Univerzita Karlova Filozofická fakulta

Provability Logic of the Alternative Set Theory

Diplomová práce

Studijní obor: logika
Vypracoval: Emil Jeřábek
Vedoucí diplomové práce: RNDr. Vítězslav Švejdar, CSc.

Katedra logiky **Praha, 2001**

Prohlašuji, že jsem diplomovou práci vypracoval samostatně s využitím uvedených pramenů a literatury.

I would like to thank RNDr. Vítězslav Švejdar for helpful comments on preliminary versions of this thesis, and for general support.

Contents

Introduction

The idea of provability logic arose in the seventies in work of G. Boolos, R. Solovay, and others, as an attempt to explore certain "modal effects" in the metamathematics of the first order arithmetic. Namely, the formal provability predicate $\mathrm{Pr}_\tau(x)$, originally constructed by Gödel, has several properties resembling the necessity operator of common modal logics: the Löb's derivability conditions,

$$T \vdash \varphi \;\;\Rightarrow\;\; T \vdash \mathrm{Pr}_\tau(\ulcorner\varphi\urcorner),$$
$$T \vdash \mathrm{Pr}_\tau(\ulcorner\varphi \to \psi\urcorner) \to (\mathrm{Pr}_\tau(\ulcorner\varphi\urcorner) \to \mathrm{Pr}_\tau(\ulcorner\psi\urcorner)),$$
$$T \vdash \mathrm{Pr}_\tau(\ulcorner\varphi\urcorner) \to \mathrm{Pr}_\tau(\ulcorner\mathrm{Pr}_\tau(\ulcorner\varphi\urcorner)\urcorner)$$

look just like an axiomatization of a subsystem of S4:

$$\vdash \varphi \;\;\Rightarrow \vdash \Box\varphi,$$
$$\vdash \Box(\varphi \to \psi) \to (\Box\varphi \to \Box\psi),$$
$$\vdash \Box\varphi \to \Box\Box\varphi.$$

We may form "arithmetical semantics" for formulas in the propositional modal language as follows: we substitute arithmetical sentences for propositional atoms, Pr_τ for boxes, and we ask whether the resulting sentence (the "arithmetical realization" or "provability interpretation" of the modal formula) is provable in our arithmetic T. The provability logic then consists of modal formulas, which are "valid" in every such "model".

Solovay showed that this simple provability logic (known as GL) has a nice axiomatization, Kripke-style semantics and decision procedure. Moreover it is very stable: almost all reasonable theories T yield the same logic.

Further investigation concentrated on generalization of the Solovay's result. In one direction, we may ask about the provability logic for theories which are not covered by the "almost all" above. This concerns e.g. theories based on the intuitionistic logic, such as **HA**, $\mathbf{HA} + MP + ECT_0$ etc., and weak classical theories, such as $I\Delta_0 + \Omega_1$ or even S^1_2.

The second direction is to change the meaning of the modal operator. We may replace Pr_τ with some more pathological provability predicate (e.g. the "Rosser's provability predicate", which enables T to prove its own consistency), provability predicates for non-r.e. theories (such as the second order arithmetic with the ω-rule), "validity in all transitive models" in strong enough set theories and so on.

More importantly, we may use a binary modal connective expressing relative interpretability over the base theory, or a similar binary relation (Π^0_1-conservativity, local interpretability, Σ^0_1-interpolability, "tolerance" etc.).

Finally, we may take two (or more) theories into account. The simplest way is to keep the modal language with one operator, translated as the provability predicate for the first theory, T, and form a logic consisting of modal formulas, such that all their arithmetical realizations are provable in the second theory, S. (A remarkable special case is $S = Th(\mathbb{N})$, the "true arithmetic", which leads to the so-called absolute provability logic of T.) These logics were completely classified for any reasonable choice of T and S, due to S. Artëmov, L. Beklemishev and others.

Another way (perhaps more natural) is to use a bimodal language, with two separate necessity operators (say, $\Box$ and $\triangle$) corresponding to provability predicates for both of the theories, Pr_τ and Pr_σ. Such a bimodal logic (denoted by $\mathrm{PRL}(T, S)$) is capable of expressing basic relationship between T and S, e.g. certain reflection principles, partial conservativity or axiomatization properties (such as finite or bounded complexity axiomatizability of one theory over the other). No general characterizations of possible bimodal logics are known, in fact only a few of them were described so far, mostly for natural pairs of subsystems of **PA**. The first known was the bimodal logic for locally essentially reflexive pairs of sound theories (e.g. $\mathrm{PRL}(\mathbf{PA}, \mathbf{ZF})$ or $\mathrm{PRL}(I\Sigma_1, \mathbf{PA})$), given by T. Carlson (see [Car86]), five other systems are due to L. Beklemishev ([Bek94] and [Bek96])—typical situations where they are applicable include $\mathrm{PRL}(I\Sigma_k, I\Sigma_\ell)$, $\mathrm{PRL}(I\Delta_0 + EXP, \mathbf{PRA})$, $\mathrm{PRL}(\mathbf{PA}, \mathbf{PA} + \mathrm{Con}(\mathbf{ZF}))$, $\mathrm{PRL}(\mathbf{PA}, \mathbf{PA} + \{\mathrm{Con}^n(\mathbf{PA});\, n \in \omega\})$, $\mathrm{PRL}(\mathbf{ZFC}, \mathbf{ZFC} + CH)$. (Here $\mathrm{Con}^n(T)$ is the iterated consistency assertion for T: $\mathrm{Con}^1(T) = \mathrm{Con}(T)$, $\mathrm{Con}^{n+1}(T) = \mathrm{Con}(T + \mathrm{Con}^n(T))$.)

The formation of a bimodal provability logic needs both theories to be formulated in one and the same language (usually, but not necessarily, the language of the arithmetic). If we use theories with different languages, such as in the example $\mathrm{PRL}(\mathbf{PA}, \mathbf{ZF})$ above, it is tacitly assumed that there is a fixed natural interpretation of the first theory in the second one (e.g. the standard model of **PA** in **ZF**), and we treat the *second* theory as the set of all sentences of the language of the *first* theory, which are provable in the second theory under this interpretation (i.e. the arithmetical sentences provable in **ZF** about ω, in our example). Alternatively, we may identify the *first* theory with the set of its axioms interpreted in the language of the *second* theory.

In this thesis, we will study an extension of the bimodal provability logic, designed for the situation of two particular theories with two different languages. We will distinguish between the two languages even at the modal level, and perhaps most importantly, we will deal with *two* different interpretations of the first theory in the second one. Thus our modal language will contain:

- two sorts of formulas, corresponding (under the "arithmetical" realization) to the two first-order languages of the theories in question,
- two modal operators, each one applicable only to formulas of one sort, corresponding to the two provability predicates of our theories,

- an additional sort-switching operator, which corresponds to one of our interpretations of the first theory in the second one.

(One would expect that there were *two* sort-switching operators, one for each interpretation. However this would decrease significantly the readability of the resulting modal formulas, and anyway four non-boolean connectives is a lot, therefore we decided not to include the second sort-switching operator into our modal language. Instead, we allow formulas of the first sort to act directly as formulas of the second sort, i.e. the second operator is "invisible". No ambiguity arises, because the context always determines uniquely the sort of a formula.)

Our two theories are Peano arithmetic (**PA**) and the Alternative Set Theory (**AST**) of P. Vopěnka (axiomatized by A. Sochor). There were several reasons for this choice:

- Both of these theories are simple enough, their metamathematical properties were thoroughly studied, especially in the case of **PA**.
- In **AST** there are two canonical natural interpretations of **PA**, given by the class of the *natural numbers* (**N**) and its proper initial segment, the class of the so-called *finite natural numbers* (**FN**). Note that this is a common situation in theories, formalizing some sort of the Nonstandard Analysis: there we have the (standard set of) internal natural numbers, which form a proper end-extension of the (external set of) standard natural numbers. However in such theories, this end-extension is usually elementary (by the Transfer Principle), which means that both types of numbers generate provably equivalent interpretations of arithmetic and are indistinguishable by means of the provability logic. We will see that this is not the case in **AST**, the interpretations given by **N** and **FN** behave very differently.
- Of course, there were also personal reasons. I like **AST** and I was aware of some strange-looking modal-like principles governing the interplay of **N** and **FN**, therefore I supposed it would be interesting to study it more deeply.

The material is organized as follows. Chapter 1 deals with the Alternative Set Theory. The goal of this chapter is to present everything about **AST** that we will need for the treatment of our provability logic. We do not expect **AST** to be a "common knowledge", hence we have included a detailed description of its axioms. Then we give some elementary facts provable in **AST** and we introduce a bit of the model theory of the classical first-order logic in **AST** (because the derivation of the most important modal principle we use depends on a construction of saturated models within **AST**). We do not go into details in this chapter, we just briefly sketch some basic steps with references to the (hopefully original) sources. A self-contained presentation would be possible, but it would be too long for our purposes and it would lead us far away from the main subject of this thesis (which is the provability logic), anyway only a small piece of chapter 1 is new here (this small piece is given with full proof, of course).

Chapter 2 investigates the provability logic. In section 2.1 we define our extension of the bimodal language and its intended "arithmetical" semantics, and we present an axiomatization of our provability logic and two auxiliary systems. Section 2.2 starts with the definition of a variant of the Kripke semantics suitable for our purposes, then we prove that the two auxiliary systems are complete w.r.t. their Kripke semantics. In section 2.3 we prove the arithmetical completeness of the provability logic, using the Kripke completeness results of section 2.2. As the proof is rather complicated, we have broken part of it into separate lemmas. We end this section with some examples, and we also put here several random facts that we considered worth mentioning, without a detailed discussion. In particular, we include here a short description of some interesting subsystems of our provability logic, which use the ordinary bimodal language and are therefore comparable to the above mentioned bimodal provability logics of L. Beklemishev and T. Carlson.

Chapter 2 is intended to be (more or less) self-contained. Apart from the very end of section 2.3, we give full proofs of everything we state here. We need of course some information on **AST** from chapter 1, but actually *everything* we use of it are the existence of the **N** and **FN** interpretations from 1.2.6 and 1.2.10, the soundness of **AST** from 1.3.6, and the contents of the theorem 1.3.7. We also assume the reader is familiar with some basics of the metamathematics of arithmetical theories, such as the Löb's theorem, Gödel's Diagonal lemma, provable Σ^0_1-completeness, representation of recursive functions and relative interpretation.

As for the notation used in this thesis, we hope it is either standard or defined here. We employ the widely used "dots-and-corners" convention, such as in the (somewhat silly) example below:

$$\mathrm{Pr}_\tau(\ulcorner \varphi \,\&\, x \to \psi(\dot{y}) \urcorner).$$

This is a formula with two free variables, x and y, and the value of x is expected to be a Gödel number of a sentence.

Given a theory T with its axiom set represented by a "primitive recursive" formula $\tau(x)$, we construct in a natural way another "primitive recursive" formula $\mathrm{Prf}_\tau(x, y)$, formalizing the predicate "x is a Gödel number of a proof in T of a formula with Gödel number y". Then the *provability predicate* for T is the Σ^0_1-formula

$$\mathrm{Pr}_\tau(y) \quad = \quad \exists x \ \mathrm{Prf}_\tau(x, y).$$

The formalized consistency statement for T is the Π^0_1-sentence

$$\mathrm{Con}_\tau \quad = \quad {\sim}\,\mathrm{Pr}_\tau(\ulcorner \bot \urcorner),$$

where $\bot$ is the Boolean constant for "falsity" or "contradiction".

When dealing with particular theories such as **PA** or **AST**, we always assume that their axiom set is defined by a formula τ constructed naturally according to the standard description of their axioms, we do not want to explore here any strange behavior arising from an unusual numeration of such theories. (The same remark applies also to the assignment of Prf_τ to τ above, of course.) In this case, we write simply Pr_T for Pr_τ, and $\mathrm{Con}(T)$ for Con_τ.

Relative interpretations are written as superscripts, so if $I : T \rhd S$ is an interpretation and φ a formula in the language of S, then φ^I is the interpretation of this formula in the language of T under I.

Chapter 1

The Alternative Set Theory

The Alternative Set Theory was developed in the seventies by Petr Vopěnka and his seminar (Antonín Sochor, Josef Mlček, Karel Čuda, Blanka Vojtášková, Pavol Zlatoš, Jiří Witzany and many others) as an approach alternative to the Cantorian view on foundation of mathematics, expressed e.g. in the classical set theory **ZFC**. We will not try to explain or defend here the philosophical and phenomenological principles governing the Alternative Set Theory, an interested reader is advised to consult excellent Vopěnka's book [Vop79]. Provability logic, which we will examine, deals rather with metamathematical properties of a formal first–order theory corresponding to the Alternative Set Theory.

Instead of a systematic treatment of the Alternative Set Theory we will present a brief survey of facts needed later in the discussion of the provability logic, with references to the original sources, because a detailed development of the Alternative Set Theory is beyond the scope of this thesis.

1.1 Axioms of AST

The Alternative Set Theory was initially used as an informal framework for doing mathematics, based on general postulates rather than axioms, and open for possibility of adding new principles where needed. There were several attempts to formalize the Alternative Set Theory more rigorously (see e.g. [Mar89]), the most prominent one is the axiomatic theory **AST** due to Antonín Sochor ([Soch79], cf. also [Soch89]; most ideas were present already in [Vop79]), which we will adopt in this thesis.

AST is a theory in the classical first–order predicate calculus with equality in the language consisting of one binary predicate $\in$, the membership relation. There are two types of objects in **AST**, *classes* and *sets,* but officially only classes are objects of the formal theory, sets being defined as classes satisfying the formula $\exists Y\ \ X \in Y$ (abbreviated as $Set(X)$), i.e. a class is a set iff it is a member of another class (cf. usual axiomatics of the von Neumann–Gödel–Bernays set theory **GB**). Traditionally, *capital Latin* letters X, Y, $\dots$ are used as general *class* variables, whereas *small Latin* letters x, y, $\dots$ are reserved for *sets* only. According to this, general formulas of the $\in$–language are denoted by *capital Greek* letters Φ, Ψ, $\ldots$, and *small Greek* letters φ, ψ, $\ldots$ are used for *set formulas,* i.e. formulas with all free variables and quantifiers restricted to sets.

In the sequel we will state the axioms of **AST**. Many of them are formulated using defined concepts, either usual in set theory or specific for **AST**, therefore we will simultaneously state some basic definitions. Of course, it is possible to rewrite all the axioms using $\in$ and $=$ only, but it would result into unintelligible clusters of symbols spanning several lines and we find it useless for our purposes.

Axiom 1. Extensionality: $\forall Z(Z \in X \leftrightarrow Z \in Y) \rightarrow X = Y$

Axiom 2. Comprehension schema: $\exists X\ \forall t\ (t \in X \leftrightarrow \Phi)$,
for all formulas Φ without a free occurrence of X

Definition 1.1.1 (AST) The class X, which is ensured to exist by the comprehension axiom for Φ, is denoted by $\{t;\ \ \Phi\}$. (This class is unique by extensionality.) Using comprehension one also defines usual operations such as $X \cap Y$, $X \cup Y$, $-X$, $\emptyset$, $\{x, y\}$, the universal class $\mathbf{V}$, etc.

Axiom 3. Existence of sets: $Set(\emptyset)\ \&\ \forall x\,\forall y\ \ Set(x \cup \{y\})$

An immediate consequence of this axiom is that the pair $\{x, y\}$ is a set whenever x and y are sets. This enables us to define the ordered pair $\langle x, y\rangle = \{\{x\}, \{x, y\}\}$ and class operations $X \times Y$, $X \restriction Y$, $\mathrm{dom}(X)$, $\mathrm{rng}(X)$, X^{-1}, $X''Y$, and $X \circ Y$ as usual. Also the definition of a *relation* and a *function* is quite standard, we write $Fnc(F)$ for "F is a function".

The form of the next axiom is a bit involved, so we give its motivation first. The idea is that sets in **AST** behave like *finite* sets internally, i.e. as long as we take only set-definable properties (set formulas) into account. In particular, we would like our sets to satisfy the following schema of induction for all *set* formulas φ:

$$\varphi(\emptyset)\ \&\ \forall x\,\forall y\,(\varphi(x) \rightarrow \varphi(x \cup \{y\})) \rightarrow \forall x\,\varphi(x).$$

However for certain technical reasons we need a stronger form of induction: namely, the induction should hold for all *formal* set formulas, which may be written approximately as

$$\forall \phi\ \ \langle \mathbf{V}, \in\rangle \models \ulcorner(\phi(\emptyset)\ \&\ \forall x\,\forall y\,(\phi(x) \rightarrow \phi(x \cup \{y\})) \rightarrow \forall x\,\phi(x))\urcorner.$$

This formulation requires a sort of coding of the logical syntax in **AST** and a formalization of the satisfaction relation $\models$. It is not desirable to develop all such techniques before stating an axiom of the theory, and fortunately it is possible to reformulate the induction axiom using the notion of Gödelian operations.

Definition 1.1.2 (AST) *Ordered pair of classes* X and Y is the class $\langle X, Y\rangle^c = (\{0\} \times X) \cup (\{1\} \times Y)$, where $0 = \emptyset$ and $1 = \{\emptyset\}$. A *coding pair* is any pair of classes $\langle K, S\rangle^c$. A class X is a *member of the system coded by the pair* $\langle K, S\rangle^c$ iff $\exists x \in K\ \ X = S''\{x\}$. By abuse of language we will speak about a *codable system of* *classes* $\mathcal{M} = \langle K, S\rangle^c$ (also called *class of classes*) instead of a coding pair $\langle K, S\rangle^c$. With this terminology we will write $X \in \mathcal{M}$ for "X is a member of the system coded by $\mathcal{M}$" and we will use notations such as $\mathcal{M} = \{X;\ X \in \mathcal{M}\}$.

A codable system $\mathcal{M}$ is *closed under Gödelian operations* iff $\mathbf{E} \in \mathcal{M}$ and for all $X, Y \in \mathcal{M}$ we have $\mathrm{rng}(X) \in \mathcal{M}$, $X^{-1} \in \mathcal{M}$, $Cnv(X) \in \mathcal{M}$, $X \setminus Y \in \mathcal{M}$ and $X \times Y \in \mathcal{M}$, where $\mathbf{E}$ denotes the class $\{\langle x, y\rangle;\ x \in y\}$ and $Cnv(X) = \{\langle x, \langle y, z\rangle\rangle;\ \langle z, \langle x, y\rangle\rangle \in X\}$.

