u + v$. If $(s, G) \in \mathcal{B}$, we proceed as in the first case. Otherwise, suppose $(u, G) \in \mathcal{E}(\mathcal{B})$. By I.H. we find that $G$ is an instance of some $U \in \|u\|^i$ for some $i$. The claim follows by $\|u\|^i \subseteq \|u + v\|^{i+1}$.
3. $s = u \cdot v$. If $(s, G) \in \mathcal{B}$, we proceed as in the first case. Otherwise, there is a formula $H$ such that
$$(u, H \to G) \in \mathcal{E}(\mathcal{B}) \quad \text{and} \quad (v, H) \in \mathcal{E}(\mathcal{B}) \ .$$
By I.H. we find that there are $V_1, U_1$ and $i_u, i_v$ such that
$$V_1 \to U_1 \in \|u\|^{i_u} \quad \text{and} \quad H \to G \text{ is an instance of } V_1 \to U_1 \ ,$$
the latter meaning that there is a substitution $\tau$ such that
$$H \to G = (V_1 \to U_1)\tau \ .$$
Also by I.H. there is $V_2$ such that
$$V_2 \in \|v\|^{i_v} \quad \text{and} \quad H \text{ is an instance of } V_2 \ .$$
Let $\sigma$ be the most general unifier of $V_1$ and $V_2$. We find that there is a substiution $\mu$ such that $\tau = \sigma\mu$, Hence $G = U_1\sigma\mu$ and thus $G$ is an instance of $U := U_1\sigma$. Moreover $U \in \|u \cdot v\|^{\max(i_u, i_v)+1}$.
4. $s = !u$. If $(s, G) \in \mathcal{B}$, we proceed as in the first case. Otherwise there is a formula $H$ such that $G = u{:}H$ and $(u, H) \in \mathcal{E}(\mathcal{B})$. By I.H. there are $V$ and $i$ such that
$$V \in \|u\|^i \quad \text{and} \quad H \text{ is an instance of } V \ .$$
We thus get $u{:}V \in \|!u\|^{i+1}$ and $u{:}H$ is an instance of $u{:}V$.

We show the direction from right to left by induction on $i$. Base case $i = 0$. There are two possible subcases: either $G = U$ and $(s, U) \in \mathcal{X}$ or $G$ is an instance of axiom scheme $U$ and $(s, G) \in \mathsf{CS}$. In both cases we find $(s, G) \in \mathcal{B}$ and thus $(s, G) \in \mathcal{E}(\mathcal{B})$.

Induction step $i = j + 1$. If $G$ is an instance of some $U \in \|s\|^j$, the claim follows by I.H. Otherwise we distinguish the following cases.

1. $s = u + v$ and $G$ is an instance of some $U \in \|u\|^j$. By I.H. we get $(u, G) \in \mathcal{E}(\mathcal{B})$. We find $(u + v, G) \in \mathcal{E}(\mathcal{B})$ by the closure conditions on $\mathcal{E}(\mathcal{B})$. The case for $s = u + v$ and $G$ is an instance of some $U \in \|v\|^j$ is analogous.

2. $s = u \cdot v$ and there are $V_1, V_2, U$ and $\sigma$ such that
   (a) $V_1 \to U \in \|u\|^j$,
   (b) $V_2 \in \|v\|^j$,
   (c) $\sigma$ is the mgu of $V_1$ and $V_2$, and
   (d) $G$ is an instance of $U\sigma$.

   Since there is a unifier $\sigma$ for $V_1$ and $V_2$, we know that there is a formula $H$ that is an instance of $V_1\sigma$ and also of $V_2\sigma$. It follows that $H \to G$ is an instance of $V_1 \to U$. By I.H. we find $(u, H \to G) \in \mathcal{E}(\mathcal{B})$. Moreover, since $H$ is an instance of $V_2$, we also find by I.H. that $(v, H) \in \mathcal{E}(\mathcal{B})$. Hence by the closure conditions on $E(\mathcal{B})$ we conclude $(u \cdot v, G) \in \mathcal{E}(\mathcal{B})$.
3. $s = !u$ and $G$ is an instance of $U = u{:}V$ for some $V$ such that $V \in \|u\|^j$. Then $G$ has the form $u{:}H$. By I.H. we find $(u, H) \in \mathcal{E}(\mathcal{B})$ and thus $(!u, u{:}H) \in \mathcal{E}(\mathcal{B})$, which is $(s, G) \in \mathcal{E}(\mathcal{B})$.

