SERGEI N. ARTEMOV & LEV D. BEKLEMISHEV

PROVABILITY LOGIC

1 INTRODUCTION

The idea of provability logic seems to originate in a short paper [Gödel, 1933]. K. Gödel was motivated by the question of providing Brouwer's intuitionistic logic, as formalized by Heyting, with an adequate semantics. According to Brouwer, intuitionistic truth means provability. Here is a summary from *Constructivism in Mathematics* ([Troelstra and van Dalen, 1988], p. 4):

> "A statement is *true* if we have a proof of it, and *false* if we can show that the assumption that there is a proof for the statement leads to a contradiction."

An axiom system for intuitionistic logic was introduced by Heyting in 1930; its full description may be found in fundamental monographs [Kleene, 1952; Troelstra and van Dalen, 1988]. In 1931–34 A. Heyting and A.N. Kolmogorov made Brouwer's definition of intuitionistic truth explicit, though informal, by introducing what is now known as the *Brouwer–Heyting–Kolmogorov (BHK) semantics* ([Heyting, 1931; Heyting, 1934; Kolmogoroff, 1932]). *BHK* semantics suggests that a formula is called true if it has a proof. Further, a proof of a compound statement is described in terms of proofs of its components:

- a proof of $A \wedge B$ consists of a proof of A and a proof of B;
- a proof of $A \vee B$ is given by presenting either a proof of A or a proof of B;
- a proof of $A \rightarrow B$ is a construction transforming proofs of A into proofs of B;
- falsehood $\perp$ is a proposition which has no proof, $\neg A$ is a shorthand for $A \rightarrow \perp$.

The *BHK* semantics is widely recognized as the intended semantics for intuitionistic logic. In [Gödel, 1933] an attempt was made to formalize the *BHK* semantics. K. Gödel introduced a modal calculus of classical provability (essentially equivalent to the Lewis modal system $\mathsf{S4}$) and defined the intuitionistic propositional logic IPC in this logic. Gödel's provability calculus is based on the classical propositional logic and has the modal axioms and rules

$\Box F \rightarrow F$,

D. Gabbay and F. Guenthner (eds.),
Handbook of Philosophical Logic, Volume 10, 1–??.

$\Box(F\to G)\to(\Box F\to\Box G)$,
$\Box F\to\Box\Box F$,
$\vdash F \ \Rightarrow \vdash \Box F$ (*necessitation* rule).

Gödel considered a translation $t(F)$ of an intuitionistic formula F into the classical modal language: "box each subformula of F" apparently regarding such a translation to be a fair formalization[1] of the Brouwer thesis

intuitionistic truth = provability.

Gödel established that

$$\mathsf{IPC}\vdash F \qquad \Rightarrow \qquad \mathsf{S4}\vdash t(F),$$

thus providing a reading of IPC-formulas as statements about classical provability. He conjectured that the converse ($\Leftarrow$) also held and concluded in 1938 (see [Gödel, 1995], p. 100–101): *Intuitionismus ist daraus ableitbar*[2]. The ($\Leftarrow$) conjecture was proved in [McKinsey and Tarski, 1948]. The ultimate goal, however, of defining IPC via classical proofs had not been achieved because S4 was left without an exact intended semantics of the provability operator $\Box$:

$$\mathsf{IPC}\hookrightarrow\mathsf{S4}\hookrightarrow\ldots\quad ?\quad \ldots\hookrightarrow \textit{CLASSICAL PROOFS.}$$

Here, *CLASSICAL PROOFS* refers to systems based on a proof predicate $\mathsf{Proof}(x,y)$ denoting "x is the code of a proof of the formula having a code y" for a classical first order theory containing Peano arithmetic PA. Gödel in [Gödel, 1933] identified a problem there and pointed out that a natural reading of $\Box F$ as the formal provability predicate $\mathsf{Provable}(F)=\exists x\,\mathsf{Proof}(x,F)$ did not work.

> Let $\bot$ be the boolean constant **false** and $\Box F$ be $\mathsf{Provable}(F)$. Then $\Box\bot\to\bot$ corresponds to the statement $\mathsf{Con}(\mathsf{PA})$ expressing consistency of PA. An S4-theorem $\Box(\Box\bot\to\bot)$ expresses the assertion that $\mathsf{Con}(\mathsf{PA})$ is provable in PA, which is false according to the second Gödel incompleteness theorem.

Thus, [Gödel, 1933] showed that S4 was a provability calculus without an exact provability semantics, whereas the interpretation of $\Box F=\mathsf{Provable}(F)$ was an exact provability semantics for modality without axiom system known. Gödel's paper left open two natural problems:

1. Find the modal logic of the formal provability predicate $\mathsf{Provable}(F)$.

[1] This translation appeared earlier in a paper by I.E. Orlov [Orlov, 1928], who applied it to a system different from S4.

[2] *Intuitionism is derivable from this.*

2. Find an exact provability semantics of S4 and thereby of IPC.

It was already clear that solutions to 1 and 2 led to essentially different models of Provability, each targeting its own set of applications. The two parts of the present paper — "Part I, Logic of Provability" (Sections 2–10) and "Part II, Logic of Proofs" (Sections 11–16) — roughly correspond to the developments around these two questions. Here in the Introduction we briefly review main achievements in both directions.

Logic of Provability

The first significant step towards a solution of Problem 1 was made by M.H. Löb [Löb, 1955] who formulated, on the basis of the previous work by D. Hilbert and P. Bernays from 1939 (see [Hilbert and Bernays, 1968]), a number of natural conditions[3] on the formal provability predicate (nowadays known as *Bernays–Löb derivability conditions*) and observed that these conditions were sufficient for the proof of Gödel's second incompleteness theorem. Moreover, under the same conditions he found an important strengthening of the Gödel theorem. He proved that the following is a valid principle of the logic of the formal provability predicate:

$$\Box(\Box F \to F) \to \Box F.$$

This powerful principle, taken together with the axioms and rules of the modal logic K4 turned out later to provide a complete axiomatization of the logic of formal provability. This system currenly bears the name GL for Gödel and Löb[4].

M.H. Löb's work, followed by significant advances in general understanding of formalization of metamathematics particularly in the hands of S. Feferman [Feferman, 1960], inspired S. Kripke, G. Boolos, D. de Jongh and others to look into the problem of exact axiomatization of the logic of provability. Independently, the same notion appeared in an algebraic context in the work of R. Magari and his school in Italy (see [Magari, 1975b]). A dramatic account of these early developments can be found in [Boolos and Sambin, 1991]. As an important early result on provability logic stands out a theorem by D. de Jongh, found independently by G. Sambin, who established that the system GL has the fixed point property (see [Smoryński, 1977b; Smoryński, 1985] and some details below).

[3] These conditions were essentially expressed by the last two axioms and the necessitation rule of the above mentioned system S4, in other words, by the modal logic K4. So, their validity must have been known to Gödel.

[4] This logic was alternatively denoted by G, L, K4.W, PrL. Neither Gödel nor Löb formulated the logic explicitly, though undeniably they established the validity of the underlying arithmetical principles. Presumably, it was T. Smiley in whose work on the foundations of ethics [Smiley, 1963] the axioms of GL appeared for the first time.

H. Friedman formulated the problem of decidability of the letterless fragment of provability logic as his Problem 35 in [Friedman, 1975a]. This question, which happened to be much easier than the general case, was immediately answered by a number of people including G. Boolos [Boolos, 1976], J. van Benthem, C. Bernardi and F. Montagna. A breakthrough came in 1976 when R. Solovay published a solution of the general problem showing that the system GL axiomatizes the provability logic for any sufficiently strong and sound formal theory [Solovay, 1976]. He also showed that the set of modal formulas expressing universally *true* principles of provability was axiomatized by a decidable extension of GL, which is usually denoted by S and is called the *truth provability logic*.

Solovay's results and his novel methods opened a new stage in the development of provability logic, with several groups of researchers, most notably in the USA (R. Solovay, G. Boolos, C. Smoryński), the Netherlands (D. de Jongh, A. Visser), Italy (R. Magari, F. Montagna, G. Sambin, L. Valentini), and USSR (S. Artëmov and his students), starting to work intensively in this area. Textbooks by G. Boolos [Boolos, 1979b] and C. Smoryński [Smoryński, 1985], the first of which appeared very early, played an important educational role.

The main thrust of the research effort went into the direction of generalizing Solovay's results to more expressive languages. Here we briefly mention some of the probems that received prominent attention. Most of them (though not all) are covered in greater detail below and roughly correspond to the sections in this paper.

First order provability logics. It was soon discovered that the first order version of GL is not arithmetically complete. G. Boolos formulated in his book the problem of axiomatizing the full first order provability logic. Improtant partial results in this direction were obtained by F. Montagna [Montagna, 1987a]. A final negative solution was given in the papers by S. Artëmov [Artemov, 1985a] and V. Vardanyan [Vardanyan, 1986]. In particular, V. Vardanyan showed that this logic is Π^0_2-complete, thus not effectively axiomatizable. Earlier S. Artëmov showed that the first order truth provability logic is not even arithmetical. Independently but somewhat later similar results were obtained by V. McGee in his Ph.D. Thesis, they were never published.[5] The later joint publication with G. Boolos [Boolos and McGee, 1987] contained a certain strengthening of Artëmov's theorem. Even more dramatically, [Artemov, 1986] showed that the first order provability logics are sensible to a particular formalization of the provability predicate and, thus, are not very robustly defined.

The material on first order provability logic is extensively covered in a

[5] We are grateful to A. Visser for providing us with this information and with a copy of V. McGee Thesis.

later textbook by G. Boolos [Boolos, 1993] and in survey [de Jongh and Japaridze, 1998], therefore we chose not to include any further details in the present survey.

Intuitionistic provability logic. The question of generalizing Solovay's results from classical theories to intuitionistic ones, such as Heyting arithmetic HA, proved to be remarkably difficult. This problem was taken up by A. Visser, D. de Jongh and their students. In [Visser, 1981] a number of nontrivial principles of the provability logic of HA were found. In [Visser, 1985] a characterization and a decision algorithm for the letterless fragment of the provability logic of HA were obtained, thus solving an intuitionistic analog of the Friedman's 35-th problem. Some significant further results were obtained in [Visser, 1985; Visser, 1994; Visser, 1999; Visser, 2002b; de Jongh and Visser, 1996; Iemhoff, 2001a; Iemhoff, 2001b; Iemhoff, 2001c] but the general problem of axiomatizing the provability logic of HA remains a major open question. It is consistent with our present knowledge, though in our opinion not very likely, that this logic is Π^0_2-complete. See below for an overview of related results.

Classification of provability logics. Solovay's theorems naturally led to the notion of *provability logic for a given theory T relative to a metatheory U*, which was suggested by S. Artëmov [Artemov, 1979; Artemov, 1980] and A. Visser [Visser, 1981]. This logic, denoted $\boldsymbol{PL}_T(U)$, is defined as the set of all propositional principles of provability in T that can be established by means of U. (Thus, the provability logic of T corresponds to $U = T$ and the truth provability logic corresponds to U being the set of all true sentences of arithmetic.) The problem of describing all possible modal logics of the form $\boldsymbol{PL}_T(U)$, where T and U range over extensions of Peano arithmetic, has become known as the Classification problem for provability logics. Partial results were obtained in [Artemov, 1980; Visser, 1984; Artemov, 1985b; Japaridze, 1986] who, in particular, discovered four main families of provability logics. The classification was completed by L. Beklemishev in [Beklemishev, 1989a] who showed that all relative provability logics occur in one of these four families.

The Classification can be extended to a broader class of theories. In fact, the same result holds for extensions of rather weak *elementary arithmetic* EA (see below). However, it remains an intriguing open question whether Solovay's theorems can be extended to bounded arithmetic theories, such as S^1_2 or S_2. Partial results were obtained in [Berarducci and Verbrugge, 1993].

Provability logics with additional operators. Theorems by Solovay have been generalized to various extensions of the propositional language by additional operators having arithmetical interpretation.

The most straightforward generalization is obtained by simultaneously considering several provability operators corresponding to different theories. Already in the simplest case of *bimodal provability logic*, the axiomatization of such logics turns out to be very difficult. The bimodal logics for many natural pairs of theories have been characterized in [Smoryński, 1985; Japaridze, 1986; Carlson, 1986; Beklemishev, 1994; Beklemishev, 1996]. However, the general classification problem for bimodal provability logics for pairs of r.e. extensions of PA remains a major open question.

There were also interesting bimodal logic studies of provability related concepts different from the standard provability predicates, such as Mostowski operator, Rosser, Feferman and Parikh provability (see [Smoryński, 1985; Visser, 1989; Shavrukov, 1991; Shavrukov, 1994; Lindström, 1996]). In a number of cases arithmetical completeness theorems à la Solovay have been obtained. These results have their origin in an important paper [Guaspari and Solovay, 1979] (see also [Smoryński, 1985]). They considered an extension of the propositional modal language by *witness comparison* operator allowing to formalize Rosser-style arguments. Similar logics have later been used in [de Jongh and Montagna, 1989; Carbone and Montagna, 1989; Carbone and Montagna, 1990] for, e.g., the study of the speed-up of proofs.

Interpretability and conservativity logics. A. Visser, following V. Švejdar, formulated another important extension of the language of provability logic. He introduced a binary modality $\varphi \rhd \psi$ to stand for the arithmetization of the statement "the theory $T + \varphi$ interprets $T + \psi$". Interpretations here are understood in the standard sense of Tarski. This new modality allows (in a classical logic context) to express provability $\Box\varphi$ by $\neg\varphi \rhd \bot$, and thus is more expressive than the ordinary $\Box$.

It turns out that the resulting *interpretability logic* substantially depends on the basis theory T. For two important classes of theories T this logic has been characterized. For finitely axiomatizable[6] theories such as $I\Sigma_1$ or ACA_0 this was done by A. Visser [Visser, 1990]. For essentially reflexive theories, such as Peano arithmetic PA, this was done independently by V. Shavrukov and A. Berarducci [Shavrukov, 1988; Berarducci, 1990]. These results substantially relied on a previous work of A. Visser, D. de Jongh and F. Veltman who, in particular, developed a suitable Kripke-style semantics for the interpretability logics.

These results remain, so far, the main successes in this area. A number of principal questions are still open. For example, interestingly enough, an axiomatization of the minimal interpretability logic, that is, of the set of interpretability principles that hold over all reasonable arithmetical theories is not known. A excellent survey of interpretability logic is given in [Visser,

[6] To be more precise, one also requires here that the theories are sufficiently strong and sequential.

1998], see also [de Jongh and Japaridze, 1998].

The $\rhd$ modality has a related *conservativity* interpretation, which leads to conservativity logics studied in [Hájek and Montagna, 1990; Hájek and Montagna, 1992; Ignatiev, 1991]. Logics of *interpolability* and of *tolerance* introduced by K. Ignatiev and G. Japaridze [Ignatiev, 1993b; Dzhaparidze, 1992; Dzhaparidze, 1993] have a related arithmetical interpretation, but a format different form that of interpretability logics. These developments fall outside the scope of the present paper, see [de Jongh and Japaridze, 1998] for an overview.

Magari algebras and propositional second order provability logic. An algebraic approach to provability logic was initiated by R. Magari and his students [Magari, 1975a; Magari, 1975b; Montagna, 1979; Montagna, 1980]. The *provability algebra* of a theory T, also called the *Magari algebra of* T, is defined as the set of T-sentences factorized modulo provable equivalence in T. This set is equipped with the usual boolean operations and the provability operator mapping a sentence F to $\mathsf{Provable}_T(F)$. Magari algebras in general are all the structures satisfying the identities of the provability algebra of PA.

Studying Magari algebras revealed many interesting properties of provability. Some of them can also be reformulated in purely logical terms, but for many other questions an algebraic context is the most natural one. An early refinement of Solovay's theorem is its so-called *uniform version* that was discovered independently in [Montagna, 1979; Artemov, 1979; Visser, 1980; Boolos, 1982; Avron, 1984]. In algebraic terms this result means that the free Magari algebra on countably many generators is embeddable into the provability algebra of any sound theory T. [Shavrukov, 1993b] proved a far-reaching generalization by characterizing all r.e. subalgebras of the provability algebra of T.

Using the notion of provability algebra one can give a provability semantics to a considerable subclass of propositional second order modal formulas, that is, modal formulas with quantifiers over arithmetical sentences. These are just the first order formulas over the provability algebra. For several years the questions of decidability of the propositional second-order provability logic, and of the first order theory of the provability algebra of PA, remained open. [Shavrukov, 1997a] gave a negative solution to these questions. His result was proved by one of the most ingenious extensions of Solovay's techniques. We note that the difficult question of decidability of the $\forall\exists$-theory of this algebra remains open.

Applications in proof theory. The logic of formal provability was designed with a hope for applications in proof theory. It considerably deepened our understanding of the behavior of formalized provability predicates.