Axiom 4. Induction: There exists a codable system $\mathcal{M}$ closed under Gödelian operations such that $\forall x\ x \in \mathcal{M}$ and

$$\forall X \in \mathcal{M}\ [\emptyset \in X \mathbin{\&} \forall x \forall y\, (x \in X \to x \cup \{y\} \in X) \to X = \mathbf{V}].$$

Definition 1.1.3 (AST) The class of the *finite natural numbers* **FN** is defined as

$$\{x;\ \forall y, z \in x\, (y \subseteq x \mathbin{\&} (y \in z \vee y = z \vee z \in y)) \mathbin{\&} \forall X \subseteq x\ \ Set(X)\}.$$

Axiom 5. Prolongation: $Fnc(F) \mathbin{\&} \mathrm{dom}(F) = \mathbf{FN} \to \exists f\ \ (Fnc(f) \mathbin{\&} F \subseteq f)$

Definition 1.1.4 (AST) A relation $R \subseteq X \times X$ is a *well-ordering* of X (written as $WO(X, R)$) iff it is a strict partial order (i.e. a transitive irreflexive relation) and every non-empty *class* $Y \subseteq X$ has an R-least element (i.e. an $x \in Y$ such that all $y \in Y$ different from x satisfy $\langle x, y\rangle \in R$).

Axiom 6. Choice: $\exists R\ \ WO(\mathbf{V}, R)$

Definition 1.1.5 (AST) A class X is *subvalent* to a class Y (in symbols $X \preceq Y$ or $|X| \leq |Y|$) iff there exists an injective function $F : X \to Y$. Classes X and Y are *equivalent* (written $X \approx Y$ or $|X| = |Y|$) iff there exists a bijection $F : X \to Y$.

Axiom 7. Cardinalities: $X \preceq \mathbf{FN} \vee X \approx \mathbf{V}$

Axiom 8. Foundation (or $\in$-induction): for any set formula φ,

$$\forall x\ \ (\forall y \in x\ \ \varphi(y) \to \varphi(x)) \to \forall x\ \ \varphi(x).$$

The Prolongation Axiom, asserting that any function from **FN** has a prolongation which is a set function, is probably the most important axiom of **AST**, it gives **AST** its special flavour, different from the classical set theory. It has many interesting consequences, e.g. there is a subclass of a set, which is not a set itself. There are also metamathematical facts showing that the Prolongation Axiom plays a key rôle in **AST**, see [Soch82] §7.

1.2 Some basic facts about AST

The material of this section belongs mainly to the folklore of the subject. General treatment of the Alternative Set Theory can be found in [Vop79], some technical details concerning the axiomatization of **AST** are in [Soch79]. An introduction to inductive definitions in **AST** is in [Tz86].

Let us start with an above-mentioned fact: the induction axiom of **AST** yields the corresponding induction schema, by an easy argument essentially equivalent to the usual proof of the normal comprehension schema in the finitely axiomatized version of **GB**.

Lemma 1.2.1 (Sochor [Soch79] §2)
Let φ be a set formula with all free variables among $x_1, \ldots, x_n$, $n \geq 1$. Then **AST** *proves*

(i) *the class $\{\langle x_1, \ldots, x_n\rangle;\ \varphi(x_1, \ldots, x_n)\}$ belongs to every codable system closed under Gödelian operations,*

(ii) *for any sets $u_2, \ldots, u_n$ the class $\{x;\ \ \varphi(x, u_2, \ldots, u_n)\}$ belongs to every codable system closed under Gödelian operations and containing all sets,*

(iii) *the set induction axiom for φ:*

$$\varphi(\emptyset)\ \&\ \forall x\, \forall y\, (\varphi(x) \rightarrow \varphi(x \cup \{y\})) \rightarrow \forall x\ \ \varphi(x),$$

the parameters x_2, ..., x_n being omitted for the sake of readability.

As shown by Vopěnka ([Vop79] ch. I sec. 1), the axioms of extensionality, existence of sets and $\in$-induction (i.e. axioms 1, 3 and 8) together with the set induction schema imply all axioms of the Zermelo-Fraenkel theory of finite sets, $\mathbf{ZF}_{\text{fin}}$, i.e. axioms of pair, sum set, power set, foundation and transitive closure, schemata of separation and replacement for all set formulas, and negation of the axiom of infinity. This leads to a straightforward construction of the natural numbers in **AST**.

Definition 1.2.2 ($\mathbf{ZF}_{\text{fin}}$) A class X is *transitive*, in symbols $Trans(X)$, if every element of X is also a subset of X. The class of the *natural numbers* is defined by

$$\mathbf{N} = \{x;\ \ Trans(x)\ \&\ \forall y, z \in x\ \ (y \in z \vee y = z \vee z \in y)\}.$$

Let $0 = \emptyset$ and $S(x) = x \cup \{x\}$. Also for $x, y \in \mathbf{N}$ we define $x < y \leftrightarrow x \in y$. We write 1 for $S(0)$, 2 for $S(1)$ etc.

Lemma 1.2.3 ($\mathbf{ZF}_{\text{fin}}$; cf. [Vop79] ch. II sec. 1) $\mathbf{N}$ *is a proper transitive class, containing* 0*, closed under* S *and linearly ordered by* $<$*. Every* $x, y \in \mathbf{N}$ *satisfy*

$$S(x) \neq 0,$$

$$S(x) = S(y) \rightarrow x = y,$$

$$x = 0 \vee \exists z \in \mathbf{N}\ \ x = S(z).$$

For any set formula φ we have the principle of ordinal induction,

$$\forall x \in \mathbf{N}\ (\forall y \in x\ \ \varphi(y) \to \varphi(x)) \to \forall x \in \mathbf{N}\ \ \varphi(x),$$

and induction for $\mathbf{N}$,

$$\varphi(0)\ \&\ \forall x \in \mathbf{N}\ \ (\varphi(x) \to \varphi(S(x))) \to \forall x \in \mathbf{N}\ \ \varphi(x).$$

⊣

Lemma 1.2.4 ($\mathbf{ZF_{fin}}$; cf. [Vop79] l.c.) *Let X be a class defined by a set formula (shortly: set-definable), $F : X \to X$ a set-definable function and $x_0 \in X$. Then there is a unique set-definable function $G : \mathbf{N} \to X$ such that $G(0) = x_0$ and $\forall n \in \mathbf{N}\ \ G(S(n)) = F(G(n))$.* ⊣

Corollary 1.2.5 ($\mathbf{ZF_{fin}}$) *There are unique set-definable functions $+ : \mathbf{N} \times \mathbf{N} \to \mathbf{N}$ and $\cdot : \mathbf{N} \times \mathbf{N} \to \mathbf{N}$ satisfying*

$$\begin{aligned} x + 0 &= x, \\ x + S(y) &= S(x + y), \\ x \cdot 0 &= 0, \\ x \cdot S(y) &= x \cdot y + x. \end{aligned}$$

⊣

Corollary 1.2.6 (cf. [Vop79] l.c.) *There is an interpretation* $^{\mathbf{N}}$ *of* **PA** *in* $\mathbf{ZF_{fin}}$ (*and* a fortiori *in* **AST**) *with absolute equality such that the domain of* $^{\mathbf{N}}$ *is the class* $\mathbf{N}$ *and the arithmetical operations are interpreted by the functions* $+$, $\cdot$, S *and* 0 *defined in* 1.2.5 *and* 1.2.2. ⊣

Definition 1.2.7 (AST) A class X is *finite* iff $\forall Y \subseteq X\ \ Set(Y)$, otherwise it is called *infinite*. X is *at most countable* iff $X \preceq \mathbf{FN}$, otherwise it is *uncountable*. X is *countable* iff $X \approx \mathbf{FN}$. We write $Fin(X)$ for "X is finite" and define $\mathbf{Fin} = \{x;\ \ Fin(x)\}$.

Remark 1.2.8 Any finite class is a set. All uncountable classes have the same cardinality as $\mathbf{V}$, by the axiom of cardinalities. Note that $\mathbf{FN} = \mathbf{N} \cap \mathbf{Fin}$, by the definition of these three classes.

The following lemma is an easy consequence of the definition of **FN** and properties of the natural numbers.

Lemma 1.2.9 (AST; [Vop79] l.c.) **FN** *is the smallest class containing* 0 *and closed under* S. ⊣

Corollary 1.2.10 (AST; cf. [Vop79] l.c.) **FN** *is closed under* $+$ *and* $\cdot$, *and the restriction of the arithmetical operations from* $\mathbf{N}$ *to* **FN** *determines an interpretation* $^{\mathbf{FN}}$ *of* **PA** *in* **AST**. *Moreover,* **FN** *satisfies the schema of full induction:*

$$\Phi(0)\ \&\ \forall n \in \mathbf{FN}\ \ (\Phi(n) \to \Phi(S(n))) \to \forall n \in \mathbf{FN}\ \ \Phi(n),$$

Φ any *formula in the language of* **AST**. ⊣

We need to develop in **AST** the logical syntax and basic model theory. One of the most important tools in mathematical logic are definitions and proofs "by structural induction". Due to the comprehension axiom, there is an elegant general framework to handle such an induction in **AST**.

Definition 1.2.11 (AST) Assume

$$\forall X_1 \cdots \forall X_n \forall x_1 \cdots \forall x_m \ \exists! Y \ \ \Phi(X_1, \ldots, X_n, x_1, \ldots, x_m, Y).$$

Then we say that the formula Φ determines a *definable operator on classes* denoted by

$$X_1, \ldots, X_n, x_1, \ldots, x_m \longmapsto Y,$$

where Y is the class satisfying $\Phi(X_1, \ldots, X_n, x_1, \ldots, x_m, Y)$.

A definable operator on classes $X \longmapsto \mathcal{L}(X)$ is called *monotonous* iff

$$\forall X \, \forall Y \ \ (X \subseteq Y \rightarrow \mathcal{L}(X) \subseteq \mathcal{L}(Y)).$$

Lemma 1.2.12 (AST) *Let $X \longmapsto \mathcal{L}(X)$ be a monotonous operator.*

(i) *There is a unique smallest class Y such that $\mathcal{L}(Y) \subseteq Y$, we will denote this class by $\mathcal{L}^*$. Moreover, $\mathcal{L}^*$ is a fix-point of $\mathcal{L}$, i.e. $\mathcal{L}(\mathcal{L}^*) = \mathcal{L}^*$.*

(ii) (*The principle of monotonous induction.*) *$\mathcal{L}(X \cap \mathcal{L}^*) \subseteq X \rightarrow \mathcal{L}^* \subseteq X$, in particular for any formula Φ,*

$$\forall x \in \mathcal{L}(\{u \in \mathcal{L}^*; \ \ \Phi(u)\}) \ \ \Phi(x) \rightarrow \forall x \in \mathcal{L}^* \ \ \Phi(x).$$

(iii) *The relation $\sqsubset$ on $\mathcal{L}^*$, defined by $y \sqsubset x \leftrightarrow \forall X \subseteq \mathcal{L}^* \ \ (x \in \mathcal{L}(X) \rightarrow y \in X)$, is well-founded.*

Proof:
Put $\mathcal{L}^* = \{x; \ \forall X \ \ (\mathcal{L}(X) \subseteq X \rightarrow x \in X)\} = \bigcap\{X; \ \ \mathcal{L}(X) \subseteq X\}$. It is only a matter of routine to show that this choice works. ⧻

Remark 1.2.13 In practice, the part (i) of 1.2.12 is used to cover an inductive definition of a term, formula etc. The part (ii) then provides the corresponding principle of induction "on the complexity of a term (formula, ...)". Finally, to deal with a definition of an object (e.g. a valuation of terms) "by recursion on the complexity of a term", we employ the part (iii) together with a construction by well-founded recursion, which works in **AST** just like in classical **ZF**:

Proposition 1.2.14 (AST; cf. [Vop79] ch. II secs. 1, 3)
(*Construction of classes by well-founded recursion.*) *Let R be a well-founded relation on U and let $x, X \longmapsto \mathcal{L}(x, X)$ be a definable operator. Then there exists a unique relation $S \subseteq U \times \mathbf{V}$ such that $S''\{x\} = \mathcal{L}(x, S \restriction (R^{-1''}\{x\}))$ for every $x \in U$.* ⧻

1.3 Logical syntax and model theory in AST

In this section we will sketch a part of the metatheory of the classical predicate calculus in **AST** and we will discuss some properties of **AST** necessary for our treatment of the provability logic for **AST**. Basic definitions of logical syntax in **AST** are already in [Vop79]. Some issues concerning proof theory and model theory in **AST** are stated in [Soch79, Soch82] and other papers, e.g. [ČV86], [RS81], [Resl79].

In view of the remark 1.2.13, a formalization of the logical syntax in **AST** is very smooth. We define a first order *language* to be a class L equipped with an *arity function* $Ar : L \to \mathbf{FN} \times 2$, where $s \in L$ is an *n-ary predicate* if $Ar(s) = \langle n, 0\rangle$ and it is an *n-ary function symbol* if $Ar(s) = \langle n, 1\rangle$. Using 1.2.12, we define the class $Term(L)$ of the *L-terms* as the smallest class containing the *variables* $\{\ulcorner x_n \urcorner; n \in \mathbf{FN}\}$ and closed under composition with function symbols: if $f \in L$ is an n-ary function symbol and $t_0, \ldots, t_{n-1}$ are L-terms, then $\ulcorner f(t_0, \ldots, t_{n-1})\urcorner$ is an L-term too. (We may put e.g. $\ulcorner x_n \urcorner = n$ and $\ulcorner f(t_0, \ldots, t_{n-1})\urcorner = \langle f, g\rangle$, where g is the function such that $g(i) = t_i$ for $i < n$.)

In a similar fashion we may define inductively the class $Form(L)$ of the *L-formulas*, the sets of bounded and free variables occuring in a formula, the substitution of a term for a variable etc. This suffices to express a simple Hilbert-style calculus for the classical first order predicate logic. We define the notion of a *theory* (just any class of sentences) and *formulas provable in the theory* (this is an inductive definition again).

A *model* $\mathcal{A}$ is a non-empty class A and an assignment of a realization $s^{\mathcal{A}}$ to every symbol $s \in L$, such that $s^{\mathcal{A}}$ is an n-ary (class) relation on A if s is an n-ary predicate, and $s^{\mathcal{A}}$ is an n-ary (class) operation on A if s is an n-ary function symbol. (All this data has to be coded into a single class somehow, but this poses no problem.) A *valuation* in $\mathcal{A}$ is any function $E : \mathbf{FN} \to A$. (Here $E(n)$ is the value assigned by E to the variable $\ulcorner x_n \urcorner$.) The system of all valuations is codable and any valuation is representable by a set, because of the prolongation axiom: for any valuation E there is a set function e such that $E \subseteq e$, conversely any function e with $\mathbf{FN} \subseteq \mathrm{dom}(e)$ and $e''\mathbf{FN} \subseteq A$ determines a valuation $E = e \restriction \mathbf{FN}$.

By recursion on complexity we may extend a valuation E uniquely to all terms $t \in Term(L)$ and we may build a satisfaction relation $\mathcal{A} \vDash \phi[E]$, using the Tarski's truth conditions. Now we know what a model of a theory T (or a formula ϕ) is, and we can define the *semantical consequence* relation, $T \vDash \phi$.

It is clear that **AST** proves basic properties of the first order logic, such as the Deduction Theorem or the soundness of the calculus wrt its semantics. More importantly, **AST** proves the Completeness Theorem:

Theorem 1.3.1 (AST; cf. [Soch79] §3)
Let T be a theory in a language L and ϕ an L-formula.

(*i*) *If T is consistent then it has a model.*

(*ii*) $T \vdash \phi \;\Leftrightarrow\; T \vDash \phi$

Proof (sketch):
As usual, it suffices to derive the first part. Given a consistent theory T, we recursively add Henkin constants to it. We obtain finally a Henkin theory $T' \supseteq T$ (in a language $L' \supseteq L$) and we prove easily that T' is consistent too. Using the axiom of choice we find a well-ordering of the class of all L'-sentences. We construct an increasing chain of consistent theories by recursion along this well-ordering (using 1.2.14) such that the union of the chain, T'', contains ϕ or $\ulcorner\sim\phi\urcorner$ for every L'-sentence ϕ. We get a consistent complete Henkin theory T'' extending T and such a theory has a canonical model. ⫨

We define a model $\mathcal{A}$ to be *countably saturated* if any countable sequence of formulas (with one free variable and with parameters from $\mathcal{A}$) is realizable in $\mathcal{A}$, provided that all its finite subsequences are realizable. (The corresponding notion in classical model theory is an $\aleph_1$-saturated model, or more precisely an $\aleph_1$-compact model, which is a bit weaker notion for uncountable languages.) We have a strengthened version of the Completeness Theorem:

Theorem 1.3.2 (AST; cf. [Soch82] §5)
Any consistent theory has a countably saturated model.

Proof (sketch):
Several methods work. We may e.g. take any model of the theory and construct its ultrapower over a uniform ultrafilter on **FN**, we may adopt the proof of 1.3.1 by adding recursively constants realizing any sequence of formulas consistent with the theory, we may use the so-called revealments (see [SV80]) etc. A crucial ingredient in all these proofs is the prolongation axiom, which enables us to code countable sequences by sets. ⫨

Lemma 1.3.3 (AST) *The structure* $\mathcal{FN} = \langle \mathbf{FN}, 0, S, +, \cdot \rangle$ *is a model of* **PA** *and* $\mathcal{V} = \langle \mathbf{V}, \mathbf{E} \rangle$ *is a model of* $\mathbf{ZF}_{\text{fin}}$.

Proof (sketch):
The first assertion follows almost directly from 1.2.10. By lemma 1.2.1, $\mathcal{V}$ is an interpretation of $\mathbf{ZF}_{\text{fin}}$. To show that it is a model of $\mathbf{ZF}_{\text{fin}}$ we have to demonstrate that it validates all (formal) instances of the set induction schema. By formalization of the proof of 1.2.1 in **AST** we find out that any codable system closed under Gödelian operations and containing all sets has to contain all classes definable in $\mathcal{V}$, hence $\mathcal{V}$ is a model of the set induction schema by the axiom 4. ⫨

Lemma 1.3.4 (PA) $\mathbf{ZF}_{\text{fin}}$ *is a conservative extension of* **PA**, *i.e. for any arithmetical sentence* φ, *if* $\mathbf{ZF}_{\text{fin}} \vdash \varphi^{\mathbf{N}}$ *then* $\mathbf{PA} \vdash \varphi$.