Hence (24) is established. By Lemma 85 we obtain

$$(s,G) \in \mathcal{E}(\mathcal{B}) \quad \text{if and only if} \quad G \text{ is an instance of some } U \in \|s\|^{\mathsf{rk}(s)}. \tag{25}$$

Now we can easily decide $(t, G) \in \mathcal{E}(\mathcal{B})$ as follows. Construct $\|t\|^{\mathsf{rk}(t)}$, which is a finite set. For each $U$ in this set, decide whether $G$ is an instance of $U$. By (25) we have $(t, G) \in \mathcal{E}(\mathcal{B})$ if and only if such a $U$ exists. ⊣

THEOREM 87. *Let* $\mathsf{CS}$ *be a decidable schematic constant specification. The satisfaction relation for finitary* $\mathsf{CS}$*-models is decidable.*

PROOF. Let $\mathcal{X} \subseteq \mathsf{Tm} \times \mathcal{L}_\mathsf{J}$ be finite and set $\mathcal{B} = \mathsf{CS} \cup \mathcal{X}$. Further let $\mathsf{val}$ be a finite valuation. We consider the finitary model $\mathcal{M} = (\mathsf{val}, \mathcal{B})$ and show that for any formula $F$, it is decidable whether $\mathcal{M} \Vdash F$.

The proof if by induction on $F$. We distinguish the following cases.

1. $F$ is an atomic proposition $P$. Since $\mathsf{val}$ is finite, it is decidable whether $P \in \mathsf{val}$, that is whether $\mathcal{M} \Vdash F$.
2. $F = \neg G$. By I.H. $\mathcal{M} \Vdash G$ is decidable and thus $\mathcal{M} \Vdash F$ is decidable, too.
3. $F = G \to H$. By I.H. $\mathcal{M} \Vdash G$ and $\mathcal{M} \Vdash H$ are decidable and thus $\mathcal{M} \Vdash F$ is decidable, too.
4. $F = t{:}G$. By the previous lemma $(t, G) \in \mathcal{E}(\mathcal{B})$ is decidable.

⊣

## 5C. Establishing decidability

Let $\mathsf{sub}(F)$ be the set of all subformulas of $F$.

DEFINITION 88 (Restricted model). Let $\mathsf{CS}$ be a constant specification and $\mathcal{M} = (\mathsf{val}, \mathcal{B})$ be a generated $\mathsf{CS}$-model. Further let $F$ be a formula. We define the restricted generated $\mathsf{CS}$-model $\mathcal{M} \restriction F = (\mathsf{val} \restriction F, \mathcal{B} \restriction F)$ by

1. $\mathsf{val} \restriction F := \{P \mid P \in \mathsf{val} \text{ and } P \in \mathsf{sub}(F)\}$
2. $\mathcal{B} \restriction F := \{(t, A) \mid (t, A) \in \mathcal{E}(\mathcal{B}) \text{ and } t{:}A \in \mathsf{sub}(F)\} \cup \mathsf{CS}$.

REMARK 89. Let $\mathsf{CS}$ be a decidable schematic constant specification and $\mathcal{M} = (\mathsf{val}, \mathcal{B})$ be a generated $\mathsf{CS}$-model. Since there are only finitely many subformulas of a given formula $F$, the set $\mathsf{sub}(F)$ is finite. Therefore, the restricted generated $\mathsf{CS}$-model $\mathcal{M} \restriction F$ is finitary.

LEMMA 90. *$\mathcal{M} = (\mathsf{val}, \mathcal{B})$ be a generated $\mathsf{CS}$-model and $F$ a formula. For all $A \in \mathsf{sub}(F)$ we have*

$$\mathcal{M} \Vdash A \quad \textit{if and only if} \quad \mathcal{M} \restriction F \Vdash A \ .$$

PROOF. First we show that for all $t{:}B \in \mathsf{sub}(F)$ we have

$$(t, B) \in \mathcal{E}(\mathcal{B}) \quad \text{if and only if} \quad (t, B) \in \mathcal{E}(\mathcal{B} \restriction F) \ . \tag{26}$$

From left to right: $(t, B) \in \mathcal{E}(\mathcal{B})$ and $t{:}B \in \mathsf{sub}(F)$ imply $(t, B) \in \mathcal{B} \restriction F$ and hence $(t, B) \in \mathcal{E}(\mathcal{B} \restriction F)$.

We show the direction from right to left for all formulas of the form $t{:}B$ (not only for those belonging to $\mathsf{sub}(F)$). We proceed by induction on $t$ and distinguish the following cases.