However, memorable applications of these methods to the study of concrete formal theories were lacking for a long time. The challenge here was to find applications to existing problems that were not *a priori* formulated in terms of formalized provability.

The situation changed in the recent years. It turned out that methods of modal logic can be useful in the study of fragments of Peano arithmetic, where the model theoretic methods were the most successful, so far. It was an open question what kind of computable functions could be proved to be total in the fragment of PA where induction was restricted to Π_2-formulas without parameters. Using provability logic methods [Beklemishev, 1999a] showed that these functions coincide with the primitive recursive ones. In general, provability logic analysis substantially clarified the behavior of parameter-free induction schemata.

Later results [Beklemishev, 2004; Beklemishev, 2003b] revealed a deeper connection between provability logic and traditional proof-theoretic questions, such as consistency proofs, ordinal analysis, and independent combinatorial principles. [Beklemishev, 2004] gives an alternative proof of the famous theorem by G. Gentzen on the proof of consistency of PA by transfinite induction up to the ordinal ϵ_0. In [Beklemishev, 2003b] and in this paper we present a simple combinatorial principle, called *the Worm principle*, which is derived from the provability logic analysis of PA and is independent from PA.

At the moment this area seems to be a promising direction for future research. The provability logic techniques used here combine several of the above mentioned concepts such as provability algebras and polymodal logics à la Japaridze [Japaridze, 1986; Boolos, 1993].

Logic of Proofs

The problem of formalizing *BHK* semantics even for propositional language was not solved until the middle of 1990s (cf. surveys [Weinstein, 1983; van Dalen, 1986; Artemov, 2001] and Section 11 of this article). The source of difficulties in provability interpretation of modality lies in the implicit nature of existential quantifier $\exists$. This phenomenon is sometimes called the $\exists$-sickness of the first-order logic: an assumption of $\exists x F(x)$ in a given formal theory does not necessarily yield $F(t)$ for some term t.

Consider, for instance, the reflection principle in PA, i.e. all formulas of type $\exists x \mathsf{Proof}(x, F) \rightarrow F$. By the second Gödel incompleteness theorem, this principle is not provable in PA, since, the consistency formula $\mathsf{Con}(\mathsf{PA})$ coincides with a special case of the reflection principle $\exists x \mathsf{Proof}(x, \bot) \rightarrow \bot$. Formula $\exists x \mathsf{Proof}(x, F)$ does not yield any specific proof of F, since this x may be a nonstandard natural number which is not a code of any actual derivation in PA. For proofs represented by explicit terms the picture is entirely different, e.g. the principle of *explicit reflection* $\mathsf{Proof}(p, F) \rightarrow F$ is provable in PA for each specific derivation p. Indeed, if $\mathsf{Proof}(p, F)$ holds, then F is evidently provable in PA, and so is formula $\mathsf{Proof}(p, F) \rightarrow F$. Otherwise, if $\mathsf{Proof}(p, F)$ is false, then $\neg\mathsf{Proof}(p, F)$ is true and provable, therefore $\mathsf{Proof}(p, F) \rightarrow F$ is also provable.

This observation suggests a remedy for the $\exists$-sickness here: representing proofs by a system of terms t in the proof formula $\mathsf{Proof}(t, F)$ instead of implicit representation of proofs by existential quantifiers in the provability formula $\exists x \mathsf{Proof}(x, F)$. In particular, it means a return to the original format of *BHK* after failed attempts to find a constructive provability semantics for IPC directly via a simpler language of modal logic. Gödel suggested using the format of explicit proofs for the interpretation of $\mathsf{S4}$ as early as 1938, but that paper remained unpublished until 1995 ([Gödel, 1995]). In a modern terminology the format of explicit proof terms is an instance of Gabbay's Labelled Deductive Systems (cf. [Gabbay, 1994]).

The logic of proofs. In [Artemov and Strassen, 1992a; Artemov and Strassen, 1992b; Artemov and Strassen, 1993] the first systems of logics of proofs in format $t{:}F$ denoting *t is a proof of F* were introduced. These first logics had no operations on proofs and were too weak for representing the modality in full.

Even before the publication of Gödel's paper of 1938 [Gödel, 1995], S. Artemov came up with a system of logic of proofs capturing the whole of $\mathsf{S4}$. In the fall of 1994 during his visit to the University of Amsterdam S.A. found the logic of proofs (which later got the name LP) and proved a theorem about realizability of $\mathsf{S4}$ by proof terms of LP called *proof polynomials*. These results were reported at the end of 1994 in Amsterdam and Münster. The first paper with complete proofs was issued as a technical report of the Mathematical Sciences Institute, Cornell University, [Artemov, 1995]. A follow up paper [Artemov, 2001] contained simplified proofs and a comprehensive survey.

Since proof polynomials enjoy a natural semantics in classical proofs, this gave a desired provability semantics to Gödel' provability calculus $\mathsf{S4}$. Combined with the above mentioned results by Gödel, MacKinsey and Tarski, the logic of proofs LP can be viewed as a formalization of the *BHK* semantics for intuitionistic propositional logic IPC completing a project initiated by Kolmogorov and Gödel. These developments resulted in the following picture of the foundations of intuitionistic logic:

$$\mathsf{IPC} \;\hookrightarrow\; \mathsf{S4} \;\hookrightarrow\; \mathsf{LP} \;\hookrightarrow\; \mathit{CLASSICAL\ PROOFS}\,,$$

where all embeddings are exact.

Models of the logic of proofs and complexity issues. The logic of proofs LP is sound and complete with respect to the natural provability semantics [Artemov, 1995; Artemov, 2001]. Still, having convenient artificial models could be very important for a successful study of LP and its applications. The first abstract models for LP (called here M-models) were introduced in [Mkrtychev, 1997] where LP was shown to be sound and complete with respect to M-models. Mkrtychev models proved to be a convenient tool for studying the logic of proofs. In particular, they helped to establish in [Mkrtychev, 1997] the decidability of LP.

[Kuznets, 2000] obtained an upper bound Σ_2^p on the satisfiability problem for LP-formulas in M-models. This bound was lower than known upper bound PSpace on the satisfiability problem in $\mathsf{S4}$. One of the possible explanations, why LP wins in complexity over closely related to it $\mathsf{S4}$, is that the satisfiability test for LP is somewhat similar to the type checking, i.e. checking the correctness of assigning types (formulas) to terms (proofs), which is known to be relatively easy in classical cases.

M-models were further explored in [Krupski(jr.), 2003], where the minimal model of LP was constructed, which completely describes derivability in LP of “modalized” formulas (i.e. formulas of type $t{:}F$). This yielded a better upper bound (NP) for the “modalized” fragment of LP. The minimal model is also used in [Krupski(jr.), 2003] to answer a well-known question about the disjunctive property of the logic of proofs:

$$\mathsf{LP} \vdash s{:}F \vee t{:}G \quad \Leftrightarrow \quad \mathsf{LP} \vdash s{:}F \;\text{ or }\; \mathsf{LP} \vdash t{:}G.$$

[Fitting, 2003b] gave a description of the canonical model for LP as a Kripke-style model. An interesting and unexpected application of the canonical model was suggested in [Fitting, 2003a], where an alternative “semantical” proof was given for the realizability theorem of $\mathsf{S4}$ in LP, whereby clarifying the role of operation “+” in this realization.

[Fitting, 2003b] gave a general definition of Kripke-style models for LP (we call them F-models here) and established soundness and completeness of the logic of proofs with respect to F-models. As it was noted by V. Krupski, completeness with respect to F-models can be attained on one-element F-models, which are M-models with the so-called full explanatory property (cf. Section 12). It is reasonable to expect to find applications of F-models in epistemic logics containing both proof polynomials and the usual $\mathsf{S4}$-modality, since Kripke models do not degenerate to singletons for such logics.

A tableau system for the logic of proofs was developed in [Renne, 2004] where completeness with respect to M-models and cut-elimination for the whole of LP was proved, though cut-elimination in LP with empty constant specifications was demonstrated in [Artemov, 2001].

Joint logics of proofs and provability. The problem of finding a joint logic of proofs and provability has been a natural next step in this direction since there are important principles formulated in a mixed language of formal provability and explicit proofs. For example, the modal principle of negative introspection $\neg\Box F \rightarrow \Box\neg\Box F$ is not valid in the provability semantics. Neither does a purely explicit version of negative introspection $\neg(x\!:\!F) \rightarrow t(x)\!:\!\neg(x\!:\!F)$ hold in the logic of proofs LP. However, a mixed explicit-implicit principle $\neg(t:F) \rightarrow \Box\neg(t:F)$ is valid in the standard provability semantics. Finding a complete axiomatization of such principles in a joint language of GL and LP has also been important for building an epistemic logic with justifications based on provability semantics.

The first joint system of provability and explicit proofs without operations on proof terms, system B, was found in [Artemov, 1994]. Arithmetically complete system BGrz of *strong provability operator* $\boxdot F = F \wedge \Box F$ and proofs without operations was found in [Nogina, 1994; Nogina, 1996].

In [Sidon, 1997; Yavorskaya (Sidon), 2002] the first arithmetically complete system of provability and explicit proofs, LPP, containing both LP and GL was found. Along with natural extensions of principles and operations from LP and GL, LPP contains some additional operations. The arithmetically complete logic, LPGL, in the joint language of LP and GL was found in [Artemov and Nogina, 2004], where it was also used for building basic systems of logic of knowledge with justifications (cf. **Applications** below and Section 13).

Logic of single-conclusion proofs. The primary use of LP is to realize modalities by proof terms (proof polynomials) thus providing a semantics of explicit proofs for modal logic $\mathsf{S4}$ and for intuitionistic logic IPC. It turned out that with respect to realizability semantics, modal logic corresponds to multi-conclusion proofs, i.e. proofs each of which can prove several different theorems (cf. Section 11). One could see easily that the set of modal principles realizable by single-conclusion proofs (so called *functional proofs*), is not compatible with any normal modal logic. For example, $x\!:\!\top \rightarrow \neg x\!:\!(\top \wedge \top)$ is valid for functional proofs, and its forgetful projection $\Box\top \rightarrow \neg\Box(\top \wedge \top)$ contradicts even the basic modal logic K.

However, the problem of finding the logic of functional proofs presented a significant interest since many proof-like objects (e.g. typed λ-terms and combinatory terms or references in databases) correspond to single-conclusion proofs. The first step in the development of the logic of functional proofs was made in [Artemov and Strassen, 1992b] where the operation-free logic of functional proofs was axiomatized. The full scale logic of functional proofs FLP was built in ([Krupski, 1997; Krupski, 2002]) and then enhanced by new operations in [Krupski, 2005], system FLP_{ref}.

The logic of the standard proof predicate. The logic of proofs LP axiomatizes all properties of propositions and proofs expressible in the propositional language and invariant with respect to the choice of a proof system

([Artemov, 2000; Artemov, 2001]). For a specific proof system some additional identities may hold. For instance, the standard "textbook" proof predicate is based on Gödel's numbering of syntax, which is monotone. In particular, the code of a given proof (a finite sequence of formulas) is greater than the code of any formula in that sequence (including the codes of theorems proven by that sequence). This property of coding prohibits self-referential assertions of sort $t\!:\!A(t)$ and in general yields the following *monotonicity axiom* introduced in [Artemov and Strassen, 1993]:

$$\neg(t_1\!:\!A_2(t_2) \wedge t_2\!:\!A_3(t_3) \wedge \ldots \wedge t_n\!:\!A_1(t_1)),$$

where t_i has to occur in $A_i(t_i)$. This axiom is valid for the standard proof predicate[7] but is not derivable in LP. It was shown in [Artemov and Strassen, 1993] that the basic logic of proofs supplied with the monotonicity axiom (system M) is complete with respect to the standard proof predicate. In [Artemov, 1994] this result is extended to a system in a richer language containing both M and the provability logic GL. A full axiomatization of the propositional logic of the standard proof predicate in the language of LP was found in [Yavorsky, 2000].

Proof polynomials for other modal logics. Systems of proof polynomials for other classical modal logics K, $\mathsf{K4}$, D, $\mathsf{D4}$, T were described in [Brezhnev, 2000; Brezhnev, 2001]. The paper [Brezhnev, 2001] should also be mentioned for its introduction of proof polynomials for Gentzen-style proof systems. The case of $\mathsf{S5} = \mathsf{S4} + (\neg\Box F \rightarrow \Box\neg\Box F)$ was special because of the presence of negative information about proofs. The paper by Artemov, Kazakov and Shapiro [Artemov *et al.*, 1999] introduced a possible system of proof terms for $\mathsf{S5}$, established realizability of the logic $\mathsf{S5}$ by these terms, decidability, and completeness of the resulting logic of proofs. However, the existence of alternative natural systems of proof terms for $\mathsf{S5}$ suggests that the problem of describing negative knowledge by operations on witnesses is far from solved.

Quantified Logics of Proofs. The arithmetical provability semantics for the logic of proofs may be naturally generalized to the first-order language and to the language of LP with quantifiers over proofs. Both possibilities of enhancing the expressive power of LP were investigated. In [Artemov and Sidon-Yavorskaya, 2001], techniques originating from [Artemov, 1985b; Vardanyan, 1986] were used to establish that the set of tautologies in the language of the first-order logic of proofs was not recursively enumerable. It was shown that a complete axiomatization of the first-order logic of proofs

[7] This holds for the usual "call-by-value" provability semantics for LP presented in this article. However, this does not necessarily hold for the "call-by-name" semantics from [Artemov, 1995] (cf. also [Artemov, 2001], Comment 6.8).

is impossible. An interesting decidable fragment of the first-order logic of the standard proof predicate was found in [Yavorsky, 2000]. Propositional logic with quantifiers over proofs was studied in [Yavorsky, 2002]. It was established that the corresponding set of formulas valid under the natural provability interpretation is not recursively enumerable, therefore propositional logic with quantifiers over proofs is not axiomatizable.

Applications. 1. We start with a discussion of a contribution to semantics of modal logic in general made by the provability logic and logic of proofs. Initially Gödel regarded the modality $\Box F$ from a provability point of view as

there exists a proof (*witness, justification*) *for* F

According to this interpretation, modality contains an informal built-in existential quantifier over proofs. Existential understanding of modality is also typical of "naive" semantics for a wide range of epistemic logics. Nonetheless, before the logic of proofs LP was discovered, major modal logics lacked an exact semantics of existential character. The first exact existential semantics of modality is given by the arithmetical provability model for system GL, which, however, does not extend to other major modal logics. Proof polynomials and the logic of proofs provide existential semantics for S4, S5 and other systems [Artemov *et al.*, 1999; Brezhnev, 2000; Brezhnev, 2001].

Decades after the above mentioned works by Gödel a semantics of a different nature was formalized for modalities, namely Kripke semantics. Modality there is similar to a universal quantifier: $\Box F$ is read as

in all possible situations F *holds.*

Semantics of this sort will be called here a *universal semantics.* Such a reading of modality naturally appears in dynamic and temporal logics aimed at describing computational processes, states of which usually form a (possibly branching) Kripke structure. To some extend, Tarski's topological semantics for S4 can be regarded as a universal semantics as well ([McKinsey and Tarski, 1946; Rasiowa and Sikorski, 1963]).

Universal semantics has been playing a prominent role in modal logic. However, it is not the only possible tool for approaching specific problems involving modal languages. In particular, universal semantics alone did not lead to a solution of the Gödel provability calculus problem because of an existential nature of the latter.

2. As we have already discussed above, a perspective area of applications of the logic of proofs is the area of logics of knowledge. A need for a logic of knowledge with justifications has been discussed in [van Benthem, 1991]. Such a logic along with the usual knowledge operators $\Box F$ (F *is*

known) should contain assertions $t\!:\!F$ (*t is an evidence of F*), thus bringing explicit and quantitative components to the logic of knowledge. The explicit character of judgments significantly expands the expressive power of epistemic logics. Because of the logical omniscience effect (cf. below), the original epistemic modality $\Box F$ should be regarded as "potential knowledge", or "knowability" rather than actual knowledge, cf. [Fitting, 2003b; Fitting, 2005]. An evidence operator $t\!:\!F$ provides a justification that F is true in all situations and hence represents a real knowledge of the agent. [Artemov and Nogina, 2004] used the provability logic with justification LPGL for building logics of knowledge with justifications. Provability logic GL itself is not compatible with the epistemic logic, mainly because arithmetical provability is not reflexive. However, $\mathsf{S4}$ can be modelled in GL by using the strong provability operator. $\mathsf{S4}$ is sound with respect to the strong provability semantics, the extension $\mathsf{S4Grz}$ of $\mathsf{S4}$ by Grzegorczyk schema $\Box(\Box(F \to \Box F) \to F) \to F$ provides a complete propositional axiomatization of strong provability [Kuznetsov and Muravitsky, 1977; Goldblatt, 1978; Boolos, 1979b; Kuznetsov and Muravitsky, 1986]. Kripke models corresponding to $\mathsf{S4Grz}$ have $\mathsf{S4}$-frames which do not distinguish possible worlds mutually accessible from each other. [Artemov and Nogina, 2004] constructed basic logics of knowledge with justifications: $\mathsf{LPS4}$, consisting of $\mathsf{S4}$ combined with LP and $t\!:\!F \to \Box F$, and $\mathsf{LPS4}^-$, which is $\mathsf{LPS4}$ augmented by the principle of negative introspection $\neg(t\!:\!F) \to \Box\neg(t\!:\!F)$.