Proof (sketch):
In **PA** we define

$$x \in^I y \quad \Leftrightarrow \quad \left\lfloor \frac{y}{2^x} \right\rfloor \text{ is odd.}$$

It is possible to check that this predicate determines an interpretation I of $\mathbf{ZF}_{\text{fin}}$ in **PA** such that

$$\mathbf{PA} \vdash \varphi \leftrightarrow (\varphi^{\mathbf{N}})^I,$$
$$\mathbf{ZF}_{\text{fin}} \vdash \psi \leftrightarrow (\psi^I)^{\mathbf{N}}$$

for any arithmetical sentence φ and any sentence ψ of the $\in$-language. This implies that our lemma holds. ⧻

Theorem 1.3.5 (PA; Sochor [Soch82] §5) *Let T be an extension of* **AST** *and $\mathcal{M}$ a countably saturated model of $\mathbf{ZF}_{\text{fin}}$ definable in T. Then there is an interpretation I of* **AST** *in T with absolute equality such that T proves $\mathcal{V}^I \simeq \mathcal{M}$ and $\mathcal{FN}^I \simeq \mathcal{FN}$, where $\mathcal{V}^I$ is the structure $\langle \mathbf{V}^I, \mathbf{E}^I \rangle$ and $\mathcal{FN}^I = \langle \mathbf{FN}^I, 0^I, S^I, +^I, \cdot^I \rangle$.*

Proof (sketch):
Let $\mathcal{M} = \langle M, \in^M \rangle$. The interpretation I is defined so that, roughly speaking, $(\text{sets})^I$ are members of M and $(\text{classes})^I$ are subclasses of M. More precisely, we identify any $x \in M$ with its extension $\tilde{x} = \{y \in M;\, y \in^M x\} \subseteq M$, thus we let the domain of I consist of all subclasses of M, and for any such $X, Y \subseteq M$ we put

$$X \in^I Y \quad \Leftrightarrow \quad \exists x \in Y \;\; \tilde{x} = X.$$

It is not hard to show that $(Set(X))^I$ iff $X = \tilde{x}$ for some $x \in M$, moreover the map $x \mapsto \tilde{x}$ is an isomorphic embedding wrt $\in$. This yields $\mathcal{V}^I \simeq \mathcal{M}$, and I is an interpretation of $\mathbf{ZF}_{\text{fin}}$ (in particular of axioms 3 and 8), because $\mathcal{M}$ is a model of $\mathbf{ZF}_{\text{fin}}$. It is clear that I is an interpretation of axioms 1 and 2 (i.e. extensionality and comprehension).

We define a function $\nu : \mathbf{FN} \to M$ such that $(\nu(0))\tilde{} = \emptyset$ and $(\nu(n+1))\tilde{} = (\nu(n))\tilde{} \cup \{\nu(n)\}$. One can show that $\text{rng}(\nu)$ equals $\mathbf{FN}^I$ and ν is an isomorphism of $\mathcal{FN}$ and $\mathcal{FN}^I$.

There is a well-ordering $\prec$ of M. We put $R = \{\langle u, v \rangle^M;\, u \prec v\}$. It follows easily that $(WO(\mathbf{V}, R))^I$, i.e. I is an interpretation of the axiom of choice (6). A similar argument shows that I interprets the axiom of cardinalities (7).

The structure $\mathcal{V}^I \simeq \mathcal{M}$ is a model of $\mathbf{ZF}_{\text{fin}}$, thus $(\mathbf{V}$ is a model of $\mathbf{ZF}_{\text{fin}})^I$. In other words, from the satisfaction relation for $\mathcal{M}$ we may construct easily a $(\text{codable system})^I$ witnessing that I interprets the induction axiom (4).

It remains to show that I is an interpretation of the prolongation axiom (5), and this is the place where the saturation of $\mathcal{M}$ is needed. Assume that $F : \mathbf{FN}^I \to M$, we have to find $f \in M$ such that $F \subseteq \tilde{f}$ and $(\tilde{f}$ is a function$)^I$. Let S be the countable sequence of formulas $\{\ulcorner Fnc(x) \,\&\, \langle n, F(n) \rangle \in x \urcorner;\;\; n \in \mathbf{FN}^I\}$. Every finite subset of S is realized in $\mathcal{M}$, hence there is $f \in M$ realizing the whole sequence, and this f works. ⧻

Proposition 1.3.6 (ZF; Sochor [Soch83] §9) **AST** *is arithmetically sound, i.e. if* $\mathbf{AST} \vdash \varphi^{\mathbf{FN}}$*, φ an arithmetical sentence, then φ holds in $\mathbb{N} = \langle \omega, 0, S, +, \cdot \rangle$, the standard model of arithmetic.*

Proof (*sketch*):
Without loss of generality we may work in $\mathbf{ZFC}+CH$. The Continuum Hypothesis implies that the standard model of $\mathbf{ZF}_{\text{fin}}$, $\langle p_\omega, \in\rangle$, has a saturated elementary extension $\mathcal{A} = \langle A, e^A\rangle$ of cardinality $\aleph_1$. Define a new model $\mathcal{B} = \langle B, e^B\rangle$ by $B = \mathcal{P}(A)$ and

$$e^B = \{\langle x, y\rangle \in B^2;\ \exists u \in y\ \ x = \{v \in A;\ \langle v, u\rangle \in e^A\}\}.$$

The same argument as in theorem 1.3.5 shows that $\mathcal{B} \models \mathbf{AST}$ and $\mathcal{FN}^B \simeq \mathbb{N}$ (moreover $\mathcal{N}^B \simeq \mathcal{N}^A \equiv \mathbb{N}$). If $\mathbf{AST} \vdash \varphi^{\mathbf{FN}}$ then $\mathcal{B} \models \varphi^{\mathbf{FN}}$, i.e. $\mathcal{FN}^B \models \varphi$, therefore $\mathbb{N} \models \varphi$. ⫩⊨

Theorem 1.3.7 *Let φ and ψ be arithmetical sentences and $\sigma(x_1, \dots, x_n)$ a Σ^0_1-formula.*

(*i*) **AST** *proves*

$$\mathrm{Pr}^{\mathbf{FN}}_{\mathbf{PA}}(\ulcorner\varphi\urcorner) \to \varphi^{\mathbf{FN}},$$
$$\mathrm{Pr}^{\mathbf{FN}}_{\mathbf{PA}}(\ulcorner\varphi\urcorner) \to \varphi^{\mathbf{N}},$$
$$\forall x_1, \dots, x_n \in \mathbf{FN}\ \ (\sigma^{\mathbf{FN}}(x_1, \dots, x_n) \to \sigma^{\mathbf{N}}(x_1, \dots, x_n)).$$

(*ii*) **PA** *proves*

$$\mathrm{Pr}_{\mathbf{AST}}(\ulcorner\varphi^{\mathbf{FN}} \dot{\to} \psi^{\mathbf{N}}\urcorner) \to \mathrm{Pr}_{\mathbf{AST}}(\ulcorner\varphi^{\mathbf{FN}} \dot{\to} \mathrm{Pr}^{\mathbf{FN}}_{\mathbf{PA}}(\ulcorner\psi\urcorner)\urcorner).$$

Proof:
Work in **AST**, and suppose that $\mathrm{Pr}^{\mathbf{FN}}_{\mathbf{PA}}(\ulcorner\varphi\urcorner)$, i.e. $\mathbf{PA} \vdash \varphi$. By 1.3.3, $\mathcal{FN} \models \varphi$, thus $\varphi^{\mathbf{FN}}$. Also $\mathbf{ZF}_{\text{fin}} \vdash \ulcorner\varphi^{\mathbf{N}}\urcorner$ by 1.2.6, therefore $\mathcal{V} \models \ulcorner\varphi^{\mathbf{N}}\urcorner$ by 1.3.3, hence $\varphi^{\mathbf{N}}$.

The last formula of (*i*) can be easily demonstrated by an induction on the complexity of σ, using only the fact that **N** is an end-extension of **FN**.

To prove the part (*ii*), it will suffice to present an interpretation of $\mathbf{AST}+\varphi^{\mathbf{FN}}+{\sim}\psi^{\mathbf{N}}$ in the theory $T = \mathbf{AST} + \varphi^{\mathbf{FN}} + \mathrm{Con}^{\mathbf{FN}}(\mathbf{PA} + {\sim}\psi)$. However, by 1.3.2 and 1.3.4, T proves that there is a countably saturated model $\mathcal{M}$ of $\mathbf{ZF}_{\text{fin}} + {\sim}\psi^{\mathbf{N}}$. By 1.3.5 there is an interpretation I of **AST** in T such that $\mathcal{FN}^I \simeq \mathcal{FN}$ and $\mathcal{V}^I \simeq \mathcal{M}$. But $\mathcal{FN} \models \varphi$ and $\mathcal{M} \models {\sim}\psi^{\mathbf{N}}$, hence I is an interpretation of $\mathbf{AST} + \varphi^{\mathbf{FN}} + {\sim}\psi^{\mathbf{N}}$ in T. (Well, in fact 1.3.5 requires a countably saturated model *definable* in the theory in question. Therefore we form a theory $T' = T+$"**M** *is a countably saturated model of* $\mathbf{ZF}_{\text{fin}} + {\sim}\psi^{\mathbf{N}}$" in a language augmented by a new constant **M**. We find an interpretation of $\mathbf{AST} + \varphi^{\mathbf{FN}} + {\sim}\psi^{\mathbf{N}}$ in T' and we realize (in **PA**) that T' is fully conservative over T.) ⫩⊨

Remark 1.3.8 Part (*ii*) of the theorem 1.3.7 is essentially the only thing of this chapter, which is due to the author of this thesis.

Chapter 2

The provability logic

2.1 Basic definitions

Our modal analysis of the provability principles of **AST** will try to explore as much as possible the interplay between the two canonical interpretations of **PA** in **AST**, therefore we chose a rather rich language:

Definition 2.1.1 The *extended bimodal language* uses the following symbols:

- propositional connectives $\rightarrow$ and $\perp$ (the others being defined in the usual way),
- unary modal operators $\triangle$ and $\Box$,
- a unary operator $^{\mathrm{N}}$,
- *arithmetical* propositional atoms p_i (for every $i \in \omega$),
- *general* propositional atoms q_i ($i \in \omega$).

There are two sorts of formulas in the extended language: the *arithmetical modal formulas,* denoted by lowercase Greek letters, and the *general modal formulas* (or simply formulas), denoted by uppercase Latin letters. These are defined inductively as follows:

- every a.m.f. is also a g.m.f.,
- $\perp$ is an a.m.f.,
- every p_i is an a.m.f. and every q_i is a g.m.f.,
- $(\varphi \rightarrow \psi)$ is an a.m.f. and $(A \rightarrow B)$ is a g.m.f. whenever φ, ψ are a.m.f. and A, B are g.m.f.,
- $\triangle A$ and $\Box\varphi$ are a.m.f. whenever A is a g.m.f. and φ an a.m.f.,
- $(\varphi)^{\mathrm{N}}$ is a g.m.f. provided that φ is an a.m.f.

Let AMF and GMF be the sets of all arithmetical and general modal formulas respectively.

Remark 2.1.2 The arithmetical formulas are the prominent ones, the g.m.f. play an auxiliary rôle. Having the provability interpretation in mind, the a.m.f. correspond to sentences of the arithmetic whereas g.m.f. represent sentences in the language of the set theory. We take **FN** as the prominent interpretation of **PA** in **AST**, an a.m.f. used in a g.m.f. context represents an arithmetical sentence interpreted in **FN**. The additional operator $^{\mathrm{N}}$ is used to override this default behavior and to force an arithmetical sentence to be interpreted in **N**. The following definition is a precise formulation of these remarks.

Definition 2.1.3 A *provability interpretation* (or *arithmetical realization*) of the extended language is a pair $* = \langle ^*, _* \rangle$, where * maps all a.m.f. to sentences of **PA**, $_*$ maps g.m.f. to sentences of **AST** and the following inductive clauses hold for every a.m.f. φ, ψ and every g.m.f. A, B:

- $\varphi_* = (\varphi^*)^{\mathbf{FN}}$,
- $\bot^* = \bot$,
- $(\varphi \to \psi)^* = \varphi^* \to \psi^*$, $(A \to B)_* = A_* \to B_*$,
- $(\triangle A)^* = \mathrm{Pr}_{\mathbf{AST}}(\ulcorner A_* \urcorner)$, $(\Box\varphi)^* = \mathrm{Pr}_{\mathbf{PA}}(\ulcorner \varphi^* \urcorner)$,
- $(\varphi^{\mathrm{N}})_* = (\varphi^*)^{\mathbf{N}}$.

The provability logics are defined as follows:

$$\mathrm{PRL}_{ext}(\mathbf{AST}, \mathbf{PA}) = \{\varphi;\ \forall * \ (* \text{ prov. int. } \Rightarrow \ \mathbf{PA} \vdash \varphi^*),\ \ \varphi \text{ is an a.m.f.}\},$$

$$\mathrm{PRL}^{+}_{ext}(\mathbf{AST}, \mathbf{PA}) = \{\varphi;\ \forall * \ (* \text{ prov. int. } \Rightarrow \ \mathbb{N} \models \varphi^*),\ \ \varphi \text{ is an a.m.f.}\}.$$

Remark 2.1.4 In order to save parentheses we adopt the convention that $^{\mathrm{N}}$ has a higher priority than other symbols of our language, so that $\triangle\varphi^{\mathrm{N}}$ reads $\triangle(\varphi^{\mathrm{N}})$. We will sometimes write $^{\mathrm{N}}$ right after the head symbol of a formula, so that $\triangle^{\mathrm{N}}\varphi = (\triangle\varphi)^{\mathrm{N}}$ (following the pattern $\sin^2 x = (\sin x)^2$).

Our main result will be a complete axiomatization of the above mentioned provability logic: we will show that

$$\mathrm{PRL}_{ext}(\mathbf{AST}, \mathbf{PA}) = \mathrm{CSRL},$$
$$\mathrm{PRL}^{+}_{ext}(\mathbf{AST}, \mathbf{PA}) = \mathrm{CSRL}^{\#},$$

where the systems CSRL and $\mathrm{CSRL}^{\#}$ are defined below.

Definition 2.1.5 The axioms of the logic CSRL are the following a.m.f.:

A1) tautologies of the Classical Propositional Calculus,

A2) $\triangle A$, $\quad A$ is a tautology of CPC, B1) $\triangle(\bot^{\mathrm{N}} \to \bot)$,

B2) $\triangle((\varphi \to \psi)^{\mathrm{N}} \to (\varphi^{\mathrm{N}} \to \psi^{\mathrm{N}}))$,

B3) $\triangle((\varphi^{\mathrm{N}} \to \psi^{\mathrm{N}}) \to (\varphi \to \psi)^{\mathrm{N}})$,

C1) $\Box(\varphi \to \psi) \to (\Box\varphi \to \Box\psi)$,

C2) $\triangle(A \to B) \to (\triangle A \to \triangle B)$,

C3) $\Box(\Box\varphi \to \varphi) \to \Box\varphi$,

C4) $\triangle A \to \Box\triangle A$,

C5) $\Box\varphi \to \triangle\varphi$,

D1) $\triangle(\Box\varphi \to \varphi^{\mathrm{N}})$,

D2) $\triangle(\varphi \to \psi^{\mathrm{N}}) \to \triangle(\varphi \to \Box\psi)$,

D3) $\triangle(\Box\varphi \to \varphi)$,

its derivation rules are Modus Ponens and the Necessitation Rule:

MP) $\dfrac{\varphi \quad \varphi \to \psi}{\psi}$

Nec) $\dfrac{\varphi}{\Box\varphi}$

The logic CSRL$^{\#}$ is the closure of CSRL and the schema

S) $\triangle\varphi \to \varphi$

under Modus Ponens.

We define also auxiliary systems L_1 and L_3: both of them include the rules MP and Nec, all axioms from the groups A, B and C, and axiom D1, moreover L_3 contains D2.

Remark 2.1.6 It may seem that thirteen axiom schemata is too much. Note that the axioms from the groups A, B, C and D have a clear meaning: group A says that the whole thing extends the Propositional Calculus, group B expresses the fact that the interpretation $^{\mathbf{N}}$ commutes with Boolean connectives, group C contains well-known axioms of the usual bimodal provability logic for extensions of theories (CSM, see remark 2.3.14), thus only the three axioms of group D give nontrivial information about our pair of theories.

2.2 Kripke completeness

In order to get the arithmetical completeness of CSRL via a Solovay-like argument (cf. [Sol76]), we need a sort of Kripke semantics for the logic discussed. The following definition modifies the Carlson models for bimodal logics ([Car86], cf. [Bek96]).

Definition 2.2.1 An (*extended Kripke*) *frame* is a structure $\mathbf{W} = \langle W, <, D, N \rangle$, where W is a non-empty set, $<$ a binary relation on W, D a subset of W and N a function $N : D \to W$. An (*extended Kripke*) *model* in the frame $\mathbf{W}$ is a pair $\langle \mathbf{W}, \Vdash \rangle$, where $\Vdash$ is a relation $\Vdash \subseteq ((W \times AMF) \cup (D \times GMF))$ satisfying the following conditions ($w \in W$, $d \in D$, $\varphi, \psi \in AMF$, $A, B \in GMF$):

- $w \nVdash \bot$,
- $w \Vdash \varphi \to \psi \iff w \nVdash \varphi$ or $w \Vdash \psi$,
  $d \Vdash A \to B \iff d \nVdash A$ or $d \Vdash B$,
- $w \Vdash \Box\varphi \iff \forall w' \in W\,(w < w' \Rightarrow w' \Vdash \varphi)$,
- $w \Vdash \triangle A \iff \forall d' \in D\,(w < d' \Rightarrow d' \Vdash A)$,
- $d \Vdash \varphi^{\mathrm{N}} \iff N(d) \Vdash \varphi$.

A formula φ is *valid in the model* $\langle \mathbf{W}, \Vdash \rangle$ if $w \Vdash \varphi$ for all $w \in W$, and it is *valid in the frame* $\mathbf{W}$ if it is valid in every model $\Vdash$ in the frame $\mathbf{W}$. The model or frame is *finite* whenever W is and it is *tree-like* if $\langle W, < \rangle$ is a tree.

At first we will deal with the Kripke frame characterization of L_1.

Definition 2.2.2 A frame $\langle W, <, D, N \rangle$ is called an L_1*-frame* if $<$ is transitive and converse well-founded and for every $d \in D$, $d < N(d)$.