1. $(t, B) \in \mathcal{B} \restriction F$. We find $(t, B) \in \mathcal{E}(\mathcal{B})$ by definition of $\mathcal{B} \restriction F$.
2. $t = r + s$ and $(r, B) \in \mathcal{E}(\mathcal{B} \restriction F)$. By I.H. we get $(r, B) \in \mathcal{E}(\mathcal{B})$ and hence $(r + s, B) \in \mathcal{E}(\mathcal{B})$. The case $t = r + s$ and $(s, B) \in \mathcal{E}(\mathcal{B} \restriction F)$ is similar.
3. $t = r \cdot s$ and there is a $C$ such that
$$(r, C \to B) \in \mathcal{E}(\mathcal{B} \restriction F) \quad \text{and} \quad (s, C) \in \mathcal{E}(\mathcal{B} \restriction F) \ .$$
By I.H. we obtain
$$(r, C \to B) \in \mathcal{E}(\mathcal{B}) \quad \text{and} \quad (s, C) \in \mathcal{E}(\mathcal{B})$$
and thus $(r \cdot s, B) \in \mathcal{E}(\mathcal{B})$.
4. $t = !s$ and $B = s{:}C$ with $(s, C) \in \mathcal{E}(\mathcal{B} \restriction F)$. We find $(s, C) \in \mathcal{E}(\mathcal{B})$ by I.H. and we conclude $(!s, s{:}C) \in \mathcal{E}(\mathcal{B})$.

Hence (26) is established for all $t{:}B \in \mathsf{sub}(F)$.

Now

$$\mathcal{M} \Vdash A \quad \text{if and only if} \quad \mathcal{M} \restriction F \Vdash A \qquad \text{for all } A \in \mathsf{sub}(F)$$

easily follows by induction on $A$. We distinguish the following cases.

1. $A = P$ for $P \in \mathsf{Prop}$. We have $\mathcal{M} \Vdash P$ iff $P \in \mathsf{val}$ iff (since $P \in \mathsf{sub}(F)$) $P \in \mathsf{val} \restriction F$ iff $\mathcal{M} \restriction F \Vdash P$.
2. $A = \neg B$ or $A = B \to C$. These cases follow immediately by I.H.

3. $A = t{:}B$. We have $\mathcal{M} \Vdash t{:}B$ iff $(t, B) \in \mathcal{E}(\mathcal{B})$ iff (since $(t, B) \in \mathsf{sub}(F)$) $(t, B) \in \mathcal{E}(\mathcal{B} \upharpoonright F)$ iff $\mathcal{M} \upharpoonright F \Vdash t{:}B$.

⊣

THEOREM 91. *Let* $\mathsf{CS}$ *be a decidable schematic constant specification. Let* $F$ *be a formula that is not derivable in* $\mathsf{LP}_{\mathsf{CS}}$*. Then there exists a finitary* $\mathsf{CS}$*-model* $\mathcal{M}$ *with* $\mathcal{M} \nVdash F$*.*

PROOF. Let $F$ be such that $\mathsf{LP}_{\mathsf{CS}} \nvdash F$. By Theorem 70 there is a generated model $\mathcal{M}$ such that $\mathcal{M} \nVdash F$. Consider the restricted model $\mathcal{M} \upharpoonright F$. By Remark 89 we know that $\mathcal{M} \upharpoonright F$ is a finitary $\mathsf{CS}$-model. Moreover, by Lemma 90 we conclude $\mathcal{M} \nVdash F$. ⊣

COROLLARY 92. $\mathsf{LP}_{\mathsf{CS}}$ *is decidable for decidable schematic constant specifications* $\mathsf{CS}$*.*

PROOF. Let $\mathbb{C}$ be the class of finitary $\mathsf{CS}$-models for a decidable schematic constant specification $\mathsf{CS}$. By Theorem 66 know that $\mathsf{LP}_{\mathsf{CS}}$ is sound with respect to $\mathbb{C}$ and Theorem 91 gives us completeness of $\mathsf{LP}_{\mathsf{CS}}$ with respect to $\mathbb{C}$. The class $\mathbb{C}$ is recursively enumerable by Corollary 78. Finally, by Theorem 87, the binary relation $\mathcal{M} \Vdash F$ between formulae and models from $\mathbb{C}$ is decidable. Thus we have established the assumptions of Lemma 76 and conclude that $\mathsf{LP}_{\mathsf{CS}}$ is decidable. ⊣

## 5D. A decidable constant specification is not enough

The requirement for the constant specification to be schematic cannot be dropped from Corollary 92.