3. The language of proof carrying formulas of the logic of proofs also suggests an approach to the *logical omniscience problem* [Parikh, 1987; Moses, 1988; Parikh, 1995; Fagin *et al.*, 1995]). *Logical omniscience* means the unrealistical assumption of epistemic logic that an intellectual agent knows all logical consequences of her data. According to this assumption each person who knows the rules of chess should also know whether or not White has a winning strategy (an example from [Fagin *et al.*, 1995]). The logical omniscience problem is to develop a mechanism in the logic of knowledge for distinguishing facts that are "easy to establish" from those which are "hard to establish." The size of a proof polynomial (possibly in a richer basis tailored to specifics of the problem) gives information about the amount of work needed to establish the given fact.

4. Another promising area of applications for the logic of proofs is the area of *typed theories and programming languages*. The usual typed λ-calculus and the typed combinatory logic equivalent to it served as a theoretical prototype for a certain class of programming languages (cf. a survey [Constable, 1998]). The logic of proofs along with the reflexive λ-calculus and the reflexive combinatory logic based on it (cf. [Alt and Artemov, 2001; Artemov, 2004] and Chapter 15) have more expressive power, including a richer system of types and self-referential methods of constructing and using them. It is natural to expect these new capabilities to find their applications in programming languages like did previous major theoretical developments

in λ-calculi.

5. Yet another area of applications of methods raising from the logic of proofs is *reflection* in artificial intelligence, automated deduction and verification. Reflection is a general term describing an ability of a formal deduction system to formalize its own meta-reasoning. This normally includes internal representation of formulas, axioms, rules and derivations, semantics, etc., and ability to represent properties of those objects by formulas of the system. The problem of building reflection in automated deduction has been discussed, e.g. in [Allen *et al.*, 1990; Constable, 1994; Constable, 1998; Harrison, 1995]. The explicit representation of proofs by proof polynomials rather than their implicit specification by quantifiers offers a new promising approach to building reflection. In particular, since explicit reflection is internally provable, this new approach allows us to avoid undesirable "reflection towers" of extensions of a theory of an increasing metamathematical strength [Artemov, 1999], which are unavoidable in the traditional theory of verification [Davis and Schwartz, 1979]. According to [McCarthy, 2004], *self-awareness* is the principle advantage of human intelligence over artificial intelligence. Logical reflection apparatus and the logic of proofs in particular could contribute to building self-aware artificial intelligence systems. In programming languages reflection can be used to naturally formalize Run Time Code Generating, RTCD. About logic analysis of RTCD cf. [Wickline *et al.*, 1998].

6. Among applications one should mention a joint paper [Artemov and Krupski, 1996] introducing a logical system for the description and the design of reference databases based on the logic of proofs. This line of research has been further pursued in [Krupski, 2005].

Part I, Logic of Provability

2 GÖDEL–LÖB PROVABILITY LOGIC: THE MODAL LOGICAL TRADITION

When formulating a new (modal) logic a number of standard questions immediately present themselves. For example, one would want to know how the logic behaves w.r.t. the following properties:

(i) Adequate semantics (completeness, finite model property);

(ii) Decidability, complexity;

(iii) Gentzen-style formulation, cut-elimination, subformula property;

(iv) Craig interpolation, Beth definability;

(v) Normal forms of (some classes of) formulas.

Now that so many systems of nonclassical and modal logic have been studied, such questions have become commonplace and are perhaps lacking certain amount of appeal. Rather, one is more interested in the other, more specific, features of the logics in question. However, the answers to these traditional questions help us to understand the system we are dealing with and provide some useful standard techniques. For the case of basic Gödel–Löb provability logic GL the answers to these standard questions constitute early work in this area. Most of them are discussed at length in the article by C. Smoryński in this Handbook [Smoryński, 2004]. We quickly recapitulate them in this section, mostly to fix the terminology.

2.1 Hilbert-style (Frege) proof system

The language of GL has propositional variables p_0, p_1, $\ldots$; boolean connectives $\to$, $\bot$, $\top$, and unary modality $\Box$. A Hilbert-style proof system for GL is given by the following axiom schemes and rules of inference.

Axiom schemes:

1. Boolean tautologies
2. $\Box(\varphi \to \psi) \to (\Box\varphi \to \Box\psi)$ (*normality*)
3. $\Box(\Box\varphi \to \varphi) \to \Box\varphi$ (*Löb's axiom*)

Rules of inference: $\varphi,\ \varphi \to \psi/\psi$ (*modus ponens*); $\varphi/\Box\varphi$ (*necessitation*).

It is well-known that GL proves the *transitivity* axiom $\Box\varphi \to \Box\Box\varphi$ and therefore extends the system K4 (see [Smoryński, 1985]). On the other hand, GL is incompatible with the *reflexivity* axiom $\Box\varphi \to \varphi$ and therefore with the system S4.

2.2 Kripke models

A *Kripke model* for GL (or simply a *model*) is a triple $\mathcal{K} := (K, \prec, \Vdash)$, where

- $\prec$ is a converse well-founded strict partial ordering on K. The poset $(K, \prec)$ is called the *frame* of $\mathcal{K}$. Elements of K are called *nodes*. We assume, unless explicitly mentioned otherwise, that every model has the minimal node, which is called the *root* of $\mathcal{K}$.

- $\Vdash$ is a *forcing relation* on $\mathcal{K}$, that is, a binary relation between the nodes of $\mathcal{K}$ and modal formulas, which satisfies the following conditions for any $x \in K$ and any formulas φ, ψ:

  1. $x \nVdash \bot$, $x \Vdash \top$;
  2. $x \Vdash \varphi \to \psi \iff (x \nVdash \varphi \text{ or } x \Vdash \psi)$;
  3. $x \Vdash \Box\varphi \iff \forall y \in K(x \prec y \Rightarrow y \Vdash \varphi)$.

By Conditions 1–3 the forcing relation on $\mathcal{K}$ is uniquely determined by its restriction to propositional variables. We say that a formula φ *holds* or *is valid* in a model $\mathcal{K}$ (denoted $\mathcal{K} \Vdash \varphi$) if it is forced at the root of $\mathcal{K}$. $\mathcal{K}, x \Vdash \varphi$ means $x \Vdash \varphi$ in $\mathcal{K}$. $\mathcal{K} \vDash \varphi$ means $x \Vdash \varphi$, for all $x \in \mathcal{K}$.

A model $\mathcal{K}$ is *treelike*, if so is the ordering $(K, \prec)$, that is, if $a, b \prec c$ implies $a \prec b$ or $b \prec a$ or $a = b$.

2.3 Gentzen-style proof system

We consider *sequents* of the form $\Gamma \Rightarrow \Delta$, where Γ and Δ are finite sets of formulas. (Thus, contraction and permutation rules are built in the definition of a sequent.) $\bigvee\Gamma$ means the formula $\varphi_1 \vee \cdots \vee \varphi_n$, if $\Gamma = \{\varphi_1, \ldots, \varphi_n\}$, and $\bot$, if $\Gamma = \varnothing$. $\bigwedge\Gamma$ is defined dually. $\Box\Gamma$ is the set $\{\Box\varphi : \varphi \in \Gamma\}$. As usual, we also write Γ, φ for $\Gamma \cup \{\varphi\}$ and $\Rightarrow \varphi$ for $\varnothing \Rightarrow \varphi$.

A Gentzen-style proof system GL^G is given by the following axioms and rules of inference.

Axioms: $\bot \Rightarrow$; $\quad \Rightarrow \top$; $\quad p \Rightarrow p$, for any variable p;

Rules of inference:

$$\frac{\Gamma, \psi \Rightarrow \Delta \quad \Gamma \Rightarrow \Delta, \varphi}{\Gamma, \varphi \to \psi \Rightarrow \Delta}\ (\to l) \qquad \frac{\Gamma, \varphi \Rightarrow \psi, \Delta}{\Gamma \Rightarrow \Delta, \varphi \to \psi}\ (\to r)$$

$$\frac{\Gamma \Rightarrow \Delta}{\Gamma, \Sigma \Rightarrow \Delta, \Pi}\ (\text{weak}) \qquad \frac{\Box\Gamma, \Gamma, \Box\varphi \Rightarrow \varphi}{\Box\Gamma \Rightarrow \Box\varphi}\ (\text{Löb})$$

As usual, the weakening rule (weak) can be eliminated at the cost of adding side formulas Σ, Π to all the axioms and the conclusion of the rule

(Löb). First, we observe the obvious *subformula property* of the Gentzen-style proof system.

PROPOSITION 1. *Any formula occurring in a* GL^G*-derivation of a sequent* $\Gamma \Rightarrow \Delta$ *is a subformula of a formula from* $\Gamma \cup \Delta$*.*

Theorem 2 below implies that the rule of *cut*

$$\frac{\Gamma, \varphi \Rightarrow \Delta \quad \Gamma \Rightarrow \varphi, \Delta}{\Gamma \Rightarrow \Delta}\ (\mathrm{cut})$$

is admissible in the system GL^G.

2.4 Joint completeness and cut-elimination theorem

K. Segerberg [Segerberg, 1971] gave the first Kripke completeness proof for GL. A correct Gentzen-style cut-free system for GL has been suggested in [Leivant, 1981], but his (syntactic) proof of cut-elimination contained a gap. Later a correct syntactic proof has been found in [Sambin and Valentini, 1982; Sambin and Valentini, 1983]. Below we present a different (semantic) proof following [Avron, 1984]. A corresponding system of natural deduction for GL was given in [Bellin, 1985].

THEOREM 2. *For any formula* φ *the following statements are equivalent:*

(i) $\mathsf{GL} \vdash \varphi$*;*

(ii) $\mathcal{K} \Vdash \varphi$*, for all models* $\mathcal{K}$*;*

(iii) $\mathcal{K} \Vdash \varphi$*, for all finite treelike models* $\mathcal{K}$*;*

(iv) $\mathsf{GL}^G \vdash\ \Rightarrow \varphi$*;*

(v) $\mathsf{GL}^G + (\mathrm{cut}) \vdash\ \Rightarrow \varphi$*.*

Proof. The implication (i)⇒(ii) is the soundness of GL w.r.t. converse well-founded Kripke models cf. [Smoryński, 2004; Smoryński, 1985]. The implications (ii)⇒(iii) and (iv)⇒(v) are obvious.

The implication (v)⇒(i) is the adequacy of the Gentzen-style formulation of GL. We have to show that all inference rules of GL^G are admissible in GL under the standard translation of sequents $\Gamma \Rightarrow \Delta$ as the formulas $\bigwedge \Gamma \to \bigvee \Delta$. This is easy for the propositional rules and the cut-rule (the latter corresponds, in a sense, to modus ponens). We derive (Löb) by the following reasoning in GL:

1. $\bigwedge \Gamma \wedge \bigwedge \Box\Gamma \wedge \Box\varphi \to \varphi$ (assumption)
2. $\bigwedge \Gamma \wedge \bigwedge \Box\Gamma \to (\Box\varphi \to \varphi)$

3. $\Box(\bigwedge\Gamma \wedge \bigwedge\Box\Gamma) \rightarrow \Box(\Box\varphi \rightarrow \varphi)$ (by normality from 2)

4. $\Box(\bigwedge\Gamma \wedge \bigwedge\Box\Gamma) \rightarrow \Box\varphi$ (by Löb's axiom from 3)

5. $\bigwedge\Box\Gamma \rightarrow \Box(\bigwedge\Gamma \wedge \bigwedge\Box\Gamma)$ (a theorem of $\mathsf{K4}$)

6. $\bigwedge\Box\Gamma \rightarrow \Box\varphi$ (from 4,5)

The central part of the proof of the theorem is (iii)⇒(iv); here is a sketch. Assume a sequent $\Gamma \Rightarrow \Delta$ is not provable in GL^G. Then it can be extended to an unprovable saturated sequent, that is, a sequent $\Gamma_1 \Rightarrow \Delta_1$ satisfying:

(i) $(\varphi \rightarrow \psi) \in \Gamma_1$ implies $\varphi \in \Delta_1$ or $\psi \in \Gamma_1$;

(ii) $(\varphi \rightarrow \psi) \in \Delta_1$ implies $\varphi \in \Gamma_1$ and $\psi \in \Delta_1$;

(iii) $\Gamma \subseteq \Gamma_1$, $\Delta \subseteq \Delta_1$ and any formula in $\Gamma_1 \cup \Delta_1$ is a subformula of a formula from $\Gamma \cup \Delta$;

(iv) $\mathsf{GL}^G \nvdash \Gamma_1 \Rightarrow \Delta_1$.

Consider the (finite) set of all such unprovable saturated sequents. Supply it with a partial ordering $\prec$ as follows: $(\Sigma_1 \Rightarrow \Pi_1) \prec (\Sigma_2 \Rightarrow \Pi_2)$ iff

(i) $\Box\varphi \in \Sigma_1$ implies $\varphi, \Box\varphi \in \Sigma_2$;

(ii) There is a $\Box\varphi \in \Sigma_2$ such that $\Box\varphi \notin \Sigma_1$.

Let $(K, \prec)$ be the restriction of this ordering to the set of all sequents above $\Gamma_1 \Rightarrow \Delta_1$. Define an assignment of propositional variables p on K by setting

$$(\Sigma \Rightarrow \Pi) \Vdash p \iff p \in \Sigma.$$

This gives us a Kripke model $\mathcal{K} = (K, \prec, \Vdash)$ with the root $\Gamma_1 \Rightarrow \Delta_1$. Now it is a matter of routine checking, for any sequent $(\Sigma \Rightarrow \Pi) \in K$ and formula $\varphi \in \Sigma \cup \Pi$, that

(i) $\varphi \in \Sigma$ implies $(\Sigma \Rightarrow \Pi) \Vdash \varphi$;

(ii) $\varphi \in \Pi$ implies $(\Sigma \Rightarrow \Pi) \nVdash \varphi$.

Therefore, we conclude: $(\Gamma_1 \Rightarrow \Delta_1) \nVdash \bigwedge\Gamma \rightarrow \bigvee\Delta$.

Notice that $(K, \prec)$ is a finite strict partial ordering, which may not yet be treelike. However, $\mathcal{K}$ can be transformed into an equivalent treelike model by the standard unravelling procedure (see [Bull and Segerberg, 2001]). ■

COROLLARY 3. GL^G *is closed under the cut-rule.*

COROLLARY 4. GL *is decidable and enjoys the finite model property.*

We also mention without proof that by a result of A.V. Chagrov [Chagrov, 1985] the set of theorems of GL is PSPACE-complete. See [Chagrov *et al.*, 2001; Švejdar, 2003] for more details.

2.5 Interpolation and definability

As an expected corollary of cut-elimination we obtain the Craig interpolation theorem for GL.

THEOREM 5 (Craig interpolation). *If* $\mathsf{GL} \vdash \varphi \to \psi$*, then there is a* θ *such that* $\mathrm{Var}(\theta) \subseteq \mathrm{Var}(\varphi) \cap \mathrm{Var}(\psi)$ *and*

$$\mathsf{GL} \vdash \varphi \to \theta \textit{ and } \mathsf{GL} \vdash \theta \to \psi.$$

Proof. Using the so-called Schütte–Maehara method we prove the following statement by induction on the depth of the GL^G-derivation: If

$$\mathsf{GL}^G \vdash \Gamma_1, \Gamma_2 \Rightarrow \Delta_1, \Delta_2,$$

then there is a formula θ such that $\mathrm{Var}(\theta) \subseteq \mathrm{Var}(\Gamma_1 \cup \Delta_1) \cap \mathrm{Var}(\Gamma_2 \cup \Delta_2)$ and

$$\mathsf{GL}^G \vdash \Gamma_1 \Rightarrow \Delta_1, \theta \text{ and } \mathsf{GL}^G \vdash \theta, \Gamma_2 \Rightarrow \Delta_2.$$

For the axioms and each of the rules the construction of θ is straightforward. Now put $\Gamma_1 = \{\varphi\}$, $\Delta_2 = \{\psi\}$, $\Gamma_2 = \Delta_1 = \varnothing$. ■

This theorem has been proved independently by C. Smoryński [Smoryński, 1978] and G. Boolos [Boolos, 1979b] by semantical arguments. The reader can find this proof in [Chagrov *et al.*, 2001]. As a standard corollary we obtain

COROLLARY 6 (Beth definability). *Assume* $\mathsf{GL} \vdash \varphi(p) \land \varphi(q) \to (p \leftrightarrow q)$*, where* q *does not occur in* $\varphi(p)$ *and the formula* $\varphi(q)$ *is obtained from* $\varphi(p)$ *by replacing all occurrences of* p *by* q*. Then there is a formula* ψ *such that* $\mathrm{Var}(\psi) \subseteq \mathrm{Var}(\varphi(p)) \setminus \{p\}$ *and* $\mathsf{GL} \vdash \varphi(p) \to (p \leftrightarrow \psi)$.