Lemma 2.2.3 *Every theorem of L_1 is valid in any L_1-frame.*

Proof:
By induction on the length of a proof. It is clear from the definition of satisfaction that the validity in a given model is preserved under MP and Nec, moreover the evaluation of Boolean connectives coincides with the usual two-valued semantics of the CPC, hence A1 and A2 are valid in every model. As for B1–B3, observe that $d \Vdash (\varphi \to \psi)^{\mathrm{N}}$ iff $N(d) \Vdash (\varphi \to \psi)$ iff $d \Vdash (\varphi^{\mathrm{N}} \to \psi^{\mathrm{N}})$. Axioms C1 and C2: if every $w' > w$ satisfies $\varphi \to \psi$ and every $w' > w$ satisfies φ, then every $w' > w$ satisfies ψ too, the case of $\triangle$ is similar.

C3: let $w \nVdash \Box\varphi$ and let $w' > w$ be a maximal element of W such that $w' \nVdash \varphi$ (it exists by the converse well-foundedness of $<$). By transitivity of $<$ and maximality of w' we have $w' \Vdash \Box\varphi$, therefore $w' \nVdash \Box\varphi \to \varphi$ and $w \nVdash \Box(\Box\varphi \to \varphi)$.

C4: suppose that $w \Vdash \triangle A$ and $w' > w$, we have to show $w' \Vdash \triangle A$. Now if $d \in D$ and $d > w'$ then $d > w$ by transitivity and $d \Vdash A$ by the hypothesis, thus $w' \Vdash \triangle A$ as required.

C5: if every $w' > w$ satisfies φ then *a fortiori* every $d > w$ from D satisfies φ.

D1: let $d \in D$ and $d \Vdash \Box\varphi$. We have $d < N(d)$, thus $N(d) \Vdash \varphi$ and $d \Vdash \varphi^{\mathrm{N}}$. ⊣⊢

Proposition 2.2.4 *Let φ be an a.m.f. The formula φ is provable in L_1 iff it is valid in all finite tree-like L_1-frames $\langle W, <, D, N \rangle$ with* $\mathrm{rng}(N) \cap D = \emptyset$.

Proof:
By the lemma it suffices to show the right-to-left implication. Assume $L_1 \nvdash \varphi$. The proof will proceed as follows: at first we construct a sort of a "universal model" of L_1, which is *not* a model in the sense of our definition, then we transform this structure into a tree-like extended Kripke model, and finally we find its finite subtree, which will turn out to be an L_1-model with all the desired properties.

A set X of a.m.f. is *consistent* provided there are no formulas $\varphi_1, \ldots, \varphi_n \in X$ such that $L_1 \vdash \sim(\varphi_1 \mathbin{\&} \cdots \mathbin{\&} \varphi_n)$. Analogously, a set Y of g.m.f. is defined to be consistent if there do not exist $A_1, \ldots, A_n \in Y$ such that $L_1 \vdash \triangle\sim(A_1 \mathbin{\&} \cdots \mathbin{\&} A_n)$. Let K be the set of all maximal consistent sets of a.m.f. (i.e. consistent sets maximal wrt inclusion) and let G be the set of all maximal consistent sets of g.m.f.

For any $Y \subseteq GMF$ put $Y_{\mathrm{N}} = \{\psi;\ \psi^{\mathrm{N}} \in Y\}$, similarly $Y_{\Box} = \{\psi;\ \Box\psi \in Y\}$ and $Y_{\triangle} = \{A;\ \triangle A \in Y\}$. If $Y \in G$ we define $F(Y) = Y \cap AMF$ and $N(Y) = Y_{\mathrm{N}}$. If $X, X' \in K$ and $Y \in G$, put

$$X < X' \iff X_{\Box} \subseteq X',$$

$$X \prec Y \iff X_{\triangle} \subseteq Y.$$

Immediately from the definition we see that for any $X \in K$, $Y \in G$ and any ψ and A either $\psi \in X$ or $\sim\psi \in X$ and similarly $A \in Y$ or $\sim A \in Y$. As a corollary we get that $F(Y)$ and $N(Y)$ belong to K whenever $Y \in G$.

Any consistent set of a.m.f. or g.m.f. is included in a maximal one, by the Zorn lemma.

Maximal consistent sets are deductively closed: given $X \in K$ and $\psi_1, \ldots, \psi_n$ in X such that $\vdash \psi_1 \mathbin{\&} \cdots \mathbin{\&} \psi_n \to \psi$, we have $\psi \in X$. Similarly if $Y \in G$, $A_1, \ldots, A_n \in Y$ and $\vdash \triangle(A_1 \mathbin{\&} \cdots \mathbin{\&} A_n \to A)$ then $A \in Y$.

Sublemma 1 *Let $X \in K$.*

(i) *If $\Box\psi \notin X$ then $X_{\Box} \cup \{\sim\psi\}$ is consistent.*

(ii) *If $\triangle A \notin X$ then $X_{\triangle} \cup \{\sim A\}$ is consistent.*

Proof:
If $X_{\triangle} \cup \{\sim A\}$ is inconsistent there exist $A_1, \ldots, A_n$ such that $\triangle A_i \in X$ and $\vdash \triangle(A_1 \mathbin{\&} \cdots \mathbin{\&} A_n \to A)$. Then $\vdash \triangle A_1 \mathbin{\&} \cdots \mathbin{\&} \triangle A_n \to \triangle A$ by propositional logic and C2, thus $\triangle A \in X$, a contradiction. The case (i) is analogous (easier). ⊣⊢

Sublemma 2 *Let $X, X', X'' \in K$ and $Y \in G$.*

(i) $X < X' < X'' \Rightarrow X < X''$,

(ii) $X < X' \prec Y \Rightarrow X \prec Y$,

(iii) $X \prec Y \Rightarrow X < F(Y)$,

(iv) $F(Y) < N(Y)$.

Proof:
Suppose that $X < X'$, $X' < X''$ and $\Box\psi \in X$. As L_1 proves[1] $\Box\psi \to \Box\Box\psi$ we have $\Box\Box\psi \in X$, therefore $\Box\psi \in X'$ and $\psi \in X''$. The rest is similar, using the axioms C4, C5 and D1.

Since we assume $\nvdash \varphi$, the set $\{\sim\varphi\}$ is consistent and we can find $X_0 \in K$ such that $\varphi \notin X_0$. We define a tree-like model $\mathbf{W} = \langle W, <, D, N, \Vdash \rangle$ by

$$W = \{\langle X_0, \dots, X_n \rangle; X_i \in K \cup G,\ X_i < X_{i+1} \ \vee\ X_i \prec X_{i+1} \ \vee \\ \vee\ F(X_i) < X_{i+1} \ \vee\ F(X_i) \prec X_{i+1}\},$$

$$D = \{\langle X_0, \dots, X_n \rangle \in W;\ \ X_n \in G\},$$

$$N(\langle X_0, \dots, X_n \rangle) = \langle X_0, \dots, X_n, N(X_n) \rangle \quad \text{(where } X_n \in G\text{)},$$

$$\langle X_0, \dots, X_n \rangle < \langle X_0, Y_1, \dots, Y_m \rangle \ \Leftrightarrow\ n < m,\ X_1 = Y_1, \dots, X_n = Y_n \quad \text{(i.e. } {<}{=}{\subset}\text{)},$$

$$\langle X_0, \dots, X_n \rangle \Vdash A \ \Leftrightarrow\ A \in X_n \quad (A \in AMF \text{ or } X_n \in G).$$

The definition of N is correct, since $F(X_n) < N(X_n)$ for any $X_n \in G$. Obviously $<$ is a strict partial order, in fact a tree with the least element $\langle X_0 \rangle$. If $d \in D$ then $d < N(d)$ and $N(d) \notin D$. For simplicity we write $\underline{X_n} = \langle X_0, \dots, X_n \rangle \in W$ and $\langle \underline{X_n}, X_{n+1}, \dots, X_m \rangle = \langle X_0, \dots, X_n, \dots, X_m \rangle$. For every $X \in K$ we put $H(X) = X$ and for $X \in G$ we define $H(X) = F(X)$ (note that K and G are disjoint, so this makes sense). We claim that $\Vdash$ defines a correct model:

We have $\underline{X_n} \nVdash \bot$, since X_n is consistent.

$\underline{X_n} \Vdash A \to B \ \Leftrightarrow\ (A \to B) \in X_n \ \Leftrightarrow\ A \notin X_n$ or $B \in X_n \ \Leftrightarrow\ \underline{X_n} \nVdash A$ or $\underline{X_n} \Vdash B$ by maximality and consistency of X_n.

$\underline{X_n} \Vdash \psi^{\mathrm{N}} \ \Leftrightarrow\ \psi^{\mathrm{N}} \in X_n \ \Leftrightarrow\ \psi \in N(X_n) \ \Leftrightarrow\ N(\underline{X_n}) \Vdash \psi$.

Suppose $\underline{X_n} \Vdash \Box\psi$ and $\underline{X_n} < \langle \underline{X_n}, X_{n+1}, \dots, X_m \rangle$. By a repeated application of (i), (ii) and (iii) of the second sublemma we get $H(X_n) < H(X_m)$ and by definition $\Box\psi \in X_n$, thus $\Box\psi \in H(X_n)$, $\psi \in H(X_m)$ and $\psi \in X_m$, so that $\langle \underline{X_n}, X_{n+1}, \dots, X_m \rangle \Vdash \psi$. For the converse suppose $\underline{X_n} \nVdash \Box\psi$, then $\Box\psi \notin H(X_n)$, thus by the first sublemma there exists an $X_{n+1} > H(X_n)$ such that $\psi \notin X_{n+1}$. We get $\underline{X_n} < \langle \underline{X_n}, X_{n+1} \rangle$ and $\langle \underline{X_n}, X_{n+1} \rangle \nVdash \psi$.

A similar argument shows that $\underline{X_n} \Vdash \triangle A$ iff $d \Vdash A$ for every $d \in D$ such that $\underline{X_n} < d$.

We have checked that $\mathbf{W}$ is a well-defined extended Kripke model. Moreover $\langle X_0 \rangle \nVdash \varphi$ and every theorem of L_1 is valid in $\mathbf{W}$ (as it is a member of any maximal consistent set). However $\mathbf{W}$ need not be an L_1-model, since the well-foundedness condition may fail for it. We will overcome this problem by taking a suitable finite restriction of $\mathbf{W}$.

Define $n(d) = N(d)$ for $d \in D$ and $n(w) = w$ for $w \in W \setminus D$. Let $\psi_1, \dots, \psi_k$ be the list of all a.m.f. such that $\Box\psi_i$ is a subformula of φ.

We will pick functions $f_1, \dots, f_k$ on W such that the following holds: if $w \Vdash \Box\psi_i$ then $f_i(w) = w$, otherwise $f_i(w) > w$ is such that $f_i(w) \nVdash \psi_i$ and $f_i(w) \Vdash \Box\psi_i$.

[1] First of all, $\Box(\varphi \& \varphi') \leftrightarrow (\Box\varphi \& \Box\varphi')$ by Nec and C1 applied to propositional tautologies $\varphi \to (\varphi' \to \varphi \& \varphi')$ and $\varphi \& \varphi' \to \varphi$, $\varphi \& \varphi' \to \varphi'$. Then $\Box\psi \to \Box(\Box(\psi \& \Box\psi) \to (\psi \& \Box\psi)) \to \Box(\psi \& \Box\psi) \to \Box\Box\psi$ by C3 and C1.

This is possible, because every $w \in K$ satisfies $\Box(\Box\psi_i \to \psi_i) \to \Box\psi_i$, in other words $\sim\Box\psi_i \to \sim\Box\sim(\Box\psi_i \,\&\, \sim\psi_i)$.

In a similar way, we let $A_1, \dots, A_\ell$ list all g.m.f. such that $\triangle A_j$ is a subformula of φ, and we choose functions $g_1, \dots, g_\ell$, so that $g_j(w) = w$ if $w \Vdash \triangle A_j$, or $g_j(w) \in D$, $g_j(w) > w$, $g_j(w) \nVdash A_j$ and $g_j(w) \Vdash \triangle A_j$. Again, we use here that the formula $\triangle(\triangle A_j \to A_j) \to \triangle A_j$ (provable[2] in L_1) is valid in every node of K.

If h is any of the functions $n, f_1, \dots, f_k, g_1, \dots, g_\ell$ then $w \le h(w)$ for every w. If $h \ne n$ and $h(w) \le v$ then $h(v) = v$, moreover $n(n(w)) = n(w)$. Therefore the closure of the set $\{\langle X_0 \rangle\}$ under the functions $n, f_1, \dots, f_k, g_1, \dots, g_\ell$, denoted by W', is a *finite* set. Put $D' = D \cap W'$ and let $<'$ and N' be the restrictions of $<$ and N on W'. If A is a propositional atom (arithmetical or general) and $w \in W'$, define $w \Vdash' A$ iff $w \Vdash A$, and extend the definition of $\Vdash'$ inductively so that $\mathbf{W}' = \langle W', <', D', N', \Vdash' \rangle$ is a model.

The relation $<'$ is a finite tree, any finite strict partial order is converse well-founded and $d <' N'(d) \notin D'$ for any $d \in D'$, therefore $\mathbf{W}'$ is a finite tree-like L_1-model with $\mathrm{rng}(N') \cap D' = \emptyset$. Moreover $\langle X_0 \rangle \in W'$ and $\langle X_0 \rangle \nVdash \varphi$, thus to complete the proof of the proposition it suffices to show that $w \Vdash A \iff w \Vdash' A$ for any $w \in W'$ and A a subformula of φ, which follows by induction on the complexity of A:

The assertion holds for atoms by definition. The induction steps for Boolean connectives and $^{\mathrm{N}}$ are straightforward as N' coincides with N on W'.

If $w \Vdash \triangle A$ and $v \in D'$, $w <' v$, then $w < v$ and $v \in D$, thus $v \Vdash A$ and $v \Vdash' A$ by the induction hypothesis, therefore $w \Vdash' \triangle A$. Suppose that $w \nVdash \triangle A$. We have $A = A_j$ for some $j = 1, \dots, \ell$. By the definition of g_j we know that $w < g_j(w)$ and $g_j(w) \nVdash A$. As W' is closed under g_j we get $w <' g_j(w) \in D'$, thus $g_j(w) \nVdash' A$ and $w \nVdash' \triangle A$.

The induction step for $\Box$ is similar. ∎

Our next task is the model completeness of the second auxiliary system, L_3.

Definition 2.2.5 Let $\mathbf{W} = \langle W, <, D, N, \Vdash \rangle$ be a model. We say that two elements $d, d' \in D$ are *arithmetically isomorphic,* written as $d \simeq d'$, if $<''\{d\} = <''\{d'\}$, $<^{-1\prime\prime}\{d\} = <^{-1\prime\prime}\{d'\}$ and $d \Vdash p_i \iff d' \Vdash p_i$ for every *arithmetical* atom p_i, i.e. $d \simeq d'$ iff d and d' have the same successors and predecessors and they agree on satisfaction of arithmetical atoms.

$\mathbf{W}$ is called an *L_3-model,* if it is an L_1-model and

$$\forall d \in D \;\, \forall w > d \;\, \exists d' \in D \;\, (d \simeq d' \,\&\, N(d') = w).$$

An L_1-model $\mathbf{W}$ is *injective,* provided that

$$\forall d, d' \in D \;\, (d \simeq d' \,\&\, N(d) = N(d') \;\Rightarrow\; d = d'),$$

i.e. the function N is injective on any equivalence class of $\simeq$.

[2] $\triangle(\triangle A \to A) \to \Box\triangle(\triangle A \to A) \to \Box(\triangle\triangle A \to \triangle A) \to \Box(\Box\triangle A \to \triangle A) \to \Box\triangle A \to \triangle\triangle A$ by C4, C2, C5 and C3, also $\triangle(\triangle A \to A) \to (\triangle\triangle A \to \triangle A)$ by C2, hence $\triangle(\triangle A \to A) \to \triangle A$.

Definition 2.2.6 The symbol $A \subseteq B$ abbreviates "A is a subformula of B". Let $\boxdot\varphi$ be the formula $\varphi \mathbin{\&} \Box\varphi$. For any a.m.f. φ the symbol U_φ denotes the formula

$$\bigwedge_{i,j} \boxdot(\triangle(\alpha_i \to \beta_j^{\mathrm{N}}) \to \triangle(\alpha_i \to \Box\beta_j)),$$

where the α_i's are all Boolean combinations of arithmetical subformulas $\psi \subseteq \varphi$ and β_j's are all Boolean combinations of all formulas ψ such that $\psi^{\mathrm{N}} \subseteq \varphi$ (there are only finitely many such things modulo logical equivalence).

The symbol R_φ denotes the formula

$$\bigwedge\{\triangle(\Box\psi \to \psi);\ \Box\psi \subseteq \varphi \text{ or } \psi \text{ is a Bool. comb. of some } \chi \text{ such that } \chi^{\mathrm{N}} \subseteq \varphi\}.$$

The formula R_φ has nothing to do with L_3, but we state the definition here because we will need some information on it, which is conveniently proved as a part of the following theorem.

Proposition 2.2.7 *Let φ be an arithmetical modal formula.*

(i) *The following conditions are equivalent:*

- (a) $L_3 \vdash \varphi$
- (b) *φ is valid in every L_3-model*
- (c) *φ is valid in every finite injective L_3-model*
- (d) $L_1 \vdash U_\varphi \to \varphi$

(ii) $L_3 \vdash R_\varphi \to \varphi$ *iff* $L_1 \vdash R_\varphi \mathbin{\&} U_\varphi \to \varphi$.

Proof:
(i-b) $\to$ (i-c) is trivial. (i-d) $\to$ (i-a) and the right-to-left implication in (ii) are easy as $L_3 \vdash U_\varphi$ (in fact, U_φ is a conjunction of formulas of the shape $\boxdot\psi$, where ψ is an instance of D2). In order to prove (i-a) $\to$ (i-b) it suffices to show that the axiom D2 is valid in all L_3-models. Let $\langle W, <, D, N, \Vdash\rangle$ be such a model and suppose $w \Vdash \triangle(\psi \to \chi^{\mathrm{N}})$, we want to derive $w \Vdash \triangle(\psi \to \Box\chi)$. Let $d > w$, $d \in D$ such that $d \Vdash \psi$ and let $u > d$, we have to show $u \Vdash \chi$. By the definition of an L_3-model there is $d' \in D$, $d \simeq d'$ such that $N(d') = u$. An easy induction shows that isomorphic nodes agree on satisfaction of all a.m.f., not necessarily atomic, hence $d' \Vdash \psi$. Moreover $d' > w$ (as $d > w$ and $d \simeq d'$), thus $d' \Vdash \chi^{\mathrm{N}}$, which means that $u = N(d') \Vdash \chi$.