THEOREM 93. *There exists a decidable constant specification* $\mathsf{CS}$ *such that* $\mathsf{LP}_{\mathsf{CS}}$ *is undecidable.*

PROOF. The proof is by reducing the Halting Problem to provability in $\mathsf{LP}_{\mathsf{CS}}$ for a particular $\mathsf{CS}$. Let $T_i$ stand for the $i$th Turing machine with one input. Let $A_1, A_2, \ldots$ be an effective enumeration of all axioms of $\mathsf{LP}$. Consider the following constant specification

$$\begin{aligned}\mathsf{CS} :=& \big\{\big(c, (A_i \to (A_j \to A_i))\big) \mid T_i(i) \text{ halts after at most } j \text{ steps}\big\}\\ &\cup \{(d, A_i) \mid i = 1, 2, \ldots\}\ .\end{aligned}$$

Clearly this $\mathsf{CS}$ is decidable. At the same time, it can easily be shown that

$$\mathsf{LP}_{\mathsf{CS}} \vdash ((c \cdot d) \cdot d){:}A_i \quad \text{if and only if} \quad T_i(i) \text{ halts}\ .$$

The right side of this equivalence is the Halting Problem, which is known to be undecidable. ⊣

CHAPTER 6

# LOGICAL OMNISCIENCE

## 6A. Logical Omniscience Tests

Modal logic of knowledge contains the epistemic closure principle in the form of axiom (K)

$$\Box(A \to B) \to (\Box A \to \Box B) \ ,$$

which yields an unrealistic feature called *logical omniscience* whereby an agent knows all logical consequences of her assumptions. In particular, a logically omniscient agent who knows the rules of chess would also know whether White has a non-losing strategy, an agent who knows the product of two primes would also know both of those primes, etc.

To handle this problem, we adopt a general, complexity-based view of logical omniscience. Acquiring knowledge consumes certain resources (space, time, attention, etc.). Any adequate model of knowledge should reflect this fact in some degree of generality. Complexity theory provides a reasonable framework for such an approach.

We assume that there is an epistemic logic system $E$ in a language capable of representing epistemic assertions such that for each valid assertion $F$ *is known*, there is a proof of $F$ in $E$. We attribute the logical omniscience effect to a situation in where for some 'short' valid knowledge assertion $F$ *is known*, it is impossible to feasibly find proofs of $F$ in $E$.

Artemov and Kuznets [1] suggested the following Logical Omniscience Test (LOT): an epistemic system $E$ is *not logically omniscient* if for any valid-in-$E$ knowledge assertion $\mathcal{A}$ of type $F$ *is known*, there is a proof of $F$ in $E$, the complexity of which is bounded by some polynom in the size of $\mathcal{A}$. LOT was inspired by the Cook-Reckhow theory of proof complexity [4, 8].

Later, they [2] suggested a more general Strong Logical Omniscience Test (SLOT) based on time complexity: an epistemic system $E$ is *strongly not logically omniscient* if for any valid-in-$E$ knowledge assertion $\mathcal{A}$ of type $F$ *is known*, a proof of $F$ in $E$ can be found in time polynomial in the size of $\mathcal{A}$.

Both LOT and SLOT connect the size of the assertion $F$ *is known* with the ability of the system to feasibly provide an adequate evidence for $F$. In LOT the feasability measure is the proof length, whereas in SLOT it is the time required to obtain a proof.

Unlike many semantic approaches that avoid logical omniscience by denying the agents some or all deductive abilities, which results in a trivialized logic of knowledge [5], the justification logic approach is axiomatic. The built-in justification terms symbolically model reasons why a given fact is known to an agent, which provides a flexible control over the agent's reasoning without imposing rigid bounds. At the same time, explicit knowledge assertions help to lower the complexity of knowledge acquisition by directing the proof search, which allows for a syntactic rather than semantic representation of the bounds of the agent's reasoning. A natural consequence is an increased length of knowledge assertions. Finding a proof is fast because a justified knowledge assertion $t{:}F$ contains the evidence term $t$, which is a symbolic footprint of such a proof.

We identify a logic $\mathsf{L}$ with the set of its theorems, that is $\mathsf{L} := \{A \mid \mathsf{L} \vdash A\}$. Thus we can define what it means to a be proof system for $\mathsf{L}$ in the following abstract way [4, 8].

DEFINITION 94. A *proof system* for a logic $\mathsf{L}$ is a binary relation $E \subset \Sigma^\star \times \mathsf{L}$ between words in some alphabet, called proofs, and theorems of $\mathsf{L}$ such that

1. $E$ is computable in polynomial time and
2. for all formulas $F$, $\mathsf{L} \vdash F$ if and only if there exists $y$ with $E(y, F)$.