Proof. We are given a formula φ satisfying $\mathsf{GL} \vdash \varphi(p) \land p \to (\varphi(q) \to q)$. Let ψ be an interpolant for this formula. ■

An interesting corollary of this general result is the Fixed Point Theorem for GL, which was thoroughly discussed in [Smoryński, 2004]. We sketch a short alternative proof due to C. Smoryński.

THEOREM 7 (Fixed points). *Let* $\varphi(p)$ *be a formula in which* p *only occurs within the scope of a* $\Box$*. Then there is a formula* ψ *such that* $\mathrm{Var}(\psi) \subseteq \mathrm{Var}(\varphi) \setminus \{p\}$ *and*

$$\mathsf{GL} \vdash \boxdot(p \leftrightarrow \varphi(p)) \leftrightarrow \boxdot(p \leftrightarrow \psi).$$

Here $\boxdot\theta$ is an abbreviation for $\theta \land \Box\theta$. The reader is invited to convince him/herself that this formulation implies both the existence and the uniqueness of fixed points, as stated in [Smoryński, 2004].

Proof. Let $\varphi(p)$ be given. First we obtain the following lemma [Bernardi, 1976]:

$$\mathsf{GL} \vdash \boxdot(p \leftrightarrow \varphi(p)) \wedge \boxdot(q \leftrightarrow \varphi(q)) \rightarrow (p \leftrightarrow q),$$

where q is a fresh variable not contained in $\varphi(p)$. For a proof of this lemma see [Smoryński, 2004] or a simple Kripke-model argument in [Boolos, 1993]. Then apply Beth's definability theorem to the formula $\boxdot(p \leftrightarrow \varphi(p))$. ∎

The Craig interpolation theorem has various extensions and strengthenings. The most well-known ones are the so-called *Lindon interpolation* and the *uniform interpolation*. Whether the Lindon interpolation holds for GL still seems to be an open question.

THEOREM 8 (Uniform interpolation). *Let a formula φ and a subset $S \subseteq \mathrm{Var}(\varphi)$ be given. Then there is a formula θ such that $\mathsf{GL} \vdash \varphi \rightarrow \theta$, $\mathrm{Var}(\theta) \subseteq S$ and for every formula ψ such that $\mathrm{Var}(\psi) \cap \mathrm{Var}(\varphi) \subseteq S$ and $\mathsf{GL} \vdash \varphi \rightarrow \psi$, we have $\mathsf{GL} \vdash \theta \rightarrow \psi$.*

This theorem was discovered by V. Shavrukov [Shavrukov, 1993b] independently from (and essentially simultaneously with) a similar result by A. Pitts [Pitts, 1992] on intuitionistic propositional logic. Shavrukov's proof was semantical rather than syntactical and relied upon the techniques of *characters*. Later A. Visser [Visser, 1996], building on the work [Ghilardi and Zawadowski, 1995], gave a more transparent semantical proof. No syntactical proof of this theorem for GL is known.

2.6 Admissible rules

A propositional inference rule

$$\frac{\varphi_1, \dots, \varphi_n}{\psi}\ (R)$$

is *admissible* in a logic L, if whenever $L \vdash \sigma(\varphi_i)$, for $i = 1, \dots, n$, there holds $L \vdash \sigma(\psi)$, where σ is any substitution of formulas for propositional variables. Typical examples of admissible rules in GL are

$$\frac{\Box p \quad \Box q}{\Box(p \wedge q)}\ (R_1) \quad \text{and} \quad \frac{\Box p}{p}\ (R_2).$$

Admissible rules must not be confused with the derivable rules in a concrete proof system $\mathcal{P}$ for L. A rule R as above is called *derivable* in $\mathcal{P}$, if there is a derivation in $\mathcal{P}$ of the formula ψ from the assumptions $\varphi_1, \dots, \varphi_n$. This notion depends not just on the set of theorems of L, but also on the choice of specific basic inference rules. Typically, all the basic rules of $\mathcal{P}$, and hence all the derivable rules, are admissible. The converse need not be the case, and is not the case for GL.

For the standard Hilbert-style proof system for GL given in Section 2.1, which we temporarily denote GL^H, the derivable rules can be easily characterized by means of the following version of Deduction theorem (see [Smoryński, 2004] or [Boolos, 1993]).

PROPOSITION 9 (Deduction theorem). *A rule* (R) *is derivable in* GL^H *iff*

$$\mathsf{GL} \vdash \boxdot\varphi_1 \wedge \cdots \wedge \boxdot\varphi_n \to \psi.$$

Thus, we see that the rule (R_1) is derivable and admissible, whereas the rule (R_2) is admissible but not derivable. (If it were, GL would prove $\boxdot\Box p \to p$ and hence $\Box p \to p$, which is not the case.) The reader can also easily check that, in contrast with (R_2), *Löb's rule*

$$\frac{\Box p \to p}{p}$$

is derivable in GL^H.

V. Rybakov [Rybakov, 1989] obtained the following important results.

THEOREM 10 (Rybakov). *The property of a rule being admissible in* GL *is decidable.*

THEOREM 11 (Rybakov). *The admissible rules in* GL *do not have a finite basis, that is, they cannot be described as derivable rules in a proof system given by finitely many axiom schemes and inference rules.*

Similar results hold for the propositional intuitionistic logic and many other modal logics. See [Chagrov *et al.*, 2001] for more details on the topic of admissibility of rules and a sketch of a proof of Rybakov's theorem. See also [Rybakov, 1997] for an in-depth monograph on admissible rules.

An alternative proof of Rybakov's theorem was obtained by methods of S. Ghilardi [Ghilardi, 1999]. Ghilardi's techniques proved to be especially useful in the study of intuitionistic provability logic (see [Ghilardi, 1999; Iemhoff, 2001b; Iemhoff, 2001c] and Section 9).

2.7 Letterless formulas and traces

A modal formula is *letterless*, if it contains no propositional variables and is thus built up from $\top$, $\bot$ using $\to$ and $\Box$. Letterless formulas have nice normal forms in GL. This fact was discovered by G. Boolos [Boolos, 1976] and independently by J. van Benthem, C. Bernardi and F. Montagna. We obtain these normal forms using the techniques of traces of modal formulas developed in [Artemov, 1980].

Let $\mathcal{K}$ be a (possibly infinite) model. The *depth function* on $\mathcal{K}$ is a mapping d from $\mathcal{K}$ to the ordinals uniquely defined by the following condition:

$$\forall x \in K \ d(x) = \sup\{d(y) + 1 \mid x \prec y\},$$

where we assume $\sup\varnothing = 0$. Recall that all models are converse well-founded, so d is a well-defined function. The *height* $h(\mathcal{K})$ of $\mathcal{K}$ is the depth of its root.

Let φ be a (not necessarily letterless) formula. *Trace* $tr(\varphi)$ of φ is the set of all numbers $n \in \omega$ such that there is a (finite) model $\mathcal{K}$ of height n such that $\mathcal{K} \nVdash \varphi$. Clearly, theorems of GL and only them leave no trace.

We define: $F_n = (\Box^{n+1}\bot \to \Box^n\bot)$. It is easy to see that $tr(F_n) = \{n\}$.

LEMMA 12. *For any formula φ, $tr(\varphi)$ is either a finite or a cofinite subset of ω.*

Proof. Assume $tr(\varphi)$ is infinite and let $\Box\varphi_1, \dots, \Box\varphi_m$ enumerate all subformulas of φ of the form $\Box\psi$. There is a model $\mathcal{K}$ such that $\mathcal{K} \nVdash \varphi$ and $h(\mathcal{K}) > m$. By Lemma 26 below there is a node $r \in \mathcal{K}$ such that $r \Vdash \Box\varphi_i \to \varphi_i$ for each i. Using this property we can 'insert' in our model a linear chain of elements at the node r without changing the forcing at the nodes of $\mathcal{K}$. Formally, for each n a new model $\mathcal{K}_n$ is defined such that K_n is the disjoint union of K and the set $\{0, \dots, n\}$. The ordering $\prec_n$ on K_n is the transitive closure of the orderings $\prec$ on K, $<$ on $\{0, \dots, n\}$, and the following relations:

(i) $x \prec_n y$, for all $y \le n$ and $x \prec r$;

(ii) $y \prec_n x$, for all $y \le n$ and $r \preceq x$.

The forcing relation for propositional variables on $\{0, \dots, n\}$ coincides with that at r and is the same as in $\mathcal{K}$ everywhere else. It is then not difficult to show that this property extends from atomic to all subformulas of φ and hence $\mathcal{K}_n \nVdash \varphi$. This holds for any n, so $tr(\varphi)$ is cofinite. ■

LEMMA 13. *If F and φ are modal formulas such that $tr(\varphi) \subseteq tr(F)$ and F is letterless, then $\mathsf{GL} \vdash F \to \varphi$.*

Proof. Consider the structure $\mathcal{N} = (\omega, >)$ as a converse well-founded (rootless, infinite) Kripke frame. The depth function d maps any finite model $\mathcal{K}$ to $\mathcal{N}$. Moreover, for any $x \in K$ and any letterless formula ψ we have

$$\mathcal{K}, x \Vdash \psi \iff \mathcal{N}, d(x) \Vdash \psi,$$

as can be easily seen by induction on ψ. Hence, $n \in tr(\psi)$ iff $\mathcal{N}, n \nVdash \psi$ for letterless ψ.

Assuming $\mathsf{GL} \nvdash F \to \varphi$ take any finite model $\mathcal{K}$ such that $\mathcal{K} \Vdash F$ and $\mathcal{K} \nVdash \varphi$ and let $n = h(\mathcal{K})$. Obviously $n \in tr(\varphi)$, but we have $\mathcal{N}, n \Vdash F$ by the previous observation. Hence, $n \notin tr(F)$, contradicting our assumption. ■

As a corollary we obtain that any letterless formula is determined by its trace up to provable equivalence:

COROLLARY 14. *Let φ, ψ be letterless. Then*

$$tr(\varphi) = tr(\psi) \iff \mathsf{GL} \vdash \varphi \leftrightarrow \psi.$$

The following corollary provides normal forms for letterless formulas.

THEOREM 15 (Normal forms). *Let φ be a letterless formula and $S = tr(\varphi)$.*

(i) If S is finite, then $\mathsf{GL} \vdash \varphi \leftrightarrow \bigwedge_{n \in S} F_n$;

(ii) If S is cofinite, then $\mathsf{GL} \vdash \varphi \leftrightarrow \bigvee_{n \notin S} \neg F_n$.

Proof. By the previous corollary we must only notice that $tr(\bigwedge_{n \in S} F_n) = S$, if S is finite, and $tr(\bigvee_{n \notin S} \neg F_n) = S$, if S is cofinite. ■

We also remark that by the proof of Lemma 12 one can effectively determine whether $tr(\varphi)$ is finite or cofinite, as well as find an upper bound to the elements in $tr(\varphi)$ (respectively, $\omega \setminus tr(\varphi)$). Testing $n \in tr(\varphi)$ is effective because, by Lemma 13,

$$n \in tr(\varphi) \iff \mathsf{GL} \nvdash \varphi \to F_n.$$

This means that the trace of a formula φ, together with the normal form of φ if φ is letterless, can be determined effectively.

3 THE INTENDED PROVABILITY SEMANTICS

R. Solovay originally formulated his completeness theorems for Peano arithmetic PA. It was immediately clear that his results applied to a wider range of theories. Applications of provability logic also required working with different systems some of which are much weaker and some much stronger than Peano arithmetic. Therefore, we will have to extend the approach taken in [Smoryński, 2004]. It will be important for us not to fix one particular theory but rather keep the possibility of different interpretations of $\Box$ open.

We shall deal with formal theories T "sufficiently strong to be able to reason about themselves." This is usually achieved by specifying a *Gödel numbering*, that is, an assignment of a numerical code $\ulcorner\tau\urcorner$ to every syntactic object τ in the language of T — variable, term, formula, proof, etc. (We shall freely identify these codes with numerals, that is, the terms representing numbers in T.) Then, if T knows enough about numbers and has the power of coding and decoding, T will be able to reason about its own syntax. Thus, one usually restricts the attention to theories T containing (a sufficiently strong fragment of) Peano arithmetic. The standard choice of such a fragment is *primitive recursive arithmetic* PRA. Its formulation can be obtained from that of PA in [Smoryński, 2004] by restricting the induction schema to quantifier-free formulas. A somewhat more economical choice is *elementary arithmetic* EA, which is also the weakest theory to date for which Solovay's theorems have been verified[8]. There are, however, yet weaker theories for which an adequate formalization of syntax has been developed. The most important among them is Buss' feasible arithmetic S^1_2 (see [Buss, 1986]). It is open, if Solovay's theorems hold for S^1_2. Therefore, we choose EA as our basic system. Readers who feel insecure about reasoning in weak arithmetics may freely read PA instead of EA for most of this chapter.

3.1 Elementary arithmetic

The language of arithmetic is a first order language containing binary predicate symbols $=$ and $\leq$; binary function symbols $+$ and $\cdot$; unary function symbols S and exp; and a constant 0. The *standard model* of arithmetic is a model with the universe $\mathbb{N} = \{0, 1, 2, \dots\}$ such that all the symbols have their usual interpretation: $=$ is the equality relation; $\leq$ is the ordering relation; $+$ and $\cdot$ are the addition and multiplication operations; S is the successor function $S(x) = x + 1$; exp is the base 2 exponentiation function $\exp(x) = 2^x$.

Formulas in the above language are called *arithmetical.* The expressions $\forall x \leq t\ \varphi(x)$ and $\exists x \leq t\ \varphi(x)$ abbreviate the formulas $\forall x\ (x \leq t \to \varphi(x))$ and $\exists x\ (x \leq t \wedge \varphi(x))$, respectively, where t is any term (not containing the variable x). Occurrences of quantifiers of this kind are called *bounded*, and Δ_0 or *elementary formulas* are those, all of whose quantifiers are bounded. Notice that, by definition, quantifier-free formulas are elementary.

Obviously, predicates definable by Δ_0-formulas in $\mathbb{N}$ are decidable. A rough estimate of the complexity of the evaluation procedure shows that such predicates are decidable in multi-exponential number of steps. The converse is also true (see [Cutland, 1980; Rose, 1984]).

Arithmetical formulas are classified according to their logical complexity into the *arithmetical hierarchy.* For $n \geq 0$ the classes of Σ_n- and Π_n-formulas are inductively defined as follows. Σ_0- and Π_0-formulas are elementary formulas. Σ_{n+1}-formulas are those of the form $\exists x_1 \dots \exists x_m A(x_1, \dots, x_m)$, where A is a Π_n-formula. Π_{n+1}-formulas are $\forall x_1 \dots \forall x_m A(x_1, \dots, x_m)$, where A is a Σ_n-formula.

From the prenex normal form theorem we know that every arithmetical formula is logically equivalent to a Σ_n-formula, for some n. By extension

[8] In various modifications this theory is also known under the names EFA (H. Friedman), ERA (W. Sieg), $I\Delta_0 + \exp$ (A. Wilkie, J. Paris), $I\Delta_0^{\exp}(\exp)$ (P. Hájek and P. Pudlák).

of terminology, we shall often call Σ_n any formula logically equivalent to a Σ_n-formula in the sense of our official definition. Modulo logical equivalence:

1. The classes Σ_n and Π_n are closed under $\vee, \wedge$.
2. $A \in \Sigma_n \iff \neg A \in \Pi_n$, and dually.
3. The class Π_n is closed under the universal quantification,
   the class Σ_n is closed under the existential quantification.

From the computational point of view, the most interesting class of formulas is Σ_1. It follows from the work of Gödel and Kleene that a relation on $\mathbb{N}$ is definable by a Σ_1-formula iff it is recursively enumerable (r.e.).

Elementary Arithmetic EA is a first order theory with equality formulated in the arithmetical language and having the following mathematical axioms:

P1. $\neg S(a) = 0$

P2. $S(a) = S(b) \rightarrow a = b$

P3. $a + 0 = a$

P4. $a + S(b) = S(a + b)$

P5. $a \cdot 0 = 0$

P6. $a \cdot S(b) = a \cdot b + a$

P7. $\exp(0) = S(0)$

P8. $\exp(S(a)) = \exp(a) + \exp(a)$

P9. $a \leq 0 \leftrightarrow a = 0$

P10. $a \leq S(b) \leftrightarrow (a \leq b \vee a = S(b))$

and the *induction axiom schema* for bounded formulas $\varphi(x, \vec{a})$:

$$\text{(Ind)} \qquad \varphi(0, \vec{a}) \wedge \forall x\, (\varphi(x, \vec{a}) \rightarrow \varphi(S(x), \vec{a})) \rightarrow \forall x \varphi(x, \vec{a}).$$

Peano arithmetic PA can be axiomatized over P1–P10 by the induction schema for arbitrary formulas $\varphi(x, \vec{a})$. One also often considers intermediate fragments of arithmetic. The restriction of the induction schema to Σ_n-formulas $\varphi(x, \vec{a})$ over P1–P10 is denoted $I\Sigma_n$. See Section 10.1 for a more detailed picture of the fragments of PA.