The implication (i-c) $\to$ (i-d): suppose that $L_1 \nvdash U_\varphi \to \varphi$, we have to find a finite injective L_3-model in which φ is not valid. By the Proposition 2.2.4 there is a finite tree-like L_1-model $\mathbf{W_0} = \langle W_0, <, D, N, \Vdash\rangle$ and $x_0 \in W_0$ such that $x_0 \nVdash \varphi$ and $x_0 \Vdash U_\varphi$, we may assume w.l.o.g. that x_0 is the root of W_0 (i.e. the least element). In L_1 one easily derives $U_\varphi \to \Box U_\varphi$, which is a general property of all formulas starting with $\boxdot$. Therefore U_φ holds in every node of W_0.

For any L_1-model $\mathbf{W} = \langle W, <, D, N, \Vdash\rangle$ define

$$\mathit{diff}(\mathbf{W}) = \{d \in D;\ \exists w > d\ \sim\exists d' \in D\ \ (d \simeq d' \mathbin{\&} N(d') = w)\},$$

$$Diff(\mathbf{W}) = \{w \in W;\ \exists d \geq w\ \ d \in diff(\mathbf{W})\}.$$

Note that $\mathbf{W}$ is an L_3-model iff $Diff(\mathbf{W}) = \emptyset$. The model $\mathbf{W_0}$ is a finite injective L_1-model with the least element $x_0 \notin D$, U_φ is valid in $\mathbf{W}$, $x_0 \nVdash \varphi$ and $<$ restricted to $Diff(\mathbf{W_0})$ is a tree. Therefore there exists a model $\mathbf{W} = \langle W, <, D, N, \Vdash \rangle$ with all these properties, which has a minimal possible cardinality of $Diff(\mathbf{W})$. It suffices to show that $Diff(\mathbf{W}) = \emptyset$.

Suppose that $Diff(\mathbf{W})$ is non-empty, we will construct a model $\mathbf{W}'$ with all the required properties such that $|Diff(\mathbf{W}')| < |Diff(\mathbf{W})|$, which yields a contradiction.

Pick a maximal element $x \in Diff(\mathbf{W})$. Clearly $x \in diff(\mathbf{W})$, in particular $x \in D$. Let $y_0, \dots, y_k$ be the list of all nodes $y > x$ such that there is no $x' \simeq x$, $N(x') = y$. Choose pairwise distinct objects $z_0, \dots, z_k$ not belonging to W. Define

$$W' = W \cup \{z_i;\ \ i \leq k\},$$

$$D' = D \cup \{z_i;\ \ i \leq k\},$$

$$N' \supseteq N, \quad N'(z_i) = y_i,\ \ i \leq k,$$

$$<' = < \cup \{\langle u, z_i \rangle;\ \ i \leq k, u < x\} \cup \{\langle z_i, u \rangle;\ \ i \leq k, x < u\}.$$

x is not the least element of W, because $x \in D$. The restriction of $<$ to $Diff(\mathbf{W})$ is a tree and $Diff(\mathbf{W})$ is a downward-closed set, therefore $x \in Diff(\mathbf{W})$ has an immediate predecessor in $<$, say r (i.e. $r < x$ and for every $u < x$, either $u = r$ or $u < r$). Put

$$\alpha = \bigwedge\{\psi;\ \ x \Vdash \psi, \psi \subseteq \varphi\} \mathbin{\&} \bigwedge\{\sim\!\psi;\ \ x \nVdash \psi, \psi \subseteq \varphi\},$$

$$\beta_i = \bigvee\{\sim\!\psi;\ \ y_i \Vdash \psi, \psi^{\mathrm{N}} \subseteq \varphi\} \vee \bigvee\{\psi;\ \ y_i \nVdash \psi, \psi^{\mathrm{N}} \subseteq \varphi\}$$

for every $i \leq k$. Then $x \nVdash \alpha \to \Box\beta_i$, thus $r \nVdash \triangle(\alpha \to \Box\beta_i)$ and $r \nVdash \triangle(\alpha \to \beta_i^{\mathrm{N}})$ (because $r \Vdash U_\varphi$). Hence there exist $w_i \in D$, $w_i > r$ such that $w_i \Vdash \alpha$ and $w_i \nVdash \beta_i^{\mathrm{N}}$. It holds then $x \Vdash \psi \ \Leftrightarrow\ w_i \Vdash \psi$ for every $\psi \subseteq \varphi$, and also $y_i \Vdash \psi \ \Leftrightarrow\ w_i \Vdash \psi^{\mathrm{N}}$ for every $\psi^{\mathrm{N}} \subseteq \varphi$.

We define

$$z_i \Vdash' p \ \Leftrightarrow\ x \Vdash p \qquad \text{and} \qquad z_i \Vdash' q \ \Leftrightarrow\ w_i \Vdash q$$

for every *arithmetical* atom p and *general* atom q, where $i \leq k$. We leave the forcing of all atoms in the nodes of W unchanged and extend the definition of $\Vdash'$ to all formulas so that $\mathbf{W}' = \langle W', <', D', N', \Vdash' \rangle$ is a model.

The relation $<'$ is a finite strict partial order (hence it is converse well-founded) and $d < N'(d)$ for every $d \in D'$, thus $\mathbf{W}'$ is a finite L_1-model. The least element x_0 of W is also the least element of W' and $x_0 \notin D'$.

The relation $\simeq$ on elements of D is unchanged in W' and all new nodes z_i are arithmetically isomorphic to each other and to x. From this it follows easily that the model $\mathbf{W}'$ is injective.

We claim that $Diff(\mathbf{W}') \subseteq Diff(\mathbf{W}) \setminus \{x\} \subsetneq Diff(\mathbf{W})$. The set $A = Diff(\mathbf{W}) \setminus \{x\}$ is downward-closed, hence it suffices to show that $diff(\mathbf{W}') \subseteq A$. If $u \in W \setminus Diff(\mathbf{W})$, $u \in D'$ and $u <' v$, then $v \in W$, $u < v$ and $u \in D$, thus there exists $u' \in D$ such that $u \simeq u'$ and $N(u') = v$. This remains true in $\mathbf{W}'$, therefore $u \notin \mathit{diff}(\mathbf{W}')$. If $u = x$ or $u \in W' \setminus W$ (i.e. $u = z_i$ for some $i \leq k$) and $u <' y$, then $y \in K$ and $x < y$. Either there exists $x' \in D$ such that $x' \simeq x$ (thus $x' \simeq u$) and $N(x') = y$ (thus $N'(x') = y$), or $y = y_j$ for some $j \leq k$. But then $u \simeq z_j \in D'$ and $N'(z_j) = y$. Hence $u \notin \mathit{diff}(\mathbf{W}')$.

In particular, $<'$ restricted to $\mathit{Diff}(\mathbf{W}')$ is a tree.

Sublemma 1 *Let A be a Boolean combination of some subformulas of φ, $i \leq k$ and $u \in W$, where either $u \in D$ or A is an a.m.f. Then*

$$z_i \Vdash' A \;\Leftrightarrow\; w_i \Vdash A,$$

$$u \Vdash' A \;\Leftrightarrow\; u \Vdash A.$$

Proof:

By induction on the complexity of the formula A.

If $A = p$ is an arithmetical atom, we have $z_i \Vdash' p \;\Leftrightarrow\; x \Vdash p \;\Leftrightarrow\; w_i \Vdash p$. The other cases for A an atom are trivial.

The induction steps for Boolean connectives are straightforward.

Let $A = \psi^{\mathrm{N}}$: if $u \in D$ we have $u \Vdash' \psi^{\mathrm{N}} \;\Leftrightarrow\; N(u) \Vdash' \psi \;\Leftrightarrow\; N(u) \Vdash \psi \;\Leftrightarrow\; u \Vdash \psi^{\mathrm{N}}$. Also $z_i \Vdash' \psi^{\mathrm{N}} \;\Leftrightarrow\; y_i \Vdash' \psi \;\Leftrightarrow\; y_i \Vdash \psi \;\Leftrightarrow\; w_i \Vdash \psi^{\mathrm{N}}$.

The induction step for $\triangle A$: we will treat at first the case $u \in W$. If $u \Vdash' \triangle A$ then $u \Vdash \triangle A$ due to the induction hypothesis and the relations $< \subseteq <'$ and $D \subseteq D'$. If $u \nVdash' \triangle A$, there exists $v >' u$ such that $v \in D'$ and $v \nVdash' A$. If $v \in W$ it follows that $u < v$, $v \in D$ and $v \nVdash A$, hence $u \nVdash \triangle A$. In the case $v = z_i$ we have $u < x$, thus $u \leq r$ and $r < w_i$, therefore $u < w_i$. By the induction hypothesis $w_i \nVdash A$, thus $u \nVdash \triangle A$, since $w_i \in D$.

The remaining case is $u = z_i \in W' \setminus W$. We have $z_i \Vdash' \triangle A \;\Leftrightarrow\; x \Vdash' \triangle A \;\Leftrightarrow\; x \Vdash \triangle A \;\Leftrightarrow\; w_i \Vdash \triangle A$: the first equivalence is due to $x \simeq z_i$, the second follows from the previous paragraph and the third by $\triangle A \subseteq \varphi$.

The induction step for $\Box\psi$ is similar. ⊣⊢

Using this sublemma we get immediately $x_0 \nVdash' \varphi$. It remains to check $x_0 \Vdash' U_\varphi$.

Suppose that $u \in W'$ and $u \Vdash' \triangle(\alpha \to \beta^{\mathrm{N}})$, where α is a Boolean combination of some $\psi \subseteq \varphi$ and β is a Boolean combination of formulas ψ such that $\psi^{\mathrm{N}} \subseteq \varphi$. Put $\tilde{u} = u$ for $u \in W$ and $\tilde{u} = x$ otherwise. If $v > \tilde{u}$, $v \in D$ and $v \Vdash \alpha$, then $v \Vdash' \alpha$, hence $v \Vdash' \beta^{\mathrm{N}}$, thus $N(v) \Vdash' \beta$ and $N(v) \Vdash \beta$. In other words $\tilde{u} \Vdash \triangle(\alpha \to \beta^{\mathrm{N}})$, therefore $\tilde{u} \Vdash \triangle(\alpha \to \Box\beta)$.

Let $u <' v \in D'$, $v \Vdash' \alpha$. Then $\tilde{v} > \tilde{u}$, $\tilde{v} \in D$ and $\tilde{v} \Vdash' \alpha$, since $v \simeq \tilde{v}$. Hence $\tilde{v} \Vdash \alpha$, thus $\tilde{v} \Vdash \Box\beta$. If $v <' w$ then $\tilde{v} < \tilde{w}$, hence $\tilde{w} \Vdash \beta$, $\tilde{w} \Vdash' \beta$ and $w \Vdash' \beta$, because $w \simeq \tilde{w}$. Therefore $v \Vdash' \Box\beta$ and $u \Vdash' \triangle(\alpha \to \Box\beta)$.

This completes the proof of $(i\text{-}c) \to (i\text{-}d)$. We have to show yet the left-to-right implication of (ii), which will be done by a modification of the preceding argument. Suppose that $L_1 \nvdash R_\varphi \,\&\, U_\varphi \to \varphi$. By 2.2.4 there is a finite tree-like L_1-model $\mathbf{W_0}$ whose root satisfies U_φ and R_φ and does not satisfy φ. We pick a model $\mathbf{W}$ with the minimal cardinality of Diff with all the properties as above with the extra condition that R_φ is valid in all nodes of the model. Again we show that this model has empty *Diff* by *reductio ad absurdum.* The only difference is that the newly constructed model $\mathbf{W}'$ should satisfy R_φ, provided that $\mathbf{W}$ does, which is done as follows:

Given a $u \in D'$ such that $u \Vdash' \Box\psi$, where $\Box\psi \subseteq \varphi$ or ψ is a Boolean combination of formulas χ such that $\chi^{\mathrm{N}} \subseteq \varphi$, and given a $v > \tilde{u}$ we have $v \Vdash' \psi$, thus $v \Vdash \psi$ and $\tilde{u} \Vdash \Box\psi$. But $\tilde{u} \in D$, hence $\tilde{u} \Vdash \psi$, therefore $\tilde{u} \Vdash' \psi$ and $u \Vdash' \psi$, because $\tilde{u} \simeq u$. ⊣

Remark 2.2.8 In contrast to L_1, the definition of an L_3-model depends not only on the underlying frame, but also on the satisfaction relation (which is used for the definition of $\simeq$). In fact, L_3 is not frame-complete. It is easy to see that L_3 corresponds to the class of all L_1-frames $\langle W, <, D, N\rangle$ such that $\forall x\,\forall y\,\forall z\ \ (x < y < z \,\&\, y \in D \Rightarrow N(y) = z)$, i.e. every non-minimal node from D has precisely one successor. Every such frame validates the formulas $\triangle(\Box\varphi \leftrightarrow \varphi^{\mathrm{N}})$, $\triangle(\Box\varphi \vee \Box{\sim}\varphi)$, $\triangle\Box\Box\bot$, which are not derivable in L_3.

In the case of CSRL the situation is even worse: the condition imposed to the model will depend on the formula we want to "disvalidate" by the model. The proof of 2.2.4 shows that any extension of L_1 closed under MP, Nec and substitution is complete w.r.t. a suitable class of models and this applies to CSRL too, but we will need also the Finite Model Property in the proof of the arithmetical completeness of CSRL. To see that CSRL is not complete w.r.t. a class of finite models, note that CSRL proves $\triangle{\sim}\Box^k\bot$ for every $k \in \omega$, thus in every finite model $\langle W, <, D, N, \Vdash\rangle$ satisfying all theorems of CSRL the set D has to be empty, i.e. all such models validate the formula $\triangle\bot$.

The system CSRL$^{\#}$ is not complete w.r.t. any class of models (even infinite), since it is not closed under Nec.

We postpone the definiton of the Kripke semantics and the Kripke completeness theorem for CSRL to the next section, since we will prove it together with the arithmetical completeness theorem. (It is possible to derive the Kripke completeness directly by examination of the models, but it is rather inconvenient and lengthy.)

2.3 Arithmetical completeness

Lemma 2.3.1 *Let φ be an a.m.f. and $* = \langle {}^*, {}_*\rangle$ an arithmetical realization.*

(i) CSRL $\vdash \varphi \ \Rightarrow\ \mathbf{PA} \vdash \varphi^*$,

(ii) CSRL$^{\#} \vdash \varphi \ \Rightarrow\ \mathbb{N} \models \varphi^*$.

Briefly, CSRL $\subseteq \mathrm{PRL}_{ext}(\mathbf{AST}, \mathbf{PA})$ *and* CSRL$^{\#} \subseteq \mathrm{PRL}^{+}_{ext}(\mathbf{AST}, \mathbf{PA})$.

Proof:
By induction on the length of the derivation of φ. The axioms A1 and A2 and Modus Ponens are sound since **PA** and **AST** contain the CPC. The axioms B1, B2 and B3 translate to the assertion that the interpretation $^{\mathbf{N}}$ commutes with the propositional connectives and this is clearly provable in **AST**. The axioms C1, C2

and the Necessitation Rule correspond to the Löb's derivability conditions. C3 is a formalization of the Löb's theorem, C4 follows from formalization of the Σ^0_1-completeness of **PA**. C5 says that $^{\mathbf{FN}}$ is an interpretation of **PA** in **AST**, which is formalizable in **PA**. Finally D1, D2 and D3 correspond to 1.3.7 and the additional axiom S of CSRL$^{\#}$ expresses the arithmetical soundness of **AST** (1.3.6). ⫨

Definition 2.3.2 Let B be a g.m.f. For every $K \in \omega$ we define an a.m.f. B^K by $B^0 = \triangle B$, $B^{K+1} = \triangle(B \vee B^K)$, i.e. $B^K = \triangle(\underbrace{B \vee \triangle(B \vee \cdots \vee \triangle(B \vee \triangle B)}_{K+1} \cdots))$.

Lemma 2.3.3 *Let φ be an a.m.f. and $K \in \omega$. Then* CSRL$^{\#} \vdash \varphi^K \to \varphi$.

Proof:
By induction on K. If $K = 0$, $\varphi^0 \to \varphi$ is $\triangle\varphi \to \varphi$, i.e. an axiom of CSRL$^{\#}$. Suppose $K > 0$. Then φ^K is $\triangle(\varphi \vee \varphi^{K-1})$, thus $\varphi^K \to \varphi \vee \varphi^{K-1}$ is an axiom and $\varphi^{K-1} \to \varphi$ is provable by the induction hypothesis, therefore $\varphi^K \to \varphi$ is also provable. ⫨

Definition 2.3.4 Let $\mathbf{W} = \langle W, <, D, N, \Vdash \rangle$ be an L_1-model, let φ be an a.m.f. We say that $\mathbf{W}$ is a φ-CSRL-*model,* if it is an L_3-model, $d \Vdash \Box\psi \to \psi$ for every $d \in D$ and ψ such that $\Box\psi \subseteq \varphi$, and for every $d \in D$ there is a $w > d$ such that $d \Vdash \psi \;\Leftrightarrow\; w \Vdash \psi$ for all ψ such that $\psi^{\mathrm{N}} \subseteq \varphi$.

A node d is φ-*reflexive* if $d \in D$ and for every $\triangle A \subseteq \varphi$ such that $d \Vdash \triangle A$ and for every $d' \in D$, $d' \simeq d$ we have $d' \Vdash A$.

Let $d, d' \in D$. We define d, d' to be *equivalent* (or φ-*equivalent*), written as $d \equiv d'$ ($d \equiv_\varphi d'$), if $d \Vdash A \;\Leftrightarrow\; d' \Vdash A$ for every g.m.f. A (or every $A \subseteq \varphi$). The model $\mathbf{W}$ is *balanced* if

$$\forall d, d' \in D \;\; (d \simeq d' \;\&\; N(d) \in D \;\&\; N(d') \in D \;\&\; N(d) \simeq N(d') \;\Rightarrow\; d \equiv d').$$

The model is φ-*nice* if for every $d \in D$ there is $w > d$, $w \notin D$ such that for every $d' \in D$ satisfying $N(d') = d$ there is $d'' \in D$ such that $d'' \simeq d'$, $N(d'') = w$ and $d'' \equiv_\varphi d'$.

Observation 2.3.5 Let $\mathbf{W}$ be an L_3-model and φ an a.m.f. Then $\mathbf{W}$ is a φ-CSRL-model iff the formula R_φ (defined in 2.2.6) is valid in $\mathbf{W}$.