DEFINITION 95. We call $\mathsf{L}$ an *epistemic system* if some subset $r\mathcal{L}$ of its language is designated as a set of *knowledge assertions*. Each knowledge assertion $\mathcal{A} \in r\mathcal{L}$ has an intended meaning *formula $F$ is known* for a unique formula $F$. Moreover, we require that the function $\mathsf{OK}$ from $r\mathcal{L}$ to the language of $\mathsf{L}$ that extracts the *object of knowledge* $F$ from a given knowledge assertion $\mathcal{A}$ is

1. computable in time polynomial in $\mathsf{size}(\mathcal{A})$ and
2. preserving $\mathsf{L}$-validity: for any $\mathcal{A} \in r\mathcal{L}$

$$\mathsf{L} \vdash \mathcal{A} \quad \text{implies} \quad \mathsf{L} \vdash \mathsf{OK}(\mathcal{A}) \ .$$

As far as formulas are concerned, we will concentrate on the two most common size measures:

1. the number of logical symbols in the formula,
2. the bit size of the formula

(the latter takes into account indices of atomic propositions and the like). The results in this chapter hold for both measures. It is also natural to extend these measures to proofs.

DEFINITION 96. Let $\mathsf{L}$ be an epistemic system with a set of knowledge assertions $r\mathcal{L}$. The *reflected fragment* $r\mathsf{L}$ is the set of all $\mathsf{L}$-valid knowledge assertions: $r\mathsf{L} := r\mathcal{L} \cap \mathsf{L}$.

By definition of an epistemic system, $\mathsf{OK}(r\mathsf{L}) \subseteq \mathsf{L}$. If $\mathsf{OK}(r\mathsf{L}) = \mathsf{L}$, we say the reflected fragment $r\mathsf{L}$ is *complete*. In other words, the reflected fragment $r\mathsf{L}$ is complete if its knowledge covers all theorems of $\mathsf{L}$.

DEFINITION 97. Let $E$ be a proof system for an epistemic system $\mathsf{L}$, or simply an *epistemic proof system* where $\mathsf{L}$ can be determined from the context.

- **Logical Omniscience Test (LOT)**: An epistemic proof system $E$ is *not logically omniscient*, or *passes LOT*, if there exists a polynomial $P$ such that for any knowledge assertion $\mathcal{A} \in r\mathsf{L}$, there is a proof of $\mathsf{OK}(\mathcal{A})$ in $E$ with the size bounded by $P(\mathsf{size}(\mathcal{A}))$.
- **Strong Logical Omniscience Test (SLOT)**: An epistemic proof system $E$ is *strongly not logically omniscient*, or *passes SLOT*, if there is a deterministic algorithm, polynomial in $\mathsf{size}(\mathcal{A})$, that, for any knowledge assertion $\mathcal{A} \in r\mathsf{L}$, is capable of restoring a proof of $\mathsf{OK}(\mathcal{A})$ in $E$.

## 6B. The case for $\mathsf{S4}$

For the language of modal logic $\mathcal{L}$, we define the set of knowledge assertions as

$$r\mathcal{L} := \{\Box F \mid F \in \mathcal{L}\}$$

with the associated object of knowledge extraction function

$$\mathsf{OK}(\Box F) := F \ .$$

Thus the reflected fragment of the modal logic $\mathsf{S4}$ is

$$r\mathsf{S4} := \{\Box F \mid \mathsf{S4} \vdash \Box F\} \ .$$

For $\mathsf{S4}$ we have that for all formulas $F$

$$\mathsf{S4} \vdash \Box F \quad \text{if and only if} \quad \mathsf{S4} \vdash F \ . \tag{27}$$

Therefore,

$$r\mathsf{S4} = \{\Box F \mid \mathsf{S4} \vdash F\} \ .$$

This equation tells us that $r\mathsf{S4}$ is very well behaved. For one thing, it is complete, which is certainly a desired property. Further, the object of knowledge extraction function $\mathsf{OK}(\cdot)$ is a one-to-one correspondence between $r\mathsf{S4}$ and $\mathsf{S4}$. Moreover, both the function and its inverse are computable in linear time.

It turns out the the flip side of this coin is logical omniscience of $\mathsf{S4}$. We have the following theorem.

THEOREM 98 ($\mathsf{S4}$ is logically omniscient).

1. *There is no epistemic proof system for* $\mathsf{S4}$ *that passes LOT unless NP=PSPACE.*
2. *There is no epistemic proof system for* $\mathsf{S4}$ *that passes SLOT unless P=PSPACE.*

PROOF. We start with showing the first statement. If a proof system $E$ for $\mathsf{S4}$ passes LOT, then there exists a polynomial $P$ such that for any assertion $\Box F$ with $\mathsf{S4} \vdash \Box F$, there is a proof of $F$ in $E$ with the size bounded by $P(\mathsf{size}(\Box F))$. By (27) and

$$\mathsf{size}(\Box F) = \mathsf{size}(F) + 1 \ ,$$

we get that for some polynomial $P$

$$\begin{aligned} \mathsf{S4} \vdash \Box F \qquad & \text{if and only if} \\ & \quad \text{there is a proof of } F \text{ in } E \\ & \quad \text{with size bounded by } P(\mathsf{size}(F)). \end{aligned}$$

It remains to note that derivability in $\mathsf{S4}$ is PSPACE-hard whereas guessing polynomial-size proof and verifying it works as an NP decision procedure.