In many respects EA is as good as PA. For example, the usual coding machinery works in EA, and EA is capable to adequately formalize syntax. On the other hand, one can show that, unlike Peano arithmetic, EA is a finitely axiomatizable theory. Moreover, the arithmetical complexity of all axioms of EA is Π_1: all occurrences of universal quantifiers in the induction axioms, except for the outer ones, can be bounded. This property puts severe constraints on the strength of EA. E.g., by a version of a theorem of [Parikh, 1971], EA cannot prove the totality of any computable function that grows faster than $\exp^{(n)}(x)$, for a fixed n (see also [Hájek and Pudlák, 1993]).

The situation is best explained in terms of the notion of *provably total computable function* introduced in Section 10.2. The provably total computable functions of EA are precisely those definable from the basic functions and predicates of our language and projection functions by composition and bounded recursion. Another name for this class of functions is *Kalmar elementary functions* (see [Rose, 1984]). These functions can be conservatively introduced as new function symbols into the language of EA, which parallels the usual process of defining primitive recursive functions in PA. Such a definitional extension respects the arithmetical hierarchy, that is, bounded formulas in the extended language can be equivalently translated into bounded formulas in the original language of EA. Thus, the classes Σ_n and Π_n for $n \geq 0$ are also preserved.

3.2 Formalizing syntax

By a *theory* we shall mean a first order theory with equality formulated in the language of EA and containing the axioms of EA. A theory T is *sound* if all theorems of T are true (hold in the standard model $\mathbb{N}$). T is Σ_n*-sound* if all its theorems of logical complexity Σ_n are true.

Most important for us is the formalization of the notions of proof and provability. Following Gödel and Feferman [Feferman, 1960] this is done in two stages. First, a theory is called *elementary presented*, if a Δ_0-formula $\mathsf{Ax}_T(x)$ is specified that is true if and only if x codes a (non-logical) axiom of T. All the usual theories such as EA or PA are elementary presented; moreover, by the so-called *Craig's trick* one can show that any r.e. theory has an equivalent elementary presentation (see [Feferman, 1960]). From $\mathsf{Ax}_T(x)$ one constructs in a standard way a Δ_0 *proof predicate* $\mathsf{Prf}_T(y, x)$ expressing "y codes a T-proof of the formula coded by x." The corresponding *provability predicate* and *consistency assertion* are then defined by

$$\mathsf{Prov}_T(x) := \exists y\, \mathsf{Prf}_T(y, x) \text{ and } \mathsf{Con}(T) := \neg\mathsf{Prov}_T(\ulcorner \bot \urcorner).$$

The formula $\mathsf{Prov}_T(x)$ satisfies the three *derivability conditions* of Bernays and Löb [Hilbert and Bernays, 1968; Löb, 1955]:

L1. $T \vdash \varphi \iff \mathsf{EA} \vdash \mathsf{Prov}_T(\ulcorner \varphi \urcorner)$

L2. $\mathsf{EA} \vdash \mathsf{Prov}_T(\ulcorner \varphi \to \psi \urcorner) \to (\mathsf{Prov}_T(\ulcorner \varphi \urcorner) \to \mathsf{Prov}_T(\ulcorner \psi \urcorner))$

L3. $\mathsf{EA} \vdash \mathsf{Prov}_T(\ulcorner \varphi \urcorner) \to \mathsf{Prov}_T(\ulcorner \mathsf{Prov}_T(\ulcorner \varphi \urcorner) \urcorner)$

Property L3 is a corollary of a more general fact known as *provable* Σ_1-*completeness*:

PROPOSITION 16.

(i) For any Σ_1*-sentence* σ*,* $\mathsf{EA} \vdash \sigma \to \mathsf{Prov}_T(\ulcorner \sigma \urcorner)$.

(ii) For any Σ_1*-formula* $\sigma(x_1, \ldots, x_n)$ *with all the free variables shown,*

$$\mathsf{EA} \vdash \sigma(x_1, \ldots, x_n) \to \mathsf{Prov}_T(\ulcorner \sigma(\dot{x}_1, \ldots, \dot{x}_n) \urcorner).$$

Here $\ulcorner \sigma(\dot{x}_1, \ldots, \dot{x}_k) \urcorner$ denotes a (Kalmar elementary) definable term for the function that, given a tuple $n_1, \ldots, n_k$, outputs the code $\ulcorner \sigma(\bar{n}_1, \ldots, \bar{n}_k) \urcorner$ of the result of substitution of the numerals $\bar{n}_1$,. . . ,$\bar{n}_k$ for variables $x_1, \ldots, x_k$ in σ.

We will also refer to the following formalized version of the Deduction theorem [Feferman, 1960].

PROPOSITION 17. *For any sentences* φ *and* ψ*,*

$$\mathsf{EA} \vdash \mathsf{Prov}_{T+\varphi}(\ulcorner \psi \urcorner) \leftrightarrow \mathsf{Prov}_T(\ulcorner \varphi \to \psi \urcorner).$$

Here we assume that $T + \varphi$ is elementary presented by the formula $\mathsf{Ax}_{T+\varphi}(x) := \mathsf{Ax}_T(x) \vee x = \ulcorner \varphi \urcorner$. Formalized Deduction theorem can also be formulated and proved with φ and ψ being free variables ranging over sentences.

As a corollary of the derivability conditions one obtains the important Löb theorem [Löb, 1955].

THEOREM 18. $T \vdash \mathsf{Prov}_T(\ulcorner \varphi \urcorner) \to \varphi \iff T \vdash \varphi$.

In view of the formalized Deduction theorem, this statement can be viewed as a version of Gödel's second incompleteness theorem for the theory $T + \neg\varphi$. Vice versa, Gödel's theorem can be obtained from Löb's theorem by setting $\varphi = \bot$. It does not hurt to repeat the most celebrated theorem in mathematical logic, so here it comes.

THEOREM 19 (Gödel). *Let* T *be an elementary presented theory containing* EA.

(i) If T *is consistent, then* $T \nvdash \mathsf{Con}(T)$.

(ii) If T *is* Σ_1*-sound, then also* $T \nvdash \neg\mathsf{Con}(T)$.

3.3 Arithmetical interpretation and its soundness

A mapping from the set of propositional letters to the set of arithmetical sentences is called an (arithmetical) *realization.* Let T be an elementary presented theory. A *T-interpretation* $f_T(\varphi)$ of a modal formula φ under a realization f is defined inductively as follows:

1. $f_T(\bot) = \bot$; $f_T(\top) = \top$;

2. $f_T(p) = f(p)$, for any propositional letter p;

3. $f_T(\theta \to \psi) = f_T(\theta) \to f_T(\psi)$,

4. $f_T(\Box\psi) = \mathsf{Prov}_T(\ulcorner f_T(\psi)\urcorner)$.

The set of formulas $\{f_T(\varphi) : f \text{ a realization}\}$ will be denoted φ^T. We write $U \vdash \varphi^T$ if U proves every formula from the set φ^T. If φ is letterless, φ^T consists of a single formula that will also be denoted φ^T. Similarly, for any set X of modal formulas, X^T will denote the set of all T-interpretations of formulas from X.

The three derivability conditions together with Löb's theorem essentially mean that GL is sound with respect to the arithmetical interpretation.

PROPOSITION 20. *If* $\mathsf{GL} \vdash \varphi$, *then* $\mathsf{EA} \vdash \varphi^T$.

A fundamental theorem of R. Solovay tells us that GL is also complete with respect to the arithmetical interpretation, that is, the converse implication also holds, provided theory T is Σ_1-sound. (In fact, the Solovay theorem holds under some yet weaker assumptions on soundness that will be introduced later.)

Before stating this theorem we would like to discuss some applications of provability logic in arithmetic that only rely on its soundness part. The soundness of GL expressed by Proposition 20 does not seem to be a very deep result. How can it be useful at all?

Modal logic provides a convenient language which, together with Kripke semantics, allows to efficiently *calculate* with certain kinds of arithmetical statements. This could in some sense be compared with what mathematicians do by applying a few simple rules — but sometimes in a very ingenious way — to symbolically compute, say, integrals. Our 'numbers' here are arithmetical sentences, and Löb's theorem is an 'equation' that yields sometimes remarkable unexpected consequences. We do not pursue this line too far right now, but give a few basic examples concerning reflection principles. Deeper applications and further uses of such results will be discussed in Section 10.

4 A MODAL VIEW OF REFLECTION PRINCIPLES

First we introduce a useful notion of *characteristic* of a theory. In some sense, this *characteristic* measures how close the theory is to being inconsistent. In the literature several other names have been used for it, including *rank, credibility extent*, and *height* of a theory. We stick to the present terminology because from the algebraic point of view this notion is a direct analog of the notion of characteristic of a field. The connection will be explained in Section 7.

4.1 Iterated consistency and characteristic

Let T be an elementary presented theory. Extensions of T by *iterated consistency* assertions are defined as follows.

$$T_0 = T, \quad T_{n+1} = T_n + \mathsf{Con}(T_n), \quad T_\omega = \textstyle\bigcup_{n\geq 0} T_n.$$

Notice that all these theories are naturally elementary presented, too.

LEMMA 21. $\mathsf{EA} \vdash \mathsf{Con}(T_n) \leftrightarrow (\neg\Box^{n+1}\bot)^T$.

Proof. By induction on n using Löb's derivability conditions and formalization of Deduction theorem. ■

$\mathsf{Con}(T_n)$ is called *n times iterated consistency assertion for T*. Whenever the initial theory T is Σ_1-sound, the theories T_n form a strictly increasing sequence of Σ_1-sound extensions of T. However, if T is not Σ_1-sound, then it is possible that T_ω is inconsistent. The *characteristic* $ch(T)$ of T is the least $n \in \omega$ such that T_n is inconsistent, if such an n exists, and ∞, otherwise. All Σ_1-sound theories have infinite characteristic. It is not difficult to see that the theory $T + (\Box^n\bot)^T$ has characteristic $n+1$, if T has characteristic ∞. Inconsistent theories and only them have characteristic 0.

4.2 Local and uniform reflection

Some early applications of provability logic in arithmetic concerned the study of reflection principles. Reflection principles were introduced in the 30's as unprovable statements generalizing Gödel's consistency assertions [Rosser, 1936; Turing, 1939]. They have later been used in proof theory for estimating the complexity of axiomatizations of formal theories [Feferman, 1960; Feferman, 1962; Kreisel and Lévy, 1968] and for obtaining conservation results and other kinds of proof-theoretic information [Schmerl, 1979; Beklemishev, 1998b; Beklemishev, 2003a]. Mostly the uniform reflection principles have been used. Provability logic was instrumental in deepening our understanding and finding applications of the local reflection principles.

The *local reflection principle* for T is the schema

$$\mathsf{Rfn}(T): \qquad \mathsf{Prov}_T(\ulcorner\varphi\urcorner) \to \varphi,$$

for all arithmetical sentences φ. The *uniform reflection principle* is the schema

$$\mathsf{RFN}(T): \qquad \forall x_1 \ldots \forall x_n\, (\mathsf{Prov}_T(\ulcorner\varphi(\dot{x}_1, \ldots, \dot{x}_n)\urcorner) \to \varphi(x_1, \ldots, x_n)),$$

for all formulas $\varphi(x_1, \ldots, x_n)$. Both schemata represent different ways of expressing the soundness of T. One cannot formulate the soundness of T as a single arithmetical formula because there is no definition of truth for the whole arithmetical language in the language itself. Σ_n-soundness of T is expressed by restricting of these principles to formulas φ of complexity Σ_n. These restricted schemata are denoted $\mathsf{Rfn}_{\Sigma_n}(T)$ and $\mathsf{RFN}_{\Sigma_n}(T)$, respectively. The corresponding schemata for the classes Π_n are similarly defined.

LEMMA 22. *Over* EA*,*

(i) $\mathsf{RFN}_{\Pi_1}(T) \equiv \mathsf{Con}(T)$*;*

(ii) $\mathsf{RFN}_{\Pi_{n+1}}(T) \equiv \mathsf{RFN}_{\Sigma_n}(T)$*, if* $n \geq 1$*.*

Proof. The inclusions from right to left in each case are clear. For the opposite inclusions reason in EA.

(i) Let $\varphi(x) \in \Pi_1$. If $\mathsf{Prov}_T(\ulcorner\varphi(\dot{x})\urcorner)$ and $\neg\varphi(x)$, then $\mathsf{Prov}_T(\ulcorner\neg\varphi(\dot{x})\urcorner)$, by Σ_1-completeness. Hence, $\mathsf{Prov}_T(\ulcorner\neg\varphi(\dot{x}) \wedge \varphi(\dot{x})\urcorner)$ and $\mathsf{Prov}_T(\ulcorner\bot\urcorner)$.

(ii) Let $\varphi(x, y) \in \Sigma_n$ and $\mathsf{Prov}_T(\ulcorner\forall y\varphi(\dot{x}, y)\urcorner)$. Then, $\forall y\, \mathsf{Prov}_T(\ulcorner\varphi(\dot{x}, \dot{y})\urcorner)$ and using Σ_n-reflection we obtain $\forall y\, \varphi(x, y)$. ■

Restricted uniform schemata $\mathsf{RFN}_{\Sigma_n}(T)$ are finitely axiomatizable over EA, which follows immediately from the existence of partial truth-definitions. We also have the following theorem [Kreisel and Lévy, 1968].

THEOREM 23 (Unboundedness). $\mathsf{Rfn}_{\Sigma_n}(T)$ *is not contained in any consistent r.e. extension of* T *by* Π_n*-sentences.*

Proof. We only prove it for extensions of T by finitely many Π_n-sentences. The general case can be reduced to the finite case by a trick, akin to Rosser's, which we omit. Let π be a Π_n-sentence such that $T + \pi$ is consistent and

$$T + \pi \vdash \mathsf{Rfn}_{\Sigma_n}(T).$$

We have

$$T + \pi \vdash \mathsf{Prov}_T(\ulcorner\neg\pi\urcorner) \to \neg\pi,$$

whence

$$T \vdash \mathsf{Prov}_T(\ulcorner\neg\pi\urcorner) \to \neg\pi,$$

and, by Löb's theorem, $T + \pi$ is inconsistent. ■

REMARK 24. A dual statement holds for $\mathsf{Rfn}_{\Pi_n}(T)$ with a similar proof.

COROLLARY 25. $\mathsf{RFN}_{\Sigma_n}(T)$ *is not contained in* any *consistent extension of* T *by* Σ_{n+1}*-sentences.*

Proof. This follows from finite axiomatizability of $\mathsf{RFN}_{\Sigma_n}(T)$ and the fact that it contains $\mathsf{Rfn}_{\Pi_{n+1}}(T)$, by Lemma 22. ■

4.3 Axiomatization results

Proofs of the results in this section are based on the following lemma from [Beklemishev, 1989a].

LEMMA 26. $\mathbf{GL} \vdash \boxdot\neg \bigwedge_{i=1}^{m}(\Box p_i \to p_i) \to \Box^m\bot$.

Proof. Rather than exhibiting a proof of the formula above we shall argue semantically using the Kripke model characterization of **GL**.

Consider any model $\mathcal{K}$ such that $\mathcal{K} \nVdash \Box^m\bot$. Then there is a sequence of nodes in $\mathcal{K}$ such that

$$r = x_{m+1} \prec x_m \prec \ldots \prec x_1,$$

where r is the root of $\mathcal{K}$. We notice that each formula $\Box p_i \to p_i$ can be false at no more than one node of the chain $x_{m+1}, \ldots, x_1$. Therefore, by the pigeonhole principle, there must exist a node z among the $m+1$ nodes x_i such that

$$z \Vdash \bigwedge_{i=1}^{m}(\Box p_i \to p_i).$$

In case z coincides with $r = x_{m+1}$ we have

$$\mathcal{K} \nVdash \neg \bigwedge_{i=1}^{m}(\Box p_i \to p_i).$$

In case $z = x_i$ for some $i \leq m$, we have $r \prec z$ by transitivity of $\prec$ and thus

$$\mathcal{K} \nVdash \Box(\neg \bigwedge_{i=1}^{m}(\Box p_i \to p_i)).$$

This proves the claim. ■

As our first application we derive the following result found in [Boolos, 1979a] and [Artemov, 1979; Artemov, 1982]. For the sake of readability we write $\Box_T\varphi$ instead of $\mathsf{Prov}_T(\ulcorner\varphi\urcorner)$.