Lemma 2.3.6 *Let φ be an a.m.f. and $K \in \omega$. Assume that there is a φ-*CSRL-*model $\mathbf{W}$ and a φ-reflexive node $x \in W$ such that $x \nVdash \varphi$. Then there exists a φ^K-*CSRL-*model $\mathbf{W}'$ and $x' \in W'$ such that $x' \nVdash \varphi^K$.*

Proof:
Let $\mathbf{W_0} = \langle W_0, <_0, D_0, N_0, \Vdash \rangle$ and $x_0 \in D_0$ be as in the hypothesis. We may assume w.l.o.g. that for every $w \in W_0$ either $w >_0 x_0$ or $w \in D_0$ and $w \simeq x_0$. By the definition of a φ-CSRL-model there exists $x' \simeq x_0$ such that for all $\psi^{\mathrm{N}} \subseteq \varphi$ the equivalence $x' \Vdash \psi \;\Leftrightarrow\; N_0(x') \Vdash \psi$ holds, we may assume that x_0 has this property (as $x' \nVdash \varphi$).

For all $x \in W_0$ we pick pairwise distinct objects $\bar{x}$ not belonging to W_0. Then we find for every $x \in W_0$ a node $\tilde{x} \in D_0$ such that the following holds: if $x >_0 x_0$ then $\tilde{x} \simeq x_0$ and $N_0(\tilde{x}) = x$, otherwise (i.e. if $x \simeq x_0$) $\tilde{x} = x_0$. We define a new model $\mathbf{W_1}$ by putting

$$W_1 = W_0 \cup \{\bar{x};\ \ x \in W_0\},$$

$$D_1 = D_0 \cup \{\bar{x};\ \ x \in W_0\},$$

$$<_1 = <_0 \cup \{\langle \bar{x}, y\rangle;\ \ x, y \in W_0\},$$

$$N_1 \text{ extends } N_0,\ \ N_1(\bar{x}) = x.$$

We leave the satisfaction of all formulas in nodes of W_0 unchanged and define

$$\bar{x} \Vdash p \ \ \Leftrightarrow\ \ x_0 \Vdash p,$$

$$\bar{x} \Vdash q \ \ \Leftrightarrow\ \ \tilde{x} \Vdash q$$

for every arithmetical atom p, general atom q and $x \in W_0$. A straightforward induction on the complexity shows that for every $A \subseteq \varphi$ we have

$$\bar{x} \Vdash A \ \ \Leftrightarrow\ \ \tilde{x} \Vdash A.$$

It follows easily that $\mathbf{W_1}$ is a φ-CSRL-model and all $\bar{x}$'s are φ-reflexive.

By repeating this construction $(K+1)$-times we get a φ-CSRL-model $\mathbf{W_{K+1}}$ containing a sequence of nodes $x_{K+1} < \cdots < x_1 < x_0$, $x_i \in D_{K+1}$ such that $x_i \nVdash \varphi$. Then $x_{K+1} \nVdash \varphi^K$ and $\mathbf{W_{K+1}}$ is a φ^K-CSRL-model, since φ and φ^K contain the same subformulas of the form $\Box\psi$ or ψ^{N}. ⊣⊢

Definition 2.3.7 $K(\varphi)$ denotes the cardinality of the set $\{A;\ \ \triangle A \subseteq \varphi\}$.

Lemma 2.3.8 *Let φ be an a.m.f. and $\tilde{\varphi} = \varphi(q_1/B_1, \ldots, q_n/B_n)$ a substitutional instance of φ which does not contain any general atom. Suppose that there is a finite injective $\tilde{\varphi}$-CSRL-model $\mathbf{W}$ such that $\tilde{\varphi}$ is not valid in $\mathbf{W}$. Then there exists a finite injective balanced φ-nice φ-CSRL-model $\mathbf{W'}$ such that φ is not valid in $\mathbf{W'}$.*

If moreover $\mathbf{W}$ does not satisfy $\tilde{\varphi}^{K(\tilde{\varphi})}$ then there is a φ-reflexive node $x \in W'$ such that $x \nVdash \varphi$.

Proof:
Let $\mathbf{W} = \langle W, <, D, N, \Vdash\rangle$ be a finite injective $\tilde{\varphi}$-CSRL-model, $x \in W$ and $x \nVdash \tilde{\varphi}$. If A is a g.m.f. define $\tilde{A} = A(q_1/B_1, \ldots, q_n/B_n, q_{n+1}/\bot, \ldots, q_m/\bot)$, where $q_{n+1}, \ldots, q_m$ are all general atoms occurring in A, distinct from $q_1, \ldots, q_n$. We define the model $\mathbf{W'} = \langle W, <, D, N, \Vdash'\rangle$ by putting

$$w \Vdash' A \ \ \Leftrightarrow\ \ w \Vdash \tilde{A}$$

for every $w \in W$ and every formula A.

$\mathbf{W'}$ is easily seen to be a finite injective L_3-model. The properties of a $\tilde{\varphi}$-CSRL-model together with the fact that satisfaction of any formula A in $\mathbf{W'}$ is determined by satisfaction of the general-atom-free formula $\tilde{A}$ in $\mathbf{W}$ imply that $\mathbf{W'}$ is a balanced φ-nice φ-CSRL-model and clearly $x \nVdash' \varphi$.

Put $K = K(\tilde{\varphi})$ and assume the extra condition $x \nVdash \tilde{\varphi}^K$. This implies that there are nodes $x_K > \cdots > x_0 > x$ such that $x_i \in D$ and $x_i \nVdash \tilde{\varphi}$. If $\psi, A \subseteq \varphi$ then $\mathbf{W}'$ satisfies the formulas $\psi \leftrightarrow \tilde{\psi}$ and $\triangle(A \leftrightarrow \tilde{A})$. Let $\triangle A \subseteq \varphi$. Then $\tilde{A}$ is equivalent to a formula

$$\bigwedge_{i<n}(\bigvee_{j<h_i} \varepsilon_i^j \psi_i^j \vee \bigvee_{j<g_i} \zeta_i^j {\chi_i^j}^{\mathrm{N}}),$$

where ε_i^j and ζ_i^j stand for either '$\sim$' or nothing and $\psi_i^j, {\chi_i^j}^{\mathrm{N}} \subseteq \tilde{A}$. Put

$$\bar{A} = \bigwedge_{i<n}(\bigvee_{j<h_i} \varepsilon_i^j \psi_i^j \vee \Box(\bigvee_{j<g_i} \zeta_i^j \chi_i^j)).$$

Then $L_3 \vdash \triangle\tilde{A} \leftrightarrow \triangle\bar{A}$ and $L_3 \vdash \triangle(\bar{A} \to \tilde{A})$.

A formula $\triangle B \to B$ has to be valid in all but possibly one node of the linear chain $x_0 < \cdots < x_K$. There are at most K formulas $\bar{A}$ such that $\triangle A \subseteq \varphi$, hence by the pigeon-hole principle there is $i \leq K$ such that $x_i \Vdash \bigwedge_{\triangle A \subseteq \varphi} (\triangle\bar{A} \to \bar{A})$. We have $x_i \nVdash' \varphi$ (since $x_i \nVdash \tilde{\varphi}$) and $x_i \in D$. We claim that x_i is a φ-reflexive node (in $\mathbf{W}'$): given $\triangle A \subseteq \varphi$ such that $x_i \Vdash' \triangle A$ and $d \simeq x_i$ we have $x_i \Vdash \triangle\tilde{A}$, $x_i \Vdash \triangle\bar{A}$ and $x_i \Vdash \bar{A}$, thus $d \Vdash \bar{A}$ (as $\bar{A}$ is an a.m.f.), hence $d \Vdash \tilde{A}$ and $d \Vdash' A$. ⊣⊢

Theorem 2.3.9 *Let φ be an arithmetical modal formula.*

(i) *The following are equivalent:*

- (a) $\mathbf{PA} \vdash \varphi^*$ *for every arithmetical realization* $*$, *i.e.* $\varphi \in \mathrm{PRL}_{ext}(\mathbf{AST}, \mathbf{PA})$
- (b) $\mathrm{CSRL} \vdash \varphi$
- (c) $L_3 \vdash R_\varphi \to \varphi$
- (d) $L_1 \vdash R_\varphi \,\&\, U_\varphi \to \varphi$
- (e) φ *is valid in all* φ-CSRL-*models*
- (f) φ *is valid in all finite injective balanced* φ-*nice* φ-CSRL-*models*

(ii) *The following are equivalent:*

- (a) $\mathbb{N} \models \varphi^*$ *for every arithmetical realization* $*$, *i.e.* $\varphi \in \mathrm{PRL}^+_{ext}(\mathbf{AST}, \mathbf{PA})$
- (b) $\mathrm{CSRL}^{\#} \vdash \varphi$
- (c) $\mathrm{CSRL} \vdash \varphi^{K(\varphi)}$
- (d) φ *is valid in all* φ-*reflexive nodes of all* φ-CSRL-*models*
- (e) φ *is valid in all* φ-*reflexive nodes of all finite injective balanced* φ-*nice* φ-CSRL-*models*

(iii) *If* $\mathrm{CSRL} \nvdash \varphi$ *then there exists a substitutional instance*

$$\tilde{\varphi} = \varphi(q_1/B_1, \ldots, q_m/B_m)$$

of φ *such that* $\tilde{\varphi}$ *contains no general atoms,* B_i *are Boolean combinations of some* p_j *and* p_k^N *and* $\mathrm{CSRL} \nvdash \tilde{\varphi}$.

Proof:
We will consider two more statements:

(*i-g*) φ is valid in all finite injective φ-CSRL-models

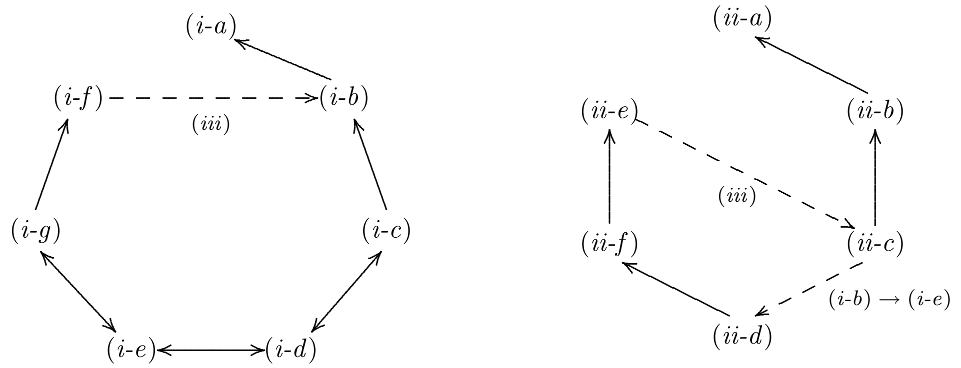
(*ii-f*) φ is valid in all φ-reflexive nodes of all finite injective φ-CSRL-models

The mutual equivalence of (*i-c*), (*i-d*), (*i-e*) and (*i-g*) follows from 2.2.7 and 2.3.5. The implication (*i-g*) $\rightarrow$ (*i-f*) is trivial, (*i-c*) $\rightarrow$ (*i-b*) is obvious as R_φ is a conjunction of axioms of CSRL and (*i-b*) $\rightarrow$ (*i-a*) follows from 2.3.1.

The implications (*ii-d*) $\rightarrow$ (*ii-f*) $\rightarrow$ (*ii-e*) are trivial, (*ii-b*) $\rightarrow$ (*ii-a*) follows from 2.3.1 and (*ii-c*) $\rightarrow$ (*ii-b*) from 2.3.3.

Assuming for the moment that (*iii*) is valid, we get (*i-f*) $\rightarrow$ (*i-b*) and (*ii-e*) $\rightarrow$ (*ii-c*) from 2.3.8 (and from (*i-c*) $\rightarrow$ (*i-b*)). Moreover assuming the implication (*i-b*) $\rightarrow$ (*i-e*) holds, we get (*ii-c*) $\rightarrow$ (*ii-d*) from 2.3.6.

The situation is shown in the following figure. The full arrows denote so far established implications, the dashed ones denote conditional implications depending on the assumption written at side:



It thus suffices to prove (*i-a*) $\rightarrow$ (*i-g*), (*ii-a*) $\rightarrow$ (*ii-f*) and (*iii*).

If (*i-g*) fails, there is a finite injective φ-CSRL-model $\mathbf{W} = \langle W, \prec, D, N, \Vdash \rangle$ and $x \in W$ such that $x \nVdash \varphi$. We may assume w.l.o.g. that $W = \{1, \ldots, n\}$, $x = 1$, $1 \notin D$ and 1 is the least element of W. Similarly if (*ii-f*) fails we find a finite injective φ-CSRL-model $\mathbf{W}$ such that $W = \{1, \ldots, n\}$, $1 \in D$, $1 \nVdash \varphi$, 1 is φ-reflexive and for all $w \in W$ either $w \simeq 1$ or $1 \prec w$. Moreover there is $x \simeq 1$ such that $\forall \psi^{\mathrm{N}} \subseteq \varphi \ (x \Vdash \psi \ \Leftrightarrow \ N(x) \Vdash \psi)$, we may assume that $x = 1$. In the sequel (#) will mark parts of the proof special to the implication (*ii-a*) $\rightarrow$ (*ii-f*).

Put $W' = W \cup \{0\}$ and define $0 \prec i$ for all $i \in W$. If $i \in D$, we pick $\hat{\imath} \succ i$ such that $\forall \psi^{\mathrm{N}} \subseteq \varphi \ (i \Vdash \psi \ \Leftrightarrow \ \hat{\imath} \Vdash \psi)$ and we find $\tilde{\imath} \in D$, $i \simeq \tilde{\imath}$ such that $N(\tilde{\imath}) = \hat{\imath}$. We arrange the choice of $\hat{\imath}$ so that $\hat{\imath} = \hat{\jmath}$ whenever $i \simeq j$. In the case of (#) we put $\tilde{1} = 1$.

Define $E = \{\tilde{\imath};\ i \in D\}$. The symbol $\vdash \cdots$ will abbreviate $\mathbf{PA} \vdash \cdots$. We define arithmetical sentences λ_i, $i \in W'$ by self-reference:

$$\vdash \lambda_i \leftrightarrow \exists x \ \forall y \geq x \ \ h(y) = \bar{\imath},$$

where $h(u) = v$ denotes a natural Σ^0_1-formula defining in **PA** the graph of the following primitive recursive function:

$$h(0) = 0,$$

$$h(x+1) = \begin{cases} i, & \text{if } i \succ h(x),\ \mathrm{Prf}_{\mathbf{PA}}(x, \ulcorner{\sim}\lambda_i\urcorner), \\ i, & \text{if } i \in E,\ i \succ h(x),\ \mathrm{Prf}_{\mathbf{AST}}(x, \ulcorner{\sim}\bigvee_{j \simeq i} \lambda_j^{\mathbf{FN}}\urcorner), \\ h(x) & \text{otherwise.} \end{cases}$$

(We assume that no number simultaneously codes a proof in **AST** and **PA**.)

We denote by $\approx$ the smallest equivalence relation on W' containing $\simeq$ and, if (#), also the pair $\langle 0, 1\rangle$. For every $i \in W$ we define an arithmetical sentence $\varkappa_i$ by

$$\varkappa_i = \bigvee_{j \approx i} \lambda_j.$$

For every $i \in D$ we define a sentence S_i of the language of **AST** by

$$S_i = \begin{cases} \varkappa_i^{\mathbf{FN}} \,\&\, \lambda^{\mathbf{N}}_{N(i)}, & i \notin E, \\ \varkappa_i^{\mathbf{FN}} \,\&\, (\varkappa_i^{\mathbf{N}} \vee \lambda^{\mathbf{N}}_{N(i)}), & i \in E. \end{cases}$$

Sublemma 1 *Let $i, j \in W$.*

(i) $i \in D \;\Rightarrow\; \mathbf{AST} \vdash S_i \to \varkappa_i^{\mathbf{FN}}$

(ii) $i \in D \setminus E \;\Rightarrow\; \mathbf{AST} \vdash S_i \to \varkappa^{\mathbf{N}}_{N(i)}, \qquad i \in E \;\Rightarrow\; \mathbf{AST} \vdash S_i \to \varkappa_i^{\mathbf{N}} \vee \varkappa^{\mathbf{N}}_{N(i)}$

(iii) $\vdash \lambda_0 \vee \varkappa_1 \vee \cdots \vee \varkappa_n$

(iv) $i \not\approx j \;\Rightarrow\; \vdash \varkappa_i \to {\sim}\varkappa_j, \qquad i, j \in D,\ i \neq j \;\Rightarrow\; \mathbf{AST} \vdash S_i \to {\sim}S_j$

(v) $i \prec j \;\Rightarrow\; \vdash \varkappa_i \to {\sim}\mathrm{Pr}_{\mathbf{PA}}(\ulcorner{\sim}\varkappa_j\urcorner), \quad i \prec j \in D \;\Rightarrow\; \vdash \varkappa_i \to {\sim}\mathrm{Pr}_{\mathbf{AST}}(\ulcorner{\sim}S_j\urcorner)$

(vi) $\vdash \varkappa_i \to \mathrm{Pr}_{\mathbf{PA}}(\ulcorner \bigvee_{j \succeq i} \varkappa_j\urcorner), \qquad i \notin D \;\Rightarrow\; \vdash \varkappa_i \to \mathrm{Pr}_{\mathbf{PA}}(\ulcorner \bigvee_{j \succ i} \varkappa_j\urcorner)$

(vii) $i \not\approx 0 \;\Rightarrow\; \vdash \varkappa_i \to \mathrm{Pr}_{\mathbf{AST}}(\ulcorner \bigvee_{\substack{j \in D \\ j \succ i}} S_j\urcorner)$

($viii$) $\nvdash {\sim}\varkappa_1$

(ix) *Assuming (#),* $\mathbf{AST} \vdash \bigvee_{j \in D} S_j$ *and* $\mathbb{N} \models \varkappa_1$

Proof:

(i) and (ii) follow immediately from the definition. The function h is monotonous and this formalizes in **PA** easily, yielding $\vdash \lambda_0 \vee \cdots \vee \lambda_n$ and this implies (iii).

Part (iv): it is clear from the definition that $\vdash \lambda_i \to {\sim}\lambda_j$ for any $i \neq j$. If $i, j \in W$, $i \not\approx j$, we have $i' \neq j'$ for every $i' \approx i$ and $j' \approx j$, hence $\vdash \varkappa_i \to {\sim}\varkappa_j$. Assume that $i, j \in D$, $i \neq j$. We distinguish two subcases. At first suppose $i \not\approx j$, then $\mathbf{AST} \vdash \varkappa_i^{\mathbf{FN}} \to {\sim}\varkappa_j^{\mathbf{FN}}$, thus $\mathbf{AST} \vdash S_i \to {\sim}S_j$ by part (i). In the other case, $i \approx j$, we have $i \simeq j$ and consequently $N(i) \neq N(j)$ (as the model **W** is injective), and either i or j does *not* belong to E (because of our choice of $\tilde{i}$). We may suppose w.l.o.g. $i \notin E$. Then we have $\mathbf{AST} \vdash \lambda^{\mathbf{N}}_{N(i)} \to \sim(\varkappa^{\mathbf{N}}_j \vee \lambda^{\mathbf{N}}_{N(j)})$ (since $k \prec N(i)$ for all $k \approx j$), therefore also $\mathbf{AST} \vdash S_i \to \sim S_j$.