For the second statement, we have that if proof system for $\mathsf{S4}$ passes SLOT, then for any assertion $\Box F$ with $\mathsf{S4} \vdash \Box F$, there is a proof of $F$ in $E$ that can be found by a deterministic algorithm polynomial in $\mathsf{size}(\Box F)$. Similar as above we thus find that there is a deterministic algorithm polynomial in $\mathsf{size}(F)$ for deciding whether $F$ is provable in $E$. Since $E$ is a proof system for $\mathsf{S4}$, we get a P decision procedure for $\mathsf{S4}$. ⊣

## 6C. Reflected fragments for $\mathsf{LP_{CS}}$

We start with the question of what is the right form of knowledge assertion in the language of justification logics. The first answer that comes to mind is, by analogy with modal logics,

$$r\mathcal{L}_\mathsf{J} := \{t{:}F \mid t \text{ is a term, } F \text{ is a formula}\}$$

with the associated object of knowledge extraction function

$$\mathsf{OK}(t{:}F) := F \ .$$

LEMMA 99. *If* $\mathsf{CS}$ *is axiomatically appropriate, then* $r\mathsf{LP_{CS}}$ *is complete for* $\mathsf{LP_{CS}}$.

Proof. Let $F$ be an arbitrary theorem of $\mathsf{LP}_{\mathsf{CS}}$. Since $\mathsf{CS}$ is axiomatically appropriate, we obtain by constructive necessitation that there exists a term $t$ such that $\mathsf{LP}_{\mathsf{CS}} \vdash t{:}F$. Because $t{:}F$ is a knowledge assertion, i.e. $t{:}F \in r\mathcal{L}_{\mathsf{J}}$, we find $t{:}F \in r\mathsf{LP}_{\mathsf{CS}}$. By $\mathsf{OK}(t{:}F) = F$ we conclude that $r\mathsf{LP}_{\mathsf{CS}}$ is complete. ⊣

Justification terms are intended to serve as a persuasive reason for the knowledge of a formula. In this respect, knowledge assertions of type $t{:}F$ are not quite satisfactory since both $t$ and $F$ may contain justification constants, the meanings of which are given only in the corresponding constant specification. But an infinite constant specification may contain an infinite amount of information, and so can each knowledge assertion $t{:}F$. Naturally, infinite information in one formula can cause logical omniscience.

To overcome this problem we introduce another form of knowledge assertions for justification logic. Let $\mathsf{CS}_{\mathsf{fin}} = \{(c_1, A_1), \dots, (c_n, A_n)\}$ be a finite constant specification. Then we write $\bigwedge \mathsf{CS}_{\mathsf{fin}}$ for the formula $c_1{:}A_1 \wedge \cdots \wedge c_n{:}A_n$.

Definition 100. The set of *comprehensive knowledge assertions* is defined by

$$cr\mathcal{L}_{\mathsf{J}} := \big\{ \bigwedge \mathsf{CS}_{\mathsf{fin}} \to t{:}F \mid t \text{ is a term, } F \text{ is a formula, } \mathsf{CS}_{\mathsf{fin}} \text{ is a finite constant specification} \big\}$$

with the associated object of knowledge extraction function

$$\mathsf{OK}(\bigwedge \mathsf{CS}_{\mathsf{fin}} \to t{:}F) := F \ .$$

By $\mathsf{LP}_{\emptyset}$ we denote $\mathsf{LP}_{\mathsf{CS}}$ with $\mathsf{CS}$ being the empty constant specification. We define the *comprehensive reflected fragment* for $\mathsf{LP}_{\mathsf{CS}}$ as

$$cr\mathsf{LP}_{\mathsf{CS}} := \big\{ \bigwedge \mathsf{CS}_{\mathsf{fin}} \to t{:}F \mid \mathsf{LP}_{\emptyset} \vdash \bigwedge \mathsf{CS}_{\mathsf{fin}} \to t{:}F \text{ for some finite } \mathsf{CS}_{\mathsf{fin}} \subseteq \mathsf{CS} \big\} \ .$$

Since each $\mathsf{LP}_{\mathsf{CS}}$-derivation uses the Axiom Necessitation rule only finitely many times, each $\mathsf{LP}_{\mathsf{CS}}$-derivation of $t{:}F$ can be turned into an $\mathsf{LP}_{\emptyset}$-derivation of $\bigwedge \mathsf{CS}_{\mathsf{fin}} \to t{:}F$ for some finite constant specification $\mathsf{CS}_{\mathsf{fin}}$.