THEOREM 27. $\mathsf{Con}(T_n)$ *is not derivable from* any n *instances of the local reflection schema for* T*, provided* $ch(T) > n$.

Proof. Assume

$$T \vdash \bigwedge_{i=1}^{n}(\Box_T\varphi_i \to \varphi_i) \to \mathsf{Con}(T_n).$$

Then, by Lemma 21 and contraposition

$$T \vdash \Box_T^{n+1}\bot \to \neg \bigwedge_{i=1}^{n}(\Box_T\varphi_i \to \varphi_i),$$

and by the derivability conditions

$$\mathsf{EA} \vdash \Box_T^{n+2}\bot \to \Box_T \neg \bigwedge_{i=1}^{n}(\Box_T\varphi_i \to \varphi_i).$$

By the arithmetical soundness of GL, from Lemma 26 we obtain

$$T \vdash \Box_T^{n+1}\bot \to \Box_T^{n}\bot.$$

By Löb's theorem, $T \vdash \Box_T^{n}\bot$ and T_n is inconsistent. ■

We remark that it is not difficult to find particular $n+1$ instances of the local reflection schema that imply $\mathsf{Con}(T_n)$: one can take all formulas $(F_i)^T$ for $i \leq n$. As a corollary we obtain

THEOREM 28. *Neither* $\mathsf{Rfn}(T)$*, nor any of the schemas* $\mathsf{Rfn}_{\Sigma_n}(T)$ *for* $n \geq 1$ *is finitely axiomatizable over* T*, provided* $ch(T) = \infty$*.*

Proof. By the previous theorem, any m instances of these schemata are insufficient to prove the formula $\mathsf{Con}(T_m)$. However, $\mathsf{Con}(T_m)$ is provable in T_ω, which is already contained in $T + \mathsf{Rfn}_{\Sigma_1}(T)$. ■

4.4 Conservation results

Another striking property of local reflection principles is the following theorem.

THEOREM 29. T *together with any* n *instances of local reflection principle for* T *is* Π_1*-conservative over* T_n*.*

Proof. Let $\pi \in \Pi_1$ and

$$T \vdash \bigwedge_{i=1}^{n}(\Box_T\varphi_i \to \varphi_i) \to \pi.$$

Then, by the derivability conditions,

$$T \vdash \boxdot_T \neg\pi \to \boxdot_T \neg \bigwedge_{i=1}^{n}(\Box_T\varphi_i \to \varphi_i).$$

From Lemma 26 we obtain $T \vdash \boxdot_T\neg\pi \to \Box_T^n\bot$, whence $T \vdash \neg\pi \to \Box_T^n\bot$, by provable Σ_1-completeness. Thus $T \vdash \neg\Box_T^n\bot \to \pi$ and $T_n \vdash \pi$. ■

The following immediate corollary is known as Goryachev's Theorem [Goryachev, 1986].

THEOREM 30 (Goryachev). $T + \mathsf{Rfn}(T)$ *and* T_ω *prove the same* Π_1*-sentences.*

By Goryachev's theorem all schemata $\mathsf{Rfn}_{\Sigma_n}(T)$, in contrast with the uniform reflection principles, prove the same Π_1-sentences. By [Beklemishev, 1997b] an even stronger conservation result holds.

THEOREM 31. *Over any elementary presented theory T,*

(i) $\mathsf{Rfn}(T)$ *and* $\mathsf{Rfn}_{\Sigma_1}(T)$ *prove the same boolean combinations of Σ_1-sentences.*

(ii) For $n > 1$, $\mathsf{Rfn}(T)$ *and* $\mathsf{Rfn}_{\Sigma_n}(T)$ *prove the same Σ_n-sentences.*

Proof. This requires a generalization of Lemma 26. Let modal formulas Q_i be defined as follows:

$$Q_1 = p, \qquad Q_{i+1} = Q_i \vee \Box Q_i,$$

where p is a propositional variable.

LEMMA 32. $\mathsf{GL} \vdash \boxdot(\bigwedge_{i=1}^{m}(\Box p_i \to p_i) \to p) \to (\bigwedge_{i=1}^{m}(\Box Q_i \to Q_i) \to p)$.

A proof is rather similar to the proof of Lemma 26, so we omit it. Also notice that Lemma 26 follows from Lemma 32, if one substitutes $\bot$ for p.

Theorem 31 is now proved as follows. Assume

$$T \vdash \textstyle\bigwedge_{i=1}^{m}(\Box_T \varphi_i \to \varphi_i) \to \pi.$$

By derivability conditions,

$$T \vdash \boxdot_T(\textstyle\bigwedge_{i=1}^{m}(\Box_T \varphi_i \to \varphi_i) \to \pi).$$

Considering an arithmetical realization f that maps the variable p to the sentence π and p_i to φ_i, by Lemma 32 we conclude that

$$T \vdash \textstyle\bigwedge_{i=1}^{m}(\Box_T \psi_i \to \psi_i) \to \pi,$$

where ψ_i denotes $f_T(Q_i)$. Now we observe that if $\pi \in \Sigma_n$ for $n > 1$, then $\psi_i \in \Sigma_n$, for all i. Hence,

$$T + \mathsf{Rfn}_{\Sigma_n}(T) \vdash \pi,$$

which proves (ii).

Let $\mathcal{B}(\Sigma_1)$ be the set of boolean combinations of Σ_1-sentences. By a similar argument, $\mathcal{B}(\Sigma_1)$ consequences of $\mathsf{Rfn}(T)$ are contained in $\mathsf{Rfn}_{\mathcal{B}(\Sigma_1)}(T)$. It is not difficult to verify using provable Σ_1-completeness that $\mathsf{Rfn}_{\mathcal{B}(\Sigma_1)}(T)$ is equivalent to $\mathsf{Rfn}_{\Sigma_1}(T)$. ■

These conservation results have been used in [Beklemishev, 1999b] to characterize the classes of provably total computable functions of fragments of Peano arithmetic with parameter-free induction and to solve some other problems in this area. See Section 10 for more details.

5 SOLOVAY THEOREMS

A central result in provability logic is the following theorem.

THEOREM 33 (Solovay). *Assume $ch(T)=\infty$. Then $\mathsf{GL}\vdash\varphi$ iff $T\vdash\varphi^T$.*

For the proof of this theorem R. Solovay [Solovay, 1976] invented the techniques of "embedding" Kripke models into arithmetic, which is currently known under the name *Solovay construction*. Variants and generalizations of this construction have been applied to obtain arithmetical completeness results for various logics with provability and interpretability semantics. We are going to describe this important construction below.

Let T be an elementary presented theory and $\mathcal{K}=(K,\prec,\Vdash)$ a finite Kripke model. We assume without loss of generality that $K=\{0,\ldots,n\}$ and 0 is the root of $\mathcal{K}$. An elementary function $h(x)$ can be defined with the aid of the arithmetical fixed point theorem to satisfy the following equations provably in EA:

$$\begin{array}{rcl} h(0) & = & 0; \\ h(m+1) & = & \left\{\begin{array}{ll} z, & \text{if } z\in K,\ h(m)\prec z \text{ and } \mathsf{Prf}_T(m,\ulcorner\ell\neq\bar{z}\urcorner); \\ h(m), & \text{otherwise.}\end{array}\right.\end{array}$$

Here $\ell=z$ denotes the arithmetical formula $\exists m\,\forall n>m\ h(n)=z$ and $\ell\neq z$ is $\neg(\ell=z)$. Formally speaking, a Δ_0-formula expressing the relation $h(x)=y$ is defined in terms of its own Gödel number (from which the Gödel number of $\ell=z$ is obtained in an elementary way).

Informally, the behavior of the function h can be illustrated by the following story.[9] Imagine a refugee who is admitted from one country to another only if he/she provides a proof not to stay there forever. If the refugee is also never allowed to go to one of the previously visited countries, he/she must eventually stop somewhere. So, an honest refugee will never be able to leave his/her country of origin... Think about $h(m)=z$ as the statement that the refugee is in country z at the moment m. Some basic facts about the moves of the refugee are expressed by the following lemma.

LEMMA 34. *The following statements are provable in EA:*

(i) $\bigvee_{z\in K}\ell=\bar{z}$;

(ii) $\forall u,v\ (\ell=u\wedge\ell=v\rightarrow u=v)$;

(iii) $\ell=\bar{z}\rightarrow\mathsf{Prov}_T(\ulcorner\bigvee_{w\succ z}\ell=\bar{w}\urcorner)$, *if $z\in K$ and $z\succ 0$;*

(iv) $\ell=\bar{z}\rightarrow\neg\mathsf{Prov}_T(\ulcorner\ell\neq\bar{u}\urcorner)$, *if $z,u\in K$ and $u\succ z$.*

[9] We were not able to trace the origins of this story, which seems to belong to the lore of provability logic.

Proof. Statements (i) and (ii) follow from the fact that (provably in EA) values of h belong to K and h is weakly increasing in the sense of the ordering $\preceq$.

To prove (iii) we reason within EA as follows:

> If $\ell = z$, then for some m, $h(m) = z$. By Σ_1-completeness, $T \vdash \exists m\ h(m) = \bar{z}$. Since h is provably monotone, we have $T \vdash \exists m \forall n > m\ h(n) \succeq \bar{z}$ and hence $T \vdash \bigvee_{w \succeq z} \ell = \bar{w}$.
>
> On the other hand, if $\ell = z$ and $z \succ 0$, then there is the least m such that $h(m+1) = z$, and by the definition of h this implies $T \vdash \ell \neq \bar{z}$. Thus, we obtain $T \vdash \bigvee_{w \succ z} \ell = \bar{w}$.

To prove (iv) we formalize the following argument in EA:

> If $\ell = z$ and $T \vdash \ell \neq \bar{u}$, where $u \succ z$, then for a sufficiently large m there holds $\forall k \geq m\ h(k) = z$ and $\mathsf{Prf}_T(m, \ulcorner \ell \neq \bar{u} \urcorner)$. But then, by the definition of h, one has $h(m+1) = u$. Since h is weakly increasing this implies $\ell \neq z$, a contradiction.

This completes the proof of Lemma 34. ∎

We call *Solovay realization* the following function:

$$f(p) = \bigvee_{z \in K,\ z \Vdash p} \ell = \bar{z}.$$

LEMMA 35. *For all formulas φ and for all $z \in K,\ z \succ 0$,*

(i) If $z \Vdash \varphi$, then $\mathsf{EA} \vdash \ell = \bar{z} \to f_T(\varphi)$;

(ii) If $z \nVdash \varphi$, then $\mathsf{EA} \vdash \ell = \bar{z} \to \neg f_T(\varphi)$.

Proof. Statements (i) and (ii) are proved simultaneously by induction on φ. We consider the most important case when φ has the form $\Box\psi$.

(i) If $z \Vdash \Box\psi$, then $\forall u \succ z\ u \Vdash \psi$. Hence, by the induction hypothesis,

$$\mathsf{EA} \vdash \bigvee_{u \succ z} \ell = \bar{u} \to f_T(\psi).$$

Using Lemma 34 (iii) we then obtain

$$\begin{aligned} \mathsf{EA} \vdash \ell = \bar{z} \quad &\to \quad \mathsf{Prov}_T(\ulcorner \textstyle\bigvee_{u \succ z} \ell = \bar{u} \urcorner) \\ &\to \quad \mathsf{Prov}_T(\ulcorner f_T(\psi) \urcorner) \\ &\to \quad f_T(\Box\psi). \end{aligned}$$

(ii) If $z \nVdash \Box\psi$, then $\exists u \succ z\ u \nVdash \psi$. By the induction hypothesis

$$\mathsf{EA} \vdash \ell = \bar{u} \to \neg f_T(\psi),$$

whence

$$\mathsf{EA} \vdash \neg\mathsf{Prov}_T(\ulcorner \ell \neq \bar{u} \urcorner) \to \neg\mathsf{Prov}_T(\ulcorner f_T(\psi) \urcorner).$$

Using Lemma 34 (iv) we obtain

$$\begin{array}{rcl} \mathsf{EA} \vdash \ell = \bar{z} & \to & \neg\mathsf{Prov}_T(\ulcorner \ell \neq \bar{u} \urcorner) \\ & \to & \neg\mathsf{Prov}_T(\ulcorner f_T(\psi) \urcorner) \\ & \to & \neg f_T(\Box\psi). \end{array}$$

■

Proof of Theorem 33. If $\mathsf{GL} \nvdash \varphi$, then there is a finite Kripke model $\mathcal{K}_0$ such that $\mathcal{K}_0 \nVdash \varphi$. We may assume that $K_0 = \{1, \dots, n\}$ and 1 is the root of $\mathcal{K}_0$. We extend $\mathcal{K}_0$ by a new node 0 stipulating $K = K_0 \cup \{0\}$ and $0 \prec z$, for all $z \in K_0$. The forcing of propositional variables is defined at 0 arbitrarily.

Applying Solovay construction to $\mathcal{K}$ yields an arithmetical realization f for which

$$\mathsf{EA} \vdash \ell = 1 \to \neg f_T(\varphi).$$

If $T \vdash f_T(\varphi)$ we then would have $T \vdash \ell \neq 1$. By Lemma 34 (iv) this implies $\ell \neq 0$, so $\ell = \bar{z}$ is true, for some $z \in K_0$. For $n = d(z)$ we have $z \Vdash \Box^{n+1}\bot$ and therefore $\mathsf{EA} \vdash \ell = \bar{z} \to (\Box^{n+1}\bot)^T$. Hence, $(\Box^{n+1}\bot)^T$ is true, that is $ch(T) \leq n$, a contradiction. ■

Solovay's theorem characterizes modal schemata provable in a theory T. What about the modal schemata true in the standard model of arithmetic? The answer is provided by so-called second Solovay theorem and the logic S. The system S is defined as the closure of GL together with the (modal) reflection axiom $\Box\varphi \to \varphi$ under modus ponens. Necessitation rule is not admissible in S because one can easily derive a contradiction from Löb's axiom and the necessitated reflection axiom: $\Box\bot \to \bot$, $\Box(\Box\bot \to \bot)$, $\Box\bot$, $\bot$.

Let $S(\varphi)$ denote the following modal formula:

$$\textstyle\bigwedge_{i=1}^{n}(\Box\varphi_i \to \varphi_i),$$

where $\Box\varphi_1, \dots, \Box\varphi_n$ enumerate all subformulas of φ of the form $\Box\psi$. A node x in a model $\mathcal{K}$ is called *φ-reflexive*, if $x \Vdash S(\varphi)$.

THEOREM 36 (Solovay, II). *Let T be a sound theory. The following statements are equivalent:*

(i) $\mathsf{S} \vdash \varphi$;

(ii) $\mathsf{GL} \vdash S(\varphi) \to \varphi$;

(iii) $\mathbb{N} \vDash \varphi^T$.

Proof. Implication (ii)⇒(i) is obvious. (i)⇒(iii) follows at once from the soundness of T. We prove (iii)⇒(ii) by contraposition. Assume $\mathsf{GL} \nvdash S(\varphi) \to \varphi$. Then there is a finite model $\mathcal{K}$ with $K = \{0, \dots, n\}$ and the root 0 such that $\mathcal{K} \Vdash S(\varphi)$ and $\mathcal{K} \nVdash \varphi$. Apply the Solovay construction to $\mathcal{K}$. Lemmas 34, 35 obviously carry through. In addition we have

LEMMA 37. *For any subformula ψ of the formula φ there holds:*

(i) If $0 \Vdash \psi$, then $\mathsf{EA} \vdash \ell = 0 \to f_T(\psi)$;

(ii) If $0 \nVdash \psi$, then $\mathsf{EA} \vdash \ell = 0 \to \neg f_T(\psi)$.

Proof. We argue by induction on ψ. Statement (ii) is proved similarly to Lemma 35 (ii). For the proof of (i) notice that if ψ has the form $\Box\theta$ and $0 \Vdash \psi$, then $\forall x \in K\ x \Vdash \theta$, by the φ-reflexivity of the node 0. Hence, by the induction hypothesis and Lemma 35 (i),

$$\mathsf{EA} \vdash \left(\textstyle\bigvee_{u \in K} \ell = \bar{u}\right) \to f_T(\theta).$$

By Lemma 34 (i), $\mathsf{EA} \vdash f_T(\theta)$ and therefore

$$\mathsf{EA} \vdash \ell = 0 \to \mathsf{Prov}_T(\ulcorner f_T(\theta) \urcorner).$$

■

From this lemma we conclude that

$$\mathsf{EA} \vdash \ell = 0 \to \neg f_T(\varphi).$$

Since T is sound, $\ell = 0$ holds in the standard model (an honest refugee does not leave the country of origin). Therefore $f_T(\varphi)$ is false, a contradiction. ■

COROLLARY 38. S *is decidable.*

6 CLASSIFICATION OF PROVABILITY LOGICS

6.1 *Provability logics relative to a theory*

Provability logic aims at describing valid laws of provability in a given system T. However, Gödel's incompleteness theorems put a significant constraint: the answer to such a question is not unique. In general, one should distinguish between the *theory* T under study and the *metatheory* U, in which one reasons about the properties of T. Different metatheories verify different properties of T. Perhaps, the most natural choice of U is the *true arithmetic* TA, which is axiomatized by the set of all sentences true in the standard model $\mathbb{N}$. Other possible choices are: T itself, EA, PA, etc. This naturally leads to the notion of *provability logic of a theory T relative to a metatheory U* that was suggested independently by S.N. Artëmov [Artemov, 1980] and A. Visser [Visser, 1981].