Part (v): assuming $i \prec j$ the definition of h implies $\vdash \mathrm{Pr}_{\mathbf{PA}}(\ulcorner \sim\lambda_j \urcorner) \to \sim\lambda_i$. If $i, j \in K$, $i \prec j$ and $i' \approx i$, then $i' \prec j$, thus $\vdash \mathrm{Pr}_{\mathbf{PA}}(\ulcorner \sim\varkappa_j \urcorner) \to \mathrm{Pr}_{\mathbf{PA}}(\ulcorner \sim\lambda_j \urcorner) \to \bigwedge\limits_{i' \approx i} \sim\lambda_{i'} \to \sim\varkappa_i$.

One can show the same way that for any $i \prec j \in D$,

$$\vdash \mathrm{Pr}_{\mathbf{AST}}(\ulcorner \sim \bigvee_{k \simeq j} \lambda^{\mathbf{FN}}_k \urcorner) \to \sim\varkappa_i.$$

To prove (v) it therefore suffices to check that for every $i \in D$,

$$\vdash \mathrm{Pr}_{\mathbf{AST}}(\ulcorner \sim S_i \urcorner) \to \mathrm{Pr}_{\mathbf{AST}}(\ulcorner \sim \bigvee_{j \simeq i} \lambda^{\mathbf{FN}}_j \urcorner).$$

Clearly $\vdash \mathrm{Pr}_{\mathbf{AST}}(\ulcorner \sim S_i \urcorner) \to \mathrm{Pr}_{\mathbf{AST}}(\ulcorner \bigvee\limits_{j \simeq i} \lambda^{\mathbf{FN}}_j \to \sim\lambda^{\mathbf{N}}_{N(i)} \urcorner)$, hence by 1.3.7 also

$$\vdash \mathrm{Pr}_{\mathbf{AST}}(\ulcorner \sim S_i \urcorner) \to \mathrm{Pr}_{\mathbf{AST}}(\ulcorner \bigvee_{j \simeq i} \lambda^{\mathbf{FN}}_j \to \mathrm{Pr}^{\mathbf{FN}}_{\mathbf{PA}}(\ulcorner \sim\lambda_{N(i)} \urcorner) \urcorner).$$

But if $j \simeq i$ then $j \prec N(i)$, thus $\vdash \mathrm{Pr}_{\mathbf{PA}}(\ulcorner \sim\lambda_{N(i)} \urcorner) \to \sim\lambda_j$. This all together implies

$$\vdash \mathrm{Pr}_{\mathbf{AST}}(\ulcorner \sim S_i \urcorner) \to \mathrm{Pr}_{\mathbf{AST}}(\ulcorner \bigvee_{j \simeq i} \lambda^{\mathbf{FN}}_j \to \sim \bigvee_{j \simeq i} \lambda^{\mathbf{FN}}_j \urcorner)$$

and the desired assertion follows.

Part (vi): the function h is provably monotonous and the formula $\exists x\, h(x) = \bar{i}$ is a Σ-sentence, therefore

$$\vdash \lambda_i \to \exists x\, h(x) = \bar{i} \to \mathrm{Pr}_{\mathbf{PA}}(\ulcorner \exists x\, h(x) = \bar{i} \urcorner) \to \mathrm{Pr}_{\mathbf{PA}}(\ulcorner \bigvee_{j \succeq i} \lambda_j \urcorner).$$

If $i' \approx i$ and $i' \preceq j'$, there exists $j \approx j'$ such that $i \preceq j$. Hence $\vdash \varkappa_i \to \mathrm{Pr}_{\mathbf{PA}}(\ulcorner \bigvee\limits_{j \succeq i} \varkappa_j \urcorner)$.

If $i \notin E$, $i \neq 0$, then clearly $\vdash \lambda_i \to \mathrm{Pr}_{\mathbf{PA}}(\ulcorner \sim\lambda_i \urcorner)$. For any $i \in W \setminus D$ we have $\varkappa_i = \lambda_i$ and $i \notin E$, thus $\vdash \varkappa_i \to \mathrm{Pr}_{\mathbf{PA}}(\ulcorner \sim\varkappa_i \urcorner)$ and $\vdash \varkappa_i \to \mathrm{Pr}_{\mathbf{PA}}(\ulcorner \bigvee\limits_{j \succ i} \varkappa_j \urcorner)$.

Part (vii): the definition of h implies that for any $i \in W$,

$$\vdash \lambda_i \to \mathrm{Pr}_{\mathbf{AST}}(\ulcorner \sim\lambda^{\mathbf{FN}}_i \urcorner).$$

By the previous paragraph

$$\vdash \lambda_i \to \mathrm{Pr}_{\mathbf{AST}}(\ulcorner \lambda^{\mathbf{FN}}_i \vee \bigvee_{\substack{j \in E \\ j \succ i}} \lambda^{\mathbf{FN}}_j \vee \bigvee_{\substack{j \notin E \\ j \succ i}} (\lambda^{\mathbf{FN}}_j \,\&\, \mathrm{Pr}^{\mathbf{FN}}_{\mathbf{PA}}(\ulcorner \sim\lambda_j \urcorner)) \urcorner),$$

hence (using 1.3.7) $\vdash \lambda_i \to \mathrm{Pr}_{\mathbf{AST}}(\ulcorner \bigvee\limits_{\substack{j \in E \\ j \succ i}} \lambda^{\mathbf{FN}}_i \urcorner)$. The formula $\exists x\, h(x) = \bar{i}$ is a Σ-sentence and h is monotonous, thus $\mathbf{AST} \vdash \lambda^{\mathbf{FN}}_i \to \bigvee\limits_{j \succeq i} \lambda^{\mathbf{N}}_j$. Consequently

$$\vdash \lambda_i \to \mathrm{Pr}_{\mathbf{AST}}(\ulcorner \bigvee_{\substack{j \in E \\ j \succ i}} (\lambda^{\mathbf{FN}}_j \,\&\, \lambda^{\mathbf{N}}_j) \vee \bigvee_{\substack{j \in E \\ j \succ i}} \bigvee_{k \succ j} (\lambda^{\mathbf{FN}}_j \,\&\, \lambda^{\mathbf{N}}_k) \urcorner).$$

If $k \succ j \succ i$, $j \in E$, there exists $j' \simeq j$ such that $N(j) = k$. This together with the definition of S_j implies $\vdash \lambda_i \to \mathrm{Pr}_{\mathbf{AST}}(\ulcorner \bigvee_{\substack{j \in D \\ j \succ i}} S_j \urcorner)$. If $0 \not\approx i$ then all $i' \approx i$ have the same successors as i, hence $\vdash \varkappa_i \to \mathrm{Pr}_{\mathbf{AST}}(\ulcorner \bigvee_{\substack{j \in D \\ j \succ i}} S_j \urcorner)$.

Part (*viii*): $\mathbb{N} \models \sim\lambda_i$ for every $i \in W$—we have $\vdash \lambda_i \to \mathrm{Pr}_{\mathbf{AST}}(\ulcorner \sim\lambda_i^{\mathbf{FN}} \urcorner)$ and $\mathbb{N} \models \mathrm{Pr}_{\mathbf{AST}}(\ulcorner \sim\lambda_i^{\mathbf{FN}} \urcorner) \to \sim\lambda_i$, since **AST** is an arithmetically sound theory. On the other hand $\vdash \lambda_0 \vee \cdots \vee \lambda_n$, thus $\mathbb{N} \models \lambda_0$. But $\vdash \lambda_0 \to \sim \mathrm{Pr}_{\mathbf{PA}}(\ulcorner \sim\lambda_1 \urcorner) \to \sim \mathrm{Pr}_{\mathbf{PA}}(\ulcorner \sim\varkappa_1 \urcorner)$, therefore $\mathbb{N} \models \sim \mathrm{Pr}_{\mathbf{PA}}(\ulcorner \sim\varkappa_1 \urcorner)$ and $\nvdash \sim\varkappa_1$.

Part (*ix*): if (#) then $0 \approx 1$, thus $\mathbb{N} \models \varkappa_1$ (since $\mathbb{N} \models \lambda_0$ by the previous paragraph). As in the proof of the part (*vii*) one checks easily that $\mathbf{AST} \vdash \lambda_0^{\mathbf{FN}} \vee \bigvee_{i \in E} \lambda_i^{\mathbf{FN}}$ and consequently

$$\mathbf{AST} \vdash \bigvee_{i \in E} (\lambda_i^{\mathbf{FN}} \,\&\, \lambda_i^{\mathbf{N}}) \vee \bigvee_{i \in D} (\lambda_i^{\mathbf{FN}} \,\&\, \lambda_{N(i)}^{\mathbf{N}}) \vee (\lambda_0^{\mathbf{FN}} \,\&\, \bigvee_{i \approx 0} \lambda_i^{\mathbf{N}}) \vee (\lambda_0^{\mathbf{FN}} \,\&\, \bigvee_{i \simeq 1} \lambda_{N(i)}^{\mathbf{N}}),$$

hence

$$\mathbf{AST} \vdash \bigvee_{i \in E} S_i \vee \bigvee_{i \in D} S_i \vee S_1 \vee \bigvee_{i \simeq 1} S_i,$$

in other words $\mathbf{AST} \vdash \bigvee_{i \in D} S_i$. ⫲

We define a provability interpretation $* = \langle {}^*, {}_* \rangle$ by putting

$$p^* = \bigvee_{i \Vdash p} \varkappa_i,$$

$$q_* = \bigvee_{\substack{i \in D \\ i \Vdash q}} S_i,$$

for every arithmetical atom p and general atom q.

Sublemma 2 *Let $i \in W$ and $\psi, A \subseteq \varphi$.*

(*i*) $i \Vdash \psi \;\Rightarrow\; \vdash \varkappa_i \to \psi^*$,
$i \nVdash \psi \;\Rightarrow\; \vdash \varkappa_i \to \sim\psi^*$.

(*ii*) $i \Vdash A \;\Rightarrow\; \mathbf{AST} \vdash S_i \to A_*$,
$i \nVdash A \;\Rightarrow\; \mathbf{AST} \vdash S_i \to \sim A_*$, *provided $i \in D$.*

Proof:

Proceed by induction on the complexity of the formulas ψ, A.

Let $\psi = p$ be an atom. If $i \Vdash p$ then $\vdash \varkappa_i \to p^*$ by the definition. If $i \nVdash p$ and $i' \in W$, $i' \approx i$, then $i' \nVdash p$, hence $\vdash \varkappa_i \to \bigwedge_{j \Vdash p} \sim\varkappa_j \to \sim p^*$ by (*iv*) of the first Sublemma.

The case when $A = q$ is a general atom is treated similarly.

If $\psi = \bot$ we have $i \nVdash \bot$ and $\vdash \varkappa_i \to \sim\bot$.

Let $\psi = (\psi_1 \to \psi_2)$. If $i \Vdash \psi$ then $i \nVdash \psi_1$ or $i \Vdash \psi_2$. By the induction hypothesis $\vdash \varkappa_i \to \sim\psi_1^*$ or $\vdash \varkappa_i \to \psi_2^*$, thus $\vdash \varkappa_i \to (\psi_1^* \to \psi_2^*)$. If $i \nVdash \psi$ then $i \Vdash \psi_1$ and $i \nVdash \psi_2$, hence $\vdash \varkappa_i \to \psi_1^*$ and $\vdash \varkappa_i \to \sim\psi_2^*$, therefore $\vdash \varkappa_i \to \sim(\psi_1^* \to \psi_2^*)$.

A similar argument applies if $A = (A_1 \to A_2)$.

Let $\psi = \Box\chi$. If $i \nVdash \psi$, there exists $j \succ i$, $j \nVdash \chi$. Then $\vdash \chi^* \to \sim\varkappa_j$ by I.H., hence $\vdash \varkappa_i \to \sim \mathrm{Pr}_{\mathbf{PA}}(\ulcorner \sim\varkappa_j \urcorner) \to \sim \mathrm{Pr}_{\mathbf{PA}}(\ulcorner \chi^* \urcorner)$ by (v) of the Sublemma. On the other hand, assume $i \Vdash \psi$. Then $j \Vdash \chi$ for every $j \succ i$, thus $\vdash \bigvee_{j \succ i} \varkappa_j \to \chi^*$. If $i \notin D$, we get immediately $\vdash \varkappa_i \to \mathrm{Pr}_{\mathbf{PA}}(\ulcorner \bigvee_{j \succ i} \varkappa_j \urcorner) \to \mathrm{Pr}_{\mathbf{PA}}(\ulcorner \chi^* \urcorner)$ by (vi) of the Sublemma. If $i \in D$ then also $i \Vdash \chi$ (since $\Box\chi \subseteq \varphi$ and $\mathbf{W}$ is a φ-CSRL-model), hence $\vdash \bigvee_{j \succeq i} \varkappa_j \to \chi^*$. Therefore $\vdash \varkappa_i \to \mathrm{Pr}_{\mathbf{PA}}(\ulcorner \bigvee_{j \succeq i} \varkappa_j \urcorner) \to \mathrm{Pr}_{\mathbf{PA}}(\ulcorner \chi^* \urcorner)$ by (vi) again.

Let $\psi = \triangle A$. If $i \nVdash \psi$, there exists $j \succ i$, $j \in D$ such that $j \nVdash A$. Thus $\mathbf{AST} \vdash A_* \to \sim S_j$ and $\vdash \varkappa_i \to \sim \mathrm{Pr}_{\mathbf{AST}}(\ulcorner \sim S_j \urcorner) \to \sim \mathrm{Pr}_{\mathbf{AST}}(\ulcorner A_* \urcorner)$ by (v). Let $i \Vdash \psi$. Assume at first $i \not\approx 0$. Then $j \Vdash A$ for every $j \succ i$, $j \in D$, thus $\mathbf{AST} \vdash \bigvee_{\substack{j \in D \\ j \succ i}} S_j \to A_*$ and $\vdash \varkappa_i \to \mathrm{Pr}_{\mathbf{AST}}(\ulcorner \bigvee_{\substack{j \in D \\ j \succ i}} S_j \urcorner) \to \mathrm{Pr}_{\mathbf{AST}}(\ulcorner A_* \urcorner)$ by (vii). Now assume $i \approx 0$, then $(\#)$, $i \simeq 1$ and the node $1 \in D$ is φ-reflexive. We have $1 \Vdash \triangle A$ and $\triangle A \subseteq \varphi$, therefore $j \Vdash A$ for all $j \simeq 1$, moreover $j \Vdash A$ for all $j \in D$, $j \succ 1$. In other words $j \Vdash A$ for every $j \in D$. Hence $\mathbf{AST} \vdash \bigvee_{j \in D} S_j \to A_*$. But $\mathbf{AST} \vdash \bigvee_{j \in D} S_j$ by (ix), thus $\mathbf{AST} \vdash A_*$ and *a fortiori* $\vdash \varkappa_i \to \mathrm{Pr}_{\mathbf{AST}}(\ulcorner A_* \urcorner)$.

Let $A = \psi$ be an a.m.f. If $i \Vdash \psi$, we have $\vdash \varkappa_i \to \psi^*$. But $\mathbf{AST} \vdash S_i \to \varkappa_i^{\mathbf{FN}}$ by (i), hence $\mathbf{AST} \vdash S_i \to \psi^{*\mathbf{FN}}$. The case $i \nVdash \psi$ is similar.

Let $A = \psi^{\mathrm{N}}$. If $i \Vdash \psi^{\mathrm{N}}$ and $i \notin E$, we have $\vdash \varkappa_{N(i)} \to \psi^*$ and $\mathbf{AST} \vdash S_i \to \varkappa_{N(i)}^{\mathbf{N}}$ by (ii), hence $\mathbf{AST} \vdash S_i \to \psi^{*\mathbf{N}}$. If $i \in E$ then also $i \Vdash \psi$ since $\psi^{\mathrm{N}} \subseteq \varphi$, therefore $\vdash \varkappa_i \vee \varkappa_{N(i)} \to \psi^*$. Moreover $\mathbf{AST} \vdash S_i \to \varkappa_i^{\mathbf{N}} \vee \varkappa_{N(i)}^{\mathbf{N}}$ by (ii), hence $\mathbf{AST} \vdash S_i \to \psi^{*\mathbf{N}}$. The situation when $i \nVdash \psi^{\mathrm{N}}$ is analogous again. ⫶⊪

We have $1 \nVdash \varphi$, thus $\mathbf{PA} \vdash \varkappa_1 \to \sim\varphi^*$. But $\mathbf{PA} \nvdash \sim\varkappa_1$, therefore $\mathbf{PA} \nvdash \varphi^*$. In the case of $(\#)$ we have $\mathbb{N} \vDash \varkappa_1$, thus $\mathbb{N} \nvDash \varphi^*$. This completes the proof of $(i\text{-}a) \to (i\text{-}g)$ and $(ii\text{-}a) \to (ii\text{-}f)$.

We have to prove the part (iii) yet. Assuming CSRL $\nvdash \varphi$ there is a finite injective φ-CSRL-countermodel to φ. By the previous part of the proof we find an arithmetical realization $*$ such that $\mathbf{PA} \nvdash \varphi^*$. Moreover the $*$ we have constructed assigns to every general atom a Boolean combination of formulas of the form $\psi^{\mathbf{FN}}$ and $\chi^{\mathbf{N}}$, where ψ and χ are arithmetical sentences. Thus there is a substitutional instance $\tilde{\varphi}$ of φ not containing general atoms and an arithmetical realization $\#$ such that $\mathbf{PA} \nvdash \tilde{\varphi}^{\#}$. This implies CSRL $\nvdash \tilde{\varphi}$ by $(i\text{-}b) \to (i\text{-}a)$. ⫶⊪

Proposition 2.3.10 *The systems L_1, L_3, CSRL and CSRL$^{\#}$ are decidable.*

Proof:
All these logics are Σ^0_1 as they are recursively axiomatized. Moreover each of them has a suitable Kripke semantics enjoying the Finite Model Property, hence they are Π^0_1 and consequently they are recursive, by the Post theorem. ⫶⊪

Remark 2.3.11 This statement may be a bit improved. All the systems above contain in a suitable sense the usual unimodal provability logic GL ([Sol76]) which is known to be $PSPACE$-complete, this gives an estimate from below to the complexity of these logics. It is possible to characterize L_1 by a reasonable sequent calculus enjoying the Cut Elimination, which gives a decision procedure for L_1 working in polynomial space, thus L_1 is $PSPACE$-complete too. It follows easily from 2.2.7 and 2.3.9 that CSRL and CSRL$^{\#}$ are linear–time reducible to each other, CSRL and CSRL$^{\#}$ are exponential–time reducible to both L_3 and L_1 and finally L_3 is exponential–time reducible to L_1, in particular CSRL, CSRL$^{\#}$ and L_3 are decidable in exponential space (more precisely they are in $SPACE(2^{\mathcal{O}(n)})$). It is an open problem if any of these three systems is in $PSPACE$ or EXP. (I conjecture a negative answer, at least for $PSPACE$.)