Lemma 101. *For any constant specification* $\mathsf{CS}$*, we have*

$$\mathsf{LP}_{\mathsf{CS}} \vdash t{:}F$$

*if and only if*

$$\mathsf{LP}_{\emptyset} \vdash \bigwedge \mathsf{CS}_{\mathsf{fin}} \to t{:}F \textit{ for some finite } \mathsf{CS}_{\mathsf{fin}} \subseteq \mathsf{CS} \ .$$

Corollary 102. *If* $\mathsf{CS}$ *is axiomatically appropriate, then the comprehensive reflected fragment* $cr\mathsf{LP}_{\mathsf{CS}}$ *is complete for* $\mathsf{LP}_{\mathsf{CS}}$.

Proof. Let $F$ be an arbitrary theorem of $\mathsf{LP}_{\mathsf{CS}}$. Since $\mathsf{CS}$ is axiomatically appropriate, we obtain by constructive necessitation that there exists a term $t$ such that $\mathsf{LP}_{\mathsf{CS}} \vdash t{:}F$. By the previous lemma there exists a finite constant specification $\mathsf{CS}_{\mathsf{fin}} \subseteq \mathsf{CS}$ with $\mathsf{LP}_{\emptyset} \vdash \bigwedge \mathsf{CS}_{\mathsf{fin}} \to t{:}F$ and thus $\mathsf{CS}_{\mathsf{fin}} \to t{:}F$ belongs to $cr\mathsf{LP}_{\mathsf{CS}}$. By $\mathsf{OK}(\mathsf{CS}_{\mathsf{fin}} \to t{:}F) = F$ we obtain the $cr\mathsf{LP}_{\mathsf{CS}}$ is complete for $\mathsf{LP}_{\mathsf{CS}}$. ⊣

Since we have

$$\mathsf{LP}_{\emptyset} \vdash \bigwedge \mathsf{CS}_{\mathsf{fin}} \to t{:}F \quad \text{iff} \quad \mathsf{LP}_{\mathsf{CS}_{\mathsf{fin}}} \vdash t{:}F \ ,$$

the comprehensive reflected fragment can also be seen as a combination of $r\mathsf{LP}_{\mathsf{CS}_{\mathsf{fin}}}$ for all possible finite $\mathsf{CS}_{\mathsf{fin}}$. None of them need to be complete for $\mathsf{LP}_{\mathsf{CS}}$, but their combination is. We need the following complexity results.

Theorem 103.

1. *Let $\mathsf{CS}$ be a schematic constant specification. Then $r\mathsf{LP}_{\mathsf{CS}}$ is in NP.*
2. *Let $\mathsf{CS}$ be an arbitrary constant specification. Then $cr\mathsf{LP}_{\mathsf{CS}}$ is in P.*

## 6D. The general case

In this section, we outline the relationship between the complexity of a reflected fragment of an epistemic system and the logical omniscience of this system.

Theorem 104. *Let $r\mathsf{L}$ be a complete reflected fragment of an epistemic system $\mathsf{L}$.*

1. *If $r\mathsf{L}$ in in P, then there exists a proof system for $\mathsf{L}$ that passes SLOT.*
2. *If $r\mathsf{L}$ is in NP, then there exists a proof system for $\mathsf{L}$ that passes LOT.*

Proof. For the first claim, suppose $r\mathsf{L}$ in in P. We define the following proof system for $\mathsf{L}$

$$E := \{(\mathcal{A}, F) \mid \mathsf{OK}(\mathcal{A}) = F \text{ and } \mathcal{A} \in r\mathsf{L}\} \ .$$

The so defined $E$ is indeed a proof system:

1. $E$ is computable in polynomial time since $r\mathsf{L}$ is in P and given $\mathcal{A}$ and $F$ testing whether $\mathsf{OK}(\mathcal{A}) = F$ is also in P.
2. For each $F$ with $\mathsf{L} \vdash F$, there exists a knowledge assertion $\mathcal{A}$ such that $E(\mathcal{A}, F)$ holds since $r\mathsf{L}$ is assumed to be complete.

Of course, $E$ passes SLOT since restoring $\mathcal{A}$ (as proof) given $\mathcal{A}$ (as knowledge assertion) is polynomial in $\mathsf{size}(\mathcal{A})$.