Let U be an arbitrary theory containing EA (not necessarily elementary presented or even r.e.) and let T be an elementary presented theory. The provability logic of T relative to U is the set $\boldsymbol{PL}_T(U)$ of all modal formulas φ such that $U \vdash \varphi^T$. Intuitively, $\boldsymbol{PL}_T(U)$ axiomatizes those principles of provability in T that can be verified by means of U. $\boldsymbol{PL}_T(T)$ is sometimes called *the* provability logic of T, and $\boldsymbol{PL}_T(\mathsf{TA})$ is called the *truth provability logic of T*. Solovay's theorems can be restated as saying that, if T is a sound theory, then $\boldsymbol{PL}_T(T) = \mathsf{GL}$ and $\boldsymbol{PL}_T(\mathsf{TA}) = \mathsf{S}$. In general, $\boldsymbol{PL}_T(U)$ is a (not necessarily normal) modal logic extending GL, for it is closed under *modus ponens* and substitution rules.

A modal logic L is called a *provability logic* if $L = \boldsymbol{PL}_T(U)$ for some T and U. A somewhat older term for the same notion, introduced by S.N. Artëmov [Artemov, 1980], is *arithmetically complete modal logic.* We explain this terminology in the following paragraph.

6.2 Inference by arithmetical interpretation

Arithmetical interpretation w.r.t. a theory T induces a natural consequence relation on the set of modal formulas. Let Γ be a set of modal formulas. Write $\Gamma \vdash^*_T \varphi$ if $\mathsf{EA} + \Gamma^T \vdash \varphi^T$, that is, if every T-interpretation of φ follows from T-interpretations of formulas from Γ. Notice that $\Gamma \vdash^*_T$ is closed under modus ponens and substitution, but, in general, not under the necessitation rule.

By Solovay's first theorem, $\varnothing \vdash^*_T \varphi$ iff $\mathsf{GL} \vdash \varphi$, and by Solovay's second theorem

$$\{\Box p \to p\} \vdash^*_T \varphi \iff \mathsf{S} \vdash \varphi \iff \{\Box p \to p\} \vdash_{\mathbf{GL},\mathrm{sub}} \varphi.$$

Here the relation $\Gamma \vdash_{\mathbf{GL},\mathrm{sub}} \varphi$ means that φ follows from Γ and axioms of GL by modus ponens and substitution rules. Later we will see that the arithmetical consequence relation $\vdash^*_T$ is, in general, much stronger than $\vdash_{\mathsf{GL},\mathrm{sub}}$.

A logic L is called *T-complete*, if it is closed under $\vdash^*_T$:

$$L \vdash^*_T \varphi \iff L \vdash \varphi,$$

for all formulas φ. The *T-completion* $[L]^T$ of L is the minimal T-complete logic containing L. The following proposition shows that T-complete logics are precisely the provability logics for T.

PROPOSITION 39. *L is T-complete iff $L = \boldsymbol{PL}_T(U)$ for some U.*

Proof. It is easy to see that logics of the form $\boldsymbol{PL}_T(U)$ are T-complete. Conversely, if L is T-complete, then $L = \boldsymbol{PL}_T(\mathsf{EA} + L^T)$. ■

6.3 Classification theorem

One of the early questions in the field of provability logic was the so-called classification problem, that is, the problem of characterizing all possible provability logics within the lattice of extensions of GL. By Proposition 39 this problem is equivalent to the question of characterizing the arithmetical consequence relation $\vdash^*_T$.

The solution to this problem is the outcome of the work of several authors: S.N. Artëmov [Artemov, 1980; Artemov, 1985b], A. Visser [Visser, 1981; Visser, 1984], G. Japaridze [Japaridze, 1986], L.D. Beklemishev [Beklemishev, 1989a]. For a set of modal formulas X, let LX denote the closure of X and all theorems of a logic L under modus ponens and substitution rules. S.N. Artëmov [Artemov, 1980] showed that any logic of the form $\mathsf{GL}X$, where X is a set of letterless formulas, is a provability logic. In [Artemov, 1985b] he showed that such extensions are exhausted by the following two specific families of logics:

$$\mathsf{GL}_\alpha = \mathsf{GL}\{F_n : n \in \alpha\}, \qquad \mathsf{GL}^-_\beta = \mathsf{GL}\{\textstyle\bigvee_{n \notin \beta} \neg F_n\},$$

where $\alpha, \beta \subseteq \omega$ and β is cofinite.

The families GL_α and GL^-_β are ordered by inclusion precisely as their indices, and GL_α is included in GL^-_α for cofinite α. The logics GL^-_β are not contained in S and therefore correspond to unsound metatheories U, if T is sound. A. Visser [Visser, 1984] showed that GL^-_β are the only provability logics not contained in S. S.N. Artëmov [Artemov, 1985b] reduced the classification problem to the interval between GL_ω and S.

G. Japaridze [Japaridze, 1986; Japaridze, 1988] found a new provability logic D within this interval:

$$\mathsf{D} = \mathsf{GL}\{\neg\Box\bot, \Box(\Box\varphi \vee \Box\psi) \to (\Box\varphi \vee \Box\psi)\}.$$

He showed that $\mathsf{D} = \boldsymbol{PL}_{\mathsf{PA}}(\mathsf{PA} + \omega\text{-}\mathsf{Con}(\mathsf{PA}))$, where $\omega\text{-}\mathsf{Con}(\mathsf{PA})$ denotes a formalization of ω-consistency of Peano arithmetic.

As the final step, L.D. Beklemishev [Beklemishev, 1989a] showed that D is the only provability logic within the interval between GL_ω and S. This completed the classification of provability logics. We denote $\mathsf{S}_\beta = \mathsf{S} \cap \mathsf{GL}^-_\beta$, $\mathsf{D}_\beta = \mathsf{D} \cap \mathsf{GL}^-_\beta$ and formulate the resulting Classification theorem [Beklemishev, 1989a].

THEOREM 40 (Classification theorem). *The provability logics are exhausted by the four families:* GL_α, GL^-_β, S_β *and* D_β, *for* $\alpha, \beta \subseteq \omega$, β *cofinite.*

Each of these logics is T-complete for any elementary presented theory T of infinite characteristic.

L.D. Beklemishev [Beklemishev, 1989a] also characterized all possible truth provability logics.

COROLLARY 41. *The truth provability logics are precisely the following ones:*

$$\mathsf{S},\ \mathsf{D},\ \mathsf{GL}_\omega,\ \textit{and}\ \mathsf{GL}\{\neg F_n\},\ n \in \omega.$$

Moreover, for any elementary presented theory T,

(i) $\boldsymbol{PL}_T(\mathsf{TA}) = \mathsf{S}$ *iff T is sound;*

(ii) $\boldsymbol{PL}_T(\mathsf{TA}) = \mathsf{D}$ *iff T is Σ_1-sound but not sound;*

(iii) $\boldsymbol{PL}_T(\mathsf{TA}) = \mathsf{GL}_\omega$ *iff T is not Σ_1-sound but $ch(T) = \infty$;*

(iv) $\boldsymbol{PL}_T(\mathsf{TA}) = \mathsf{GL}\{\neg F_n\}$ *iff $ch(T) = n$ (for $n < \infty$).*

Proof. If $ch(T) = n < \omega$, then formula $(\neg F_n)^T$ is true, hence $\boldsymbol{PL}_T(\mathsf{TA}) \supseteq \mathsf{GL}\{\neg F_n\}$. But $\mathsf{GL}\{\neg F_n\}$ is a maximal logic among consistent provability logics, so $\boldsymbol{PL}_T(\mathsf{TA}) = \mathsf{GL}\{\neg F_n\}$. If $ch(T) = \infty$, then all formulas F_n^T are true, therefore $\boldsymbol{PL}_T(\mathsf{TA}) \supseteq \mathsf{GL}_\omega$. By Classification theorem, there are only three consistent provability logics containing GL_ω: GL_ω, D or S. If T is Σ_1-sound, clearly $\boldsymbol{PL}_T(\mathsf{TA}) \supseteq \mathsf{D}$. Corollary 52 (i) below implies that $\mathsf{EA} + \mathsf{D}^T \vdash \mathsf{Rfn}_{\Sigma_1}(T)$, hence $\boldsymbol{PL}_T(\mathsf{TA}) \supseteq \mathsf{D}$ if and only if T is Σ_1-sound. Together with the obvious (i) this proves (ii) and (iii). ■

For the sake of completeness we also formulate a rudimentary variant of the Classification theorem for theories T of finite characteristic. This incorporates a result of A. Visser [Visser, 1981; Visser, 1984] describing the set of provability logics of the form $\boldsymbol{PL}_T(T)$, for such theories T.

COROLLARY 42. *Let $ch(T) = n < \infty$. Then*

$$\begin{aligned} \boldsymbol{PL}_T(\mathsf{EA}) &= \mathsf{GL}\{\Box^{n+1}\bot\}, \\ \boldsymbol{PL}_T(T) &= \mathsf{GL}\{\Box^{n}\bot\}, \end{aligned}$$

and T-complete logics are precisely the logics GL^-_α for $\omega \setminus \alpha \subseteq \{0, \dots, n\}$.

This statement easily follows from the Classification theorem and Statement (iv) of the previous corollary.

Finally, we mention an important corollary of the proof of the Classification theorem that will be discussed below.

COROLLARY 43. *The consequence relation $\Gamma \vdash^*_T \varphi$, as a relation between a finite set of formulas Γ and a formula φ, is decidable. Moreover, for any such Γ one can effectively find a formula Γ^* such that for any φ,*

$$\Gamma \vdash^*_T \varphi \iff \Gamma^* \vdash_{\mathsf{GL},\mathrm{sub}} \varphi.$$

As such a formula Γ^* one can take the axiom of the logic $[\Gamma]^T$ that happens to be finitely axiomatizable for finite Γ.

REMARK 44. The relation $\Gamma \vdash_{\mathsf{GL},\mathrm{sub}} \varphi$ is undecidable. This follows from the existence of a finitely axiomatizable undecidable logic extending GL (see [Chagrov *et al.*, 2001]). From the Classification theorem we conclude that all finitely axiomatizable provability logics are decidable.

6.4 Proof of the Classification theorem

A full proof of the Classification theorem would exceed the limits of this survey, so we shall skip some more technical parts. The missing details can be found in the dissertation of the second author translated by AMS in [Beklemishev *et al.*, 1999].

The proof roughly falls into three main steps, which use different techniques and ideas. Step 1 is the techniques of traces developed by S.N. Artëmov, which allows to reduce the Classification problem to the interval of logics between GL_ω and S. Step 2 is the result that there are no provability logics between D and S, which is mostly based on Kripke models for D and their characteristic formulas. Finally, Step 3 is the fact that there are no provability logics between GL_ω and D, which is based on a modification of the Solovay construction.

Classification by traces

Recall the definition of trace of a formula from Section 2.7. The *trace* $tr(L)$ *of a modal logic* L extending GL is the union of traces of all theorems of L. It is not difficult to see that $tr(\mathsf{GL}_\alpha) = tr(\mathsf{GL}_\alpha^-) = \alpha$ and $tr(\mathsf{S}) = tr(\mathsf{D}) = \omega$.

LEMMA 45. *If* α *is coinfinite, then* GL_α *is the strongest logic with trace* α. *If* α *is cofinite, then the strongest logic with trace* α *is* GL_α^-.

Proof. Let L be a logic, $tr(L) = \alpha$, and α coinfinite. If $L \vdash \varphi$, then $tr(\varphi) \subseteq \alpha$, and therefore, by Lemma 12, $tr(\varphi)$ is finite. By Lemma 13,

$$\mathsf{GL} \vdash \left(\bigwedge\nolimits_{n\in tr(\varphi)} F_n\right) \to \varphi,$$

whence $\mathsf{GL}_\alpha \vdash \varphi$. So we have proved that $L \subseteq \mathsf{GL}_\alpha$.

Now assume that α is cofinite. Then $L \subseteq \mathsf{GL}_\alpha^-$ because by Lemma 13 the letterless formula $\bigvee_{n\notin\alpha} \neg F_n$, whose trace is α, implies all theorems of L. ■

Now we show that, if $\alpha \subseteq \omega$ is coinfinite, then the only *provability logic* with trace α is GL_α. If, on the other hand, α is cofinite, then any provability logic with trace α either coincides with GL_α^-, or is contained in the interval between GL_α and S_α. These results are based on the following lemma.

LEMMA 46. *Let T be an elementary presented theory and φ a modal formula. If $n \in \mathrm{tr}(\varphi)$, then there exists a realization f such that*

$$\mathsf{EA} \vdash f_T(\varphi) \to (F_n)^T.$$

Proof. This lemma is a direct application of the Solovay construction. Let $n \in tr(\varphi)$ and $\mathcal{K}_0$ be a model of height n falsifying φ, where $K_0 = \{1, \ldots, m\}$ and 1 is the root of $\mathcal{K}_0$. As in the proof of Solovay's theorem, attach a new root 0 to $\mathcal{K}_0$. Obviously, in the new model $\mathcal{K}$ the node 1 is the only node of depth n, $d(0) = n + 1$, and $1 \nVdash \varphi$.

Apply the Solovay construction to $\mathcal{K}$ and let f be the corresponding realization. By Lemma 35,

$$\mathsf{EA} \vdash \ell = 1 \to \neg f_T(\varphi).$$

Since 1 is the only node of depth n, it is not difficult to see that

$$\mathsf{EA} \vdash (\neg F_n)^T \to \ell = 1,$$

which proves the lemma. ■

From this lemma we obtain the following corollaries.

COROLLARY 47. *If L is a provability logic and $n \in \mathrm{tr}(L)$, then $L \vdash F_n$.*

COROLLARY 48. *If L is a provability logic, $\mathrm{tr}(L) = \alpha$, and α is coinfinite, then $L = \mathsf{GL}_\alpha$.*

Proof. $L \subseteq \mathsf{GL}_\alpha$, by Lemma 45; $\mathsf{GL}_\alpha \subseteq L$, by Corollary 47. ■

In a similar manner the following result is obtained.

LEMMA 49. *Let L be a provability logic such that $L \nsubseteq \mathsf{S}$. Then $tr(L) = \alpha$ is cofinite and $L = \mathsf{GL}^-_\alpha$.*

Proof. Let $\alpha = tr(L)$. If α is coinfinite, then GL_α is the strongest logic with trace α. But $\mathsf{GL}_\alpha \subseteq \mathsf{S}$, therefore $L \subseteq \mathsf{S}$. Hence, α is cofinite and $L \subseteq \mathsf{GL}^-_\alpha$.

Let us prove that $L \supseteq \mathsf{GL}^-_\alpha$. Since $L \nsubseteq \mathsf{S}$, there is a formula φ such that $L \vdash \varphi$, but $\mathsf{S} \nvdash \varphi$. Clearly, in this case $\mathsf{GL} \nvdash S(\varphi) \to \varphi$, hence there is a model $\mathcal{K}$ with a root 0 such that $0 \nVdash \varphi$ and the node 0 is φ-reflexive.

Set

$$\psi = \varphi \wedge \bigwedge_{n<h(\mathcal{K}),\ n\in\alpha} F_n.$$

By Corollary 47, $L \vdash \psi$. Apply Solovay's construction to the model $\mathcal{K}$ and let f be the corresponding realization. From the properties of the Solovay realization it is possible to conclude that

$$\mathsf{EA} \vdash f_T(\psi) \to \left(\bigvee\nolimits_{k\notin\alpha} \neg F_k\right)^T. \qquad (1)$$

Now assume that $L = \boldsymbol{PL}_T(U)$, for some theories T and U. Since $L \vdash \psi$, for any realization f we have $U \vdash f_T(\psi)$. In particular, this holds for the Solovay realization. Then, by (1), we obtain

$$U \vdash \left(\bigvee_{k \notin \alpha} \neg F_k\right)^T,$$

whence $L \vdash \bigvee_{k \notin \alpha} \neg F_k$. ■

COROLLARY 50. *Any consistent provability logic L of trace ω satisfies*

$$\mathsf{GL}_\omega \subseteq L \subseteq \mathsf{S}.$$

Finally, we reduce the classification problem to the interval between GL_ω and S. Notice that the intersection of any two T-complete logics $L_1 \cap L_2$ is T-complete. Moreover, if $tr(L) = \omega$, then $tr(L \cap \mathsf{GL}^-_\alpha) = \alpha$. So, the function

$$\eta_\alpha : L \longmapsto L \cap \mathsf{GL}^-_\alpha$$

maps provability logics of trace ω to those of (cofinite) trace α. It turns out that η_α is a bijection. The inverse mapping is given by the formula

$$\xi_\alpha : L \longmapsto L\{F_n : n \notin \alpha\}.$$

The fact that η_α and ξ_α are mutually inverse can be inferred from Corollary 47 and Lemma 45. Together with Corollary 50 this implies that we only have to describe the provability logics in the interval between GL_ω and S.