Remark 2.3.12 The arithmetical completeness of CSRL and the Gödel's Diagonal Lemma imply that CSRL is closed under the Diagonalization Rules (DiR)

$$\boxdot(p \leftrightarrow \psi(p)) \to \varphi \qquad / \qquad \varphi,$$
$$\triangle(q \leftrightarrow B(q)) \to \varphi \qquad / \qquad \varphi,$$

where the atom p (resp. q) does not occur in φ and every its occurence in ψ (resp. B) is in the scope of a box (resp. triangle). This can be alternatively established by a purely syntactic argument—as in the case of GL, the logic L_1 has *unique definable fixpoints:* there is an a.m.f. χ (resp. a g.m.f. C) not containing p (resp. q) such that L_1 proves

$$\boxdot(p \leftrightarrow \psi(p)) \leftrightarrow \boxdot(p \leftrightarrow \chi),$$
$$\triangle(q \leftrightarrow B(q)) \leftrightarrow \triangle(q \leftrightarrow C).$$

Example 2.3.13 Let φ be an arithmetical sentence. We know that **FN** and **N** are models of arithmetic in **AST**, therefore we may ask what relation there is between $\varphi^{\mathbf{FN}}$ and $\varphi^{\mathbf{N}}$. We will try to find out, whether

- **AST** proves $\varphi^{\mathbf{FN}} \leftrightarrow \varphi^{\mathbf{N}}$,
- **AST** proves $\varphi^{\mathbf{FN}} \to \varphi^{\mathbf{N}}$.

There is a simple answer to the first question: it holds if and only if the formula φ is decidable in **PA**. Clearly, if $\mathbf{PA} \vdash \varphi$ or $\mathbf{PA} \vdash {\sim}\varphi$ then (provably in **AST**) either φ or ${\sim}\varphi$ holds simultaneously in **FN** and **N**, since both are models of **PA**. On the other hand, suppose that $\mathbf{AST} \vdash \varphi^{\mathbf{FN}} \leftrightarrow \varphi^{\mathbf{N}}$. Using twice the axiom D2 we see that CSRL proves

$$\triangle(p \leftrightarrow p^{\mathrm{N}}) \to \triangle(p \to \Box p) \,\&\, \triangle({\sim}p \to \Box{\sim}p),$$

hence also

$$\triangle(p \leftrightarrow p^{\mathrm{N}}) \to \triangle(\Box p \vee \Box{\sim}p).$$

Therefore $\triangle(p \leftrightarrow p^{\mathrm{N}}) \to (\Box p \vee \Box{\sim}p)$ is a valid principle of $\mathrm{PRL}^{+}_{ext}(\mathbf{AST}, \mathbf{PA})$, i.e. either φ or ${\sim}\varphi$ is provable in **PA**.

The second question is more complicated. The answer is positive if φ is a Σ^0_1-sentence, because **N** is an end-extension of **FN**. The same holds for formulas logically equivalent to a Σ^0_1-sentence, and this suggests that φ could be Σ^0_1 in a stronger theory, say **PA** or **AST**. Consider the following statements:

(i) $\mathbf{PA} \vdash \varphi \leftrightarrow \sigma$ for some $\sigma \in \Sigma^0_1$,

(ii) $\mathbf{AST} \vdash \varphi^{\mathbf{FN}} \to \varphi^{\mathbf{N}}$,

(iii) $\mathbf{AST} \vdash (\varphi \leftrightarrow \sigma)^{\mathbf{FN}}$ for some $\sigma \in \Sigma^0_1$.

We will show that (i) $\to$ (ii) $\to$ (iii), but neither of these two implications can be reversed. Note first that (ii) is equivalent to

(ii') $\mathbf{AST} \vdash \varphi^{\mathbf{FN}} \leftrightarrow \mathrm{Pr}^{\mathbf{FN}}_{\mathbf{PA}}(\ulcorner\varphi\urcorner)$,

because CSRL $\vdash \triangle(p \to p^{\mathrm{N}}) \leftrightarrow \triangle(p \leftrightarrow \Box p)$ (use D2 and D3 from left to right and D1 from right to left). Moreover (iii) is equivalent to

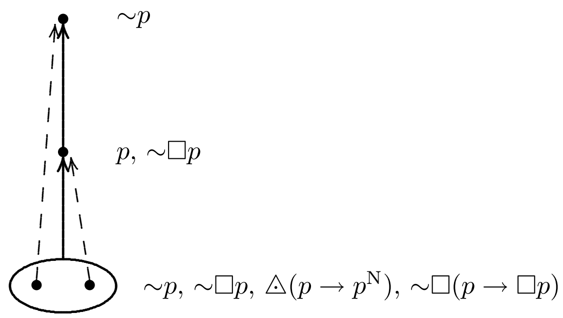
(iii') $\mathbf{AST} \vdash \varphi^{\mathbf{FN}} \leftrightarrow \mathrm{Pr}^{\mathbf{FN}}_{\mathbf{PA}}(\ulcorner\psi\urcorner)$ for some arithmetical sentence ψ.

The implication (iii') $\to$ (iii) is trivial as $\mathrm{Pr}_{\mathbf{PA}}(\ulcorner\psi\urcorner)$ is always Σ^0_1. On the other hand if $\sigma \in \Sigma^0_1$ then $\mathbf{AST} \vdash \sigma^{\mathbf{FN}} \leftrightarrow \mathrm{Pr}^{\mathbf{FN}}_{\mathbf{PA}}(\ulcorner\sigma\urcorner)$ by provable Σ^0_1-completeness and D3, hence $\mathbf{AST} \vdash (\varphi \leftrightarrow \sigma)^{\mathbf{FN}}$ implies $\mathbf{AST} \vdash \varphi^{\mathbf{FN}} \leftrightarrow \mathrm{Pr}^{\mathbf{FN}}_{\mathbf{PA}}(\ulcorner\sigma\urcorner)$.

If $\mathbf{PA} \vdash \varphi \leftrightarrow \sigma$, $\sigma \in \Sigma^0_1$, then $\mathbf{PA} \vdash \varphi \to \mathrm{Pr}_{\mathbf{PA}}(\ulcorner\varphi\urcorner)$ by provable Σ^0_1-completeness, hence (i) $\to$ (ii'). The implication (ii') $\to$ (iii') is trivial, thus (i) $\to$ (ii) and (ii) $\to$ (iii).

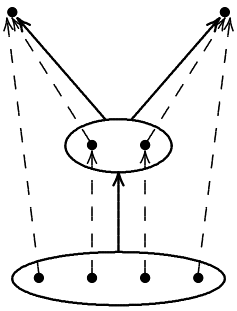
In order to demonstrate that (iii) does not in general imply (ii), or equivalently (iii') $\not\to$ (ii), it suffices to show that the formula $\alpha = \triangle(p \leftrightarrow \Box p') \to \triangle(p \to p^{\mathrm{N}})$ is not a valid principle of $\mathrm{PRL}^+_{ext}(\mathbf{AST}, \mathbf{PA})$, i.e. $\mathrm{CSRL}^{\#} \nvdash \alpha$. Similarly if ($ii$) $\to$ (i) then (ii) would imply $\mathbf{PA} \vdash \varphi \to \mathrm{Pr}_{\mathbf{PA}}(\ulcorner\varphi\urcorner)$, hence we need to show that $\mathrm{CSRL}^{\#} \nvdash \beta$, where β is the formula $\triangle(p \to p^{\mathrm{N}}) \to \Box(p \to \Box p)$.

The following figure shows a countermodel to β.



This represents a model $\mathbf{W} = \langle W, <, D, N, \Vdash\rangle$ as follows: bullets denote nodes of W, ovals embrace $\simeq$-equivalence classes of nodes from D, full arrows make up $<$, and dashed arrows the function N. The graph of $<$ is simplified—we omit arrows which follow by transitivity, and we treat all nodes in an oval as one, because arithmetically isomorphic nodes have the same successors and predecessors. The forcing of selected formulas, recorded at the side of the drawing, applies to the nearest node only, however all nodes of an oval force the same *arithmetical* formulas. Now, using the relevant definitions, it is easy to see that $\mathbf{W}$ is a β-CSRL-model, its bottom nodes are β-reflexive and do not force β, hence $\mathrm{CSRL}^{\#} \nvdash \beta$. (In fact, this model is also injective, balanced and β-nice.)

Here is a counterexample for α.



Again, it is an injective, balanced and α-nice α-CSRL-model and its bottom nodes, forcing $\sim\alpha$, are α-reflexive.

Remark 2.3.14 CSRL has some interesting subsystems obtained by restricting its language. Namely, we can form bimodal provability logics for **AST** and **PA** comparable to the bimodal logics for other pairs of theories. This is important as only a small number of them are understood so far, see [Car86], [Bek94], [Bek96], [JJ98].

Usual bimodal provability logics with a single type of atoms and formulas require both theories involved to have the same language, however we study here **PA**, having the arithmetical language, and **AST**, which has the set-theoretical language permitting (at least) two canonical interpretations of the language of **PA**. We resolve this problem by considering two "arithmetics", $\mathbf{AST^{FN}} = \{\varphi;\ \mathbf{AST} \vdash \varphi^{\mathbf{FN}}\}$ and $\mathbf{AST^{N}} = \{\varphi;\ \mathbf{AST} \vdash \varphi^{\mathbf{N}}\}$. In the fragment of CSRL without general atoms the provability predicates for **PA** and $\mathbf{AST^{FN}}$ are represented by the modal operators $\Box$ and $\triangle$, the provability predicate for $\mathbf{AST^{N}}$ may be represented by an additional modality $\triangledown$ which is introduced by putting $\triangledown\varphi = \triangle\varphi^{\mathrm{N}}$.

Note that **AST** is a conservative extension of **PA**, i.e. $\mathbf{AST^{N}} = \mathbf{PA}$ ($\triangle\psi^{\mathrm{N}} \to \triangle\Box\psi$ is an instance of D2, hence CSRL$^{\#}$ proves $\triangledown\psi \to \Box\psi$), but this fact is not formalizable in **PA** (or **AST**) itself (one can show that CSRL$^{\#}$ proves $\triangle(\triangledown\psi \to \Box\psi) \leftrightarrow \Box\psi$, thus **AST** can establish the conservativity just for sentences which are in fact theorems of **PA**). In other words the provability predicate of $\mathbf{AST^{N}}$ acts as an alternative numeration of **PA** which is not provably equivalent to the standard one. However the inclusions $\mathbf{PA} \subseteq \mathbf{AST^{N}} \subseteq \mathbf{AST^{FN}}$ are provable in **PA**.

It is possible to form three pairs of theories from **PA**, $\mathbf{AST^{FN}}$ and $\mathbf{AST^{N}}$, thus three bimodal provability logics arise here.

The logic PRL($\mathbf{PA}, \mathbf{AST^{FN}}$) is well-known. It is obtained from the minimal bimodal provability logic for extension of theories, CSM (see [Smo85]; it is given by axioms A1, C1–C5, MP and Nec), by adding the axiom schema

ER) $\triangle(\Box\varphi \to \varphi)$.

(ER stands for "essentially reflexive". Note that in our notation ER=D3.) Due to Carlson [Car86], this is true for every (locally) essentially reflexive pair of Σ^0_1-sound theories extending $I\Delta_0 + EXP$, the pair **PA**, $\mathbf{AST^{FN}}$ obviously meets this requirement.

The logic PRL($\mathbf{PA}, \mathbf{AST^N}$), with $\Box$ and $\triangledown$ as primitive operators, is axiomatizable by CSM (with $\triangledown$ instead of $\triangle$) plus the axiom schema

Q) $\triangledown\varphi \leftrightarrow \triangledown\Box\varphi$.

As in the case of CSM+(ER), this logic is maximal: $\mathrm{PRL}(T, S) = \mathrm{CSM} + (\mathrm{Q})$ whenever $\mathrm{PRL}(T, S) \supseteq \mathrm{CSM} + (\mathrm{Q})$, the theories T and S are Σ^0_1-sound and contain $I\Delta_0 + EXP$. (Moreover under these requirements the theories T and S actually coincide, but not provably so. We have already mentioned that **PA** and $\mathbf{AST^N}$ have this property.)

The logic PRL($\mathbf{AST^N}, \mathbf{AST^{FN}}$) is axiomatizable by CSM (with $\triangledown$ in the place of $\Box$) and the schema

Σ-C) $\triangle\sigma \to \triangledown\sigma$,

where σ is a disjunction of formulas starting with $\triangledown$ or $\triangle$, including the empty disjunction $\bot$. (Σ-C stands for "Σ^0_1-conservative".) This logic is not maximal, it is properly included in the trivial provability logic containing the axiom $\triangle\varphi \leftrightarrow \triangledown\varphi$ or in the Beklemishev's system CSM+(B_1-Cons) (this axiom looks like Σ-C but applies to all Boolean combinations of formulas starting with $\triangledown$ or $\triangle$, see [Bek96]).

It is also possible to form a trimodal provability logic PRL($\mathbf{PA}, \mathbf{AST^N}, \mathbf{AST^{FN}}$) with all the three operators $\Box$, $\triangledown$, $\triangle$. It turns out that $\triangledown$ is definable from the remaining modalities by

$$\triangledown\varphi \leftrightarrow \triangle\Box\varphi,$$

and it is easy to see that this schema together with CSM+(ER) axiomatizes the logic (because of the arithmetical completeness of CSM+(ER)).

The absolute (true) provability logics PRL^+($\mathbf{PA}, \mathbf{AST^{FN}}$), PRL^+($\mathbf{PA}, \mathbf{AST^N}$), PRL^+($\mathbf{AST^N}$, $\mathbf{AST^{FN}}$) and PRL^+($\mathbf{PA}, \mathbf{AST^N}, \mathbf{AST^{FN}}$) are axiomatized by the soundness schema

S) $\triangle\varphi \to \varphi$

(or $\triangledown\varphi \to \varphi$ in the case of PRL^+($\mathbf{PA}, \mathbf{AST^N}$)) over the set of all theorems of the corresponding PRL with Modus Ponens as the sole rule of inference.

Remark 2.3.15 The system L_1 satisfies the Craig's interpolation property. (This may be demonstrated e.g. by an easy induction on the length of a cut-free proof in the above mentioned sequent calculus.) If we restrict ourselves to formulas without general atoms, then systems L_3, CSRL and $\mathrm{CSRL}^{\#}$ have interpolation too, but this is no longer true when we drop this restriction. There is no logic between L_3 and $\mathrm{CSRL}^{\#}$ with the interpolation property, indeed the following theorem of L_3,

$$\triangle(q \leftrightarrow p_1) \to (\triangle(q \leftrightarrow p_2^{\mathrm{N}}) \to \triangle(\Box p_2 \vee \Box{\sim}p_2)),$$

does not have an interpolant in $\mathrm{CSRL}^{\#}$. (This was established by a model-theoretic argument.)

References

CMUC = Commentationes Mathematicae Universitatis Carolinae, Prague

[Bek94] L. D. Beklemishev, On bimodal logics of provability, *Annals of Pure and Applied Logic,* vol. 68 (1994), no. 2, pp. 115–159

[Bek96] L. D. Beklemishev, Bimodal logics for extensions of arithmetical theories, *Journal of Symbolic Logic,* vol. 61 (1996), no. 1, pp. 91–124

[Car86] T. Carlson, Modal logics with several operators and provability interpretations, *Israel Journal of Mathematics,* vol. 54 (1986), no. 1, pp. 14–24

[ČV86] K. Čuda, B. Vojtášková, Model–theoretical constructions in **AST**, *CMUC,* vol. 27 (1986), pp. 581–604

[JJ98] G. Japaridze, D. H. J. de Jongh, The Logic of Provability, in: S. R. Buss (ed.), *Handbook of Proof Theory,* Elsevier, North-Holland, Amsterdam 1998, pp. 475–546

[Mar89] C. Marchini, Formalization and models for the Alternative Set Theory, in [Proc89], pp. 21–23

[Proc89] J. Mlček, M. Benešová, B. Vojtášková (eds.), *Mathematics in the Alternative Set Theory,* Proceedings of the 1st Symposium on the Alternative Set Theory in Stará Lesná, Association of Slovak Mathematicians and Physicists, Bratislava 1989

[Resl79] M. Resl, On models in the Alternative Set Theory, *CMUC,* vol. 20 (1979), pp. 723–736

[RS81] M. Resl, A. Sochor, Provability in the Alternative Set Theory, *CMUC,* vol. 22 (1981), pp. 655–660

[Smo85] C. Smoryński, *Self-reference and modal logic,* Springer-Verlag, Berlin 1985

[Soch79] A. Sochor, Metamathematics of the alternative set theory I, *CMUC,* vol. 20 (1979), pp. 697–722

[Soch82] A. Sochor, Metamathematics of the alternative set theory II, *CMUC,* vol. 23 (1982), pp. 55–79

[Soch83] A. Sochor, Metamathematics of the alternative set theory III, *CMUC,* vol. 24 (1983), pp. 137–154

[Soch89] A. Sochor, Bases of alternative set theory, in [Proc89], pp. 43–70

[Sol76] R. M. Solovay, Provability interpretations of modal logic, *Israel Journal of Mathematics,* vol. 25 (1976), pp. 287–304

[SV80] A. Sochor, P. Vopěnka, Revealments, *CMUC,* vol. 21 (1980), pp. 97–118

[Tz86] A. Tzouvaras, Countable inductive definition in AST, *CMUC,* vol. 27 (1986), pp. 17–33

[Vop79] P. Vopěnka, *Mathematics in the Alternative Set Theory,* Teubner Texte, Leipzig 1979

[Vop89] P. Vopěnka, *Úvod do matematiky v alternatívnej teórii množín,* Alfa, Bratislava 1989 (in Slovak)

Index