To show the second claim, suppose that $M$ is a non-deterministic Turing machine for deciding $r\mathsf{L}$. Let $\lambda_{\mathcal{A}}$ be the sequence of choices made by $M$ given $\mathcal{A}$ as its input. We define a proof system for $\mathsf{L}$ as follows.

$$E := \{((\mathcal{A}, \lambda_{\mathcal{A}}), F) \mid \mathsf{OK}(\mathcal{A}) = F \text{ and } \mathcal{A} \in r\mathsf{L}\} \ .$$

The so defined $E$ is indeed a proof system:

1. First we show that $E$ is computable in polynomial-time. Clearly a deterministic polynomial-time Turing machine can emulate the non-deterministic machine $M$ given the sequence of choices along one of $M$'s branches. Hence given the input $(\mathcal{A}, \lambda_{\mathcal{A}})$ it is decidable in polynomial-time whether $\mathcal{A} \in r\mathsf{L}$. Moreover, again testing whether $\mathsf{OK}(\mathcal{A}) = F$ is also in P given $\mathcal{A}$ and $F$.
2. As before, any theorem of $\mathsf{L}$ has a proof in $E$ since $r\mathsf{L}$ is assumed to be complete.

Note that the size of $\lambda_{\mathcal{A}}$ is polynomial in the size of $\mathcal{A}$ since the Turing machine $M$ runs in time polynomial in $\mathcal{A}$ and thus its computation can involve only polynomially many choices. Therefore, the size of the proof $(\mathcal{A}, \lambda_{\mathcal{A}})$ is polynomial in $\mathcal{A}$ and hence the proof system $E$ passes LOT. ⊣

COROLLARY 105. 1. *If* $\mathsf{CS}$ *is an axiomatically appropriate constant specification, then* $\mathsf{LP}_{\mathsf{CS}}$ *as an epistemic system with the comprehensive reflected fragment* $cr\mathsf{LP}_{\mathsf{CS}}$ *passes SLOT (with respect to a certain proof system).*

2. *If* $\mathsf{CS}$ *is a schematic axiomatically appropriate constant specification, then* $\mathsf{LP}_{\mathsf{CS}}$ *as an epistemic system with the reflected fragment* $r\mathsf{LP}_{\mathsf{CS}}$ *passes LOT (with respect to a certain proof system).*

PROOF. The two claims follow immediately from Theorem 104 and Theorem 103 together with Corollary 102 and Lemma 99, respectively. ⊣

## 6E. Notes

This chapter is based on [1, 2] where the logical omniscience tests LOT and SLOT have been introduced and studied in depth.

# REFERENCES


[1] Sergei [N.] Artemov and Roman Kuznets, *Logical omniscience via proof complexity*, ***Computer Science Logic, 20th international workshop, CSL 2006, 15th annual conference of the EACSL, Szeged, Hungary, September 25–29, 2006, proceedings*** (Zoltán Ésik, editor), Lecture Notes in Computer Science, vol. 4207, Springer, 2006, pp. 135–149.

[2] ———, *Logical omniscience as a computational complexity problem*, ***Theoretical Aspects of Rationality and Knowledge, proceedings of the twelfth conference (TARK 2009)*** (Stanford University, California) (Aviad Heifetz, editor), ACM, July 6–8, 2009, pp. 14–23.

[3] Vladimir [N.] Brezhnev and Roman Kuznets, *Making knowledge explicit: How hard it is*, ***Theoretical Computer Science***, vol. 357 (2006), no. 1–3, pp. 23–34.

[4] Stephen Cook and Robert Reckhow, *On the length of proofs in the propositional calculus (preliminary version)*, ***Conference record of sixth annual acm symposium on theory of computing 1974***, ACM Press, 1974, pp. 135–148.

[5] Joseph Y. Halpern and Riccardo Pucella, *Dealing with logical omniscience*, ***Proceedings of the 11th conference on theoretical aspects of rationality and knowledge (tark'07)*** (Dov Samet, editor), ACM, 2007, pp. 169–176.

[6] Roman Kuznets, *Complexity issues in justification logic*, ***Ph.D. thesis***, CUNY Graduate Center, May 2008.

[7] ———, *Self-referentiality of justified knowledge*, ***Computer science — theory and applications, third international Computer Science symposium in Russia, CSR 2008, Moscow, Russia, June 7–12, 2008, proceedings*** (Edward A. Hirsch, Alexander A. Razborov, Alexei Semenov, and Anatol Slissenko, editors), Lecture Notes in Computer Science, vol. 5010, Springer, 2008, pp. 228–239.

[8] Pavel Pudlák, *The lengths of proofs*, ***Handbook of proof theory*** (Sam Buss, editor), Elsevier, 1998, pp. 209–335.