Logics between GL_ω and D

We formulate the main lemma and its corollary right away.

LEMMA 51. *Let T be an elementary presented theory and $\mathsf{GL}_\omega \nvdash \varphi$. Then for any arithmetical Σ_1-sentence σ there is a realization f such that*

$$\mathsf{EA} + \{\neg\Box^n\bot : n \in \omega\}^T + f_T(\varphi) \vdash \mathsf{Prov}_T(\ulcorner\sigma\urcorner) \to \sigma.$$

COROLLARY 52. *If $\mathsf{GL}_\omega \nvdash \varphi$, then*

(i) $\mathsf{EA} + \{\neg\Box^n\bot : n \in \omega\}^T + \varphi^T \vdash \mathsf{Rfn}_{\Sigma_1}(T)$*;*

(ii) $\mathsf{GL}_\omega\{\varphi\} \vdash^*_T \mathsf{D}$.

Proof. Statement (i) is just a weakening of Lemma 51; (ii) follows from (i) because arithmetical interpretations of the axioms of D are instances of local Σ_1-reflection principle. ■

Proof of Lemma 51. We apply a modification of the Solovay construction. Assume $\mathsf{GL}_\omega \nvdash \varphi$. Then, by Lemma 26,

$$\mathsf{GL} \nvdash \Diamond S(\varphi) \to \varphi.$$

Hence, there is a model $\mathcal{K}_0$ with $K_0 = \{1, \dots, k\}$ and the root 1 such that there is a φ-reflexive node $r \succ 1$ and $\mathcal{K}_0 \nVdash \varphi$. As usual, consider a model $\mathcal{K}$ obtained by adding to $\mathcal{K}_0$ a new root 0.

Let σ be an arbitrary Σ_1-sentence. We may assume that σ has the form $\exists x \sigma_0(x)$, where σ_0 is a Δ_0-formula. Define an elementary function h as follows:

$$\begin{array}{rcl} h(0) & = & 0; \\ h(m+1) & = & \begin{cases} z, & \text{if } \mathsf{Prf}_T(m, \ulcorner \ell \neq \bar{z} \urcorner),\ z \succ h(m) \text{ and } z \neq r; \\ \text{otherwise} & \begin{cases} r, & \text{if } h(m) = 1 \text{ and } \exists x \leq m\ \sigma_0(x); \\ h(m), & \text{otherwise.} \end{cases} \end{cases} \end{array}$$

Here $\ell = z$ denotes, as usual, the formula expressing $\lim_{m \to \infty} h(m) = z$. Informally speaking, among various countries visited by the refugee there is now a dangerous one, designated by node 1 and a friendly one, designated by node r. If the alarm rings ($\exists x \leq m\ \sigma_0(x)$ is true) while the refugee resides the dangerous country, he/she has to leave immediately. In this case the refugee goes directly to the friendly country r. In all other cases he/she behaves as in the original Solovay construction.

We define a realization f as before:

$$f(p) = \bigvee_{z \in K,\ z \Vdash p} \ell = \bar{z}.$$

LEMMA 53. *For all $z \in K$, $z \succ 0$, and all subformulas ψ of the formula φ there holds:*

(i) If $z \Vdash \psi$, then $\mathsf{EA} \vdash \ell = \bar{z} \to f_T(\psi)$;

(ii) If $z \nVdash \psi$, then $\mathsf{EA} \vdash \ell = \bar{z} \to \neg f_T(\psi)$.

Proof. This lemma is proved very similarly to Lemma 35 by induction on ψ. We essentially rely on the fact that in the case of emergency the refugee jumps to a φ-reflexive node. Let $\psi = \Box\theta$.

(i) If $z \Vdash \Box\theta$ and $z \neq r$, reason as in Lemma 35. If $z = r$, we have $\forall u \succ z\ u \Vdash \theta$, but also $z \Vdash \theta$ by φ-reflexivity of r. Hence, by induction hypothesis,

$$\mathsf{EA} \vdash (\textstyle\bigvee_{u \succeq r} \ell = \bar{u}) \to f_T(\theta)$$

and

$$\mathsf{EA} \vdash \mathsf{Prov}_T(\ulcorner \textstyle\bigvee_{u \succeq r} \ell = \bar{u} \urcorner) \to \mathsf{Prov}_T(\ulcorner f_T(\theta) \urcorner).$$

Therefore, by the provable monotonicity of h,

$$\begin{aligned}\mathsf{EA} \vdash \ell = \bar{r} &\rightarrow \mathsf{Prov}_T(\ulcorner \textstyle\bigvee_{u \succeq r} \ell = \bar{u} \urcorner) \\ &\rightarrow f_T(\Box\theta).\end{aligned}$$

(ii) If $z \nVdash \Box\theta$, then there is a node $u \succ z$ such that $u \nVdash \theta$. Pick a maximal such node. We claim that $u \neq r$. Indeed, by maximality,

$$u \nVdash \theta \text{ and } \forall w \succ u \; w \Vdash \theta.$$

Hence $u \Vdash \Box\theta$, and thus u cannot be a φ-reflexive node, whereas the node r is φ-reflexive. By the induction hypothesis, we have

$$\mathsf{EA} \vdash \ell = \bar{u} \rightarrow \neg f_T(\theta),$$

whence

$$\begin{aligned}\mathsf{EA} \vdash \ell = \bar{z} &\rightarrow \neg\mathsf{Prov}(\ulcorner \ell \neq \bar{u} \urcorner) \\ &\rightarrow \neg\mathsf{Prov}_T(\ulcorner f_T(\theta) \urcorner) \\ &\rightarrow \neg f_T(\Box\theta),\end{aligned}$$

as in the proof of Lemma 35. ■

LEMMA 54. $\mathsf{EA} + \{\neg\Box^n\bot : n \in \omega\}^T$ *proves the following statements:*

(i) $\ell \in \{0, 1, r\}$*;*

(ii) $\mathsf{Prov}_T(\ulcorner\sigma\urcorner) \wedge \neg\sigma \rightarrow \ell = 1$.

Proof. (i) On the set of nodes $K \setminus \{0, 1, r\}$ the function h behaves exactly as the Solovay function. In particular, for any $z \notin \{0, 1, r\}$ we have

$$\mathsf{EA} \vdash \ell = \bar{z} \rightarrow (\Box^{d(z)+1}\bot)^T,$$

which can be proved by induction on the depth of z. It follows that for $n > h(\mathcal{K})$

$$\begin{aligned}\mathsf{EA} \vdash (\neg\Box^n\bot)^T &\rightarrow \bigwedge_{z \notin \{0,1,r\}} \ell \neq \bar{z} \\ &\rightarrow \ell \in \{0, 1, r\}.\end{aligned}$$

(ii) First of all, we notice that

$$\mathsf{EA} \vdash \sigma \rightarrow \ell \neq 1,$$

because if the alarm rings while the refugee resides in 1, he/she is forced to leave. This yields

$$\begin{array}{rcl}\mathsf{EA} \vdash \mathsf{Prov}_T(\ulcorner\sigma\urcorner) & \to & \mathsf{Prov}_T(\ulcorner\ell \neq 1\urcorner) \\ & \to & \ell \neq 0.\end{array}$$

It is also obvious that

$$\mathsf{EA} \vdash \neg\sigma \to \ell \neq r,$$

because the only way for a refugee to get to the node r is by jumping from 1 when the alarm rings. Summing this up with Statement (i) we obtain

$$\mathsf{EA} + \{\neg\Box^n\bot : n \in \omega\}^T \vdash \mathsf{Prov}_T(\ulcorner\sigma\urcorner) \wedge \neg\sigma \to \ell = 1,$$

as required. ■

We finish the proof of Lemma 51 as follows. By Lemma 53,

$$\mathsf{EA} \vdash \ell = 1 \to \neg f_T(\varphi),$$

and combining this with Lemma 54 (ii) we obtain

$$\mathsf{EA} + \{\neg\Box^n\bot : n \in \omega\}^T \vdash \mathsf{Prov}_T(\ulcorner\sigma\urcorner) \wedge \neg\sigma \to \neg f_T(\varphi).$$

So, by propositional logic,

$$\mathsf{EA} + \{\neg\Box^n\bot : n \in \omega\}^T + f_T(\varphi) \vdash \mathsf{Prov}_T(\ulcorner\sigma\urcorner) \to \sigma,$$

as required. ■

COROLLARY 55. *There are no provability logics strictly between* GL_ω *and* D*.*

Logics between D *and* S

The weight of the arithmetical component in this part of the proof is relatively low. Analysis of Kripke models for the logics D [Beklemishev, 1989b] and S [Visser, 1984] and their characteristic formulas yields the following property whose proof can be found in [Beklemishev *et al.*, 1999].

LEMMA 56. *Let φ be a modal formula such that $\mathsf{D} \nvdash \varphi$. Then there is a formula ψ such that* $\mathrm{Var}(\psi) = \mathrm{Var}(\varphi)$*, $\mathsf{S} \nvdash \psi$ and $\mathsf{D}\{\varphi\} \vdash \psi \vee (\Box p \to p)$, where $p \notin \mathrm{Var}(\varphi)$.*

From this lemma we infer

LEMMA 57. *Let T be an elementary presented theory and φ a modal formula such that $\mathsf{D} \nvdash \varphi$. Then $\mathsf{D}\{\varphi\} \vdash_T^* \Box p \to p$.*

Proof. For a given φ apply Lemma 56 and obtain a formula ψ. Consider the T-completion $L = [\mathsf{D}\{\psi\}]^T$ of the logic $\mathsf{D}\{\psi\}$. By Lemma 49, since $\mathsf{S} \nvdash \psi$, L coincides with GL^-_β for some cofinite $\beta \subseteq \omega$. Let F denote the formula $\bigvee_{n \notin \beta} \neg F_n$. Obviously, $L \vdash F$ and therefore $\mathsf{D}\{\psi\} \vdash^*_T F$. But we also have $\mathsf{D} \vdash \neg F$ because D contains GL_ω, hence $\mathsf{D}\{\psi\} \vdash^*_T \bot$. Now using the fact that $p \notin \mathrm{Var}(\psi)$ we conclude that

$$\mathsf{D}\{\psi \vee (\Box p \to p)\} \vdash^*_T \Box p \to p.$$

Thus, $\mathsf{D}\{\varphi\} \vdash^*_T \Box p \to p$, as required. ■

COROLLARY 58. *There are no provability logics strictly between* D *and* S.

6.5 Examples and discussion

So far we have excluded all non-provability logics, but we have not yet shown that the remaining ones — $\mathsf{GL}_\alpha, \mathsf{GL}^-_\beta, \mathsf{S}_\beta$, and D_β — are arithmetically complete. However, having done all the technical work in the previous section, this is now easy.

The fact that logics of the form GL_α and GL^-_β (α coinfinite, β cofinite) are T-complete for any T of infinite characteristic follows from Lemma 45. Indeed, any of these logics L is maximal among consistent provability logics, therefore $[L]^T = L$.

We need two more examples.

EXAMPLE 59. $\boldsymbol{PL}_T(T_\omega) = \mathsf{GL}_\omega$, if T has infinite characteristic.

Proof. The containment ($\supseteq$) is clear. If $\boldsymbol{PL}_T(T_\omega) \neq \mathsf{GL}_\omega$, then $\boldsymbol{PL}_T(T_\omega) \supseteq \mathsf{D}$ and hence

$$T_\omega \vdash \mathsf{D}^T \vdash \mathsf{Rfn}_{\Sigma_1}(T),$$

by Corollary 52. However, by Theorem 23, $\mathsf{Rfn}_{\Sigma_1}(T)$ is not contained in any consistent r.e. extension of T by Π_1-sentences, in particular, in T_ω if $ch(T) = \infty$. ■

EXAMPLE 60. $\boldsymbol{PL}_T(T + \mathsf{Rfn}_{\Sigma_1}(T)) = \mathsf{D}$, if T has infinite characteristic.

Proof. ($\supseteq$) is easy. Inequality would imply $\boldsymbol{PL}_T(T + \mathsf{Rfn}_{\Sigma_1}(T)) \supseteq \mathsf{S}$ and hence

$$T + \mathsf{Rfn}_{\Sigma_1}(T) \vdash \mathsf{Rfn}(T).$$

If $ch(T) = \infty$, then $T + \mathsf{Rfn}_{\Sigma_1}(T)$ is a consistent (by Corollary 30) r.e. extension of T of complexity Π_2. Hence, by Theorem 23, it cannot contain $\mathsf{Rfn}(T)$. ■

Since the class of T-complete logics is closed under intersection, we conclude that all logics $\mathsf{GL}_\alpha, \mathsf{GL}^-_\beta, \mathsf{S}_\beta$, and D_β are T-complete, for any T of infinite characteristic.

The Classification theorem shows that the provability logic $\boldsymbol{PL}_T(U)$ is essentially determined by the amount of reflection for T that is provable in U. To find out, given T and U, how much reflection for T is provable in U can be rather difficult. However, this question for many natural pairs of theories has already been investigated using traditional proof-theoretic methods (see Section 10 for more examples and applications of such results).

EXAMPLE 61. $\boldsymbol{PL}_{\mathsf{EA}}(\mathsf{PA}) = \boldsymbol{PL}_{I\Sigma_n}(\mathsf{PA}) = \mathsf{S}$.

Proof. By a well-known theorem of G. Kreisel and A. Lévy [Kreisel and Lévy, 1968], $\mathsf{PA} \vdash \mathsf{RFN}(I\Sigma_n)$. Classification theorem then leaves only one possibility. ∎

EXAMPLE 62. $\boldsymbol{PL}_{I\Sigma_m}(I\Sigma_n) = \mathsf{D}$, for $m < n$.

Proof. A theorem of D. Leivant [Leivant, 1983; Hájek and Pudlák, 1993] states that for all $n \geq 1$,

$$I\Sigma_{n+1} \vdash \mathsf{RFN}_{\Sigma_{n+2}}(I\Sigma_n).$$

On the other hand,

$$I\Sigma_{n+1} \nvdash \mathsf{Rfn}(I\Sigma_n),$$

because $I\Sigma_{n+1}$ is a finitely axiomatizable extension of $I\Sigma_n$. ∎

EXAMPLE 63. $\boldsymbol{PL}_{\mathsf{PA}}(\mathsf{PA} + \mathsf{Con}(\mathsf{ZF})) = \boldsymbol{PL}_{I\Sigma_1}(I\Sigma_1 + \mathsf{Con}(\mathsf{PA})) = \mathsf{GL}_\omega$.

Proof. Obviously, if an elementary presented theory U contains the local Σ_1-reflection schema for T, then $T + \mathsf{Con}(U) \supseteq T_\omega$. Yet, a consistent theory $T + \mathsf{Con}(U)$ cannot prove $\mathsf{Rfn}_{\Sigma_1}(T)$ because $\mathsf{Con}(U)$ is Π_1. So, $\boldsymbol{PL}_T(T + \mathsf{Con}(U)) \neq \mathsf{D}$ and has to coincide with GL_ω. ∎

7 PROVABILITY ALGEBRAS

Provability algebras were introduced by R. Magari [Magari, 1975a; Magari, 1975b] as an alternative way of looking at provability logic.[10] Some (though

[10] R. Magari used the term 'diagonalizable algebras'. Later the term 'Magari algebras' has been used on a par with the original one. This latter was officialised at a gathering of provability logicians at the Magari memorial conference in Siena in 1994. In this paper we have a need in two terms: the one for the Lindenbaum algebra associated with a formal theory T equipped with the operator of provability, which we call *the provability algebra of* T, and another for the more general algebras satisfying the same identities, which we call *Magari algebras*. Thus, provability algebras are a particular kind of Magari algebras naturally associated with arithmetical theories.

not all) of the results in provability logic can be translated from the logical language to the algebraic language and vice versa. Very often the choice of language is more or less a matter of taste. However, there are some advantages to the algebraic point of view: firstly, it is closer to the way of looking at things in mathematics, it emphasizes the underlying *structures* and thus helps to formulate proper analogies and questions to be answered. Secondly, this approach is very flexible. It allows to naturally incorporate certain additional features of arithmetical theories that are, in particular, necessary for further applications in proof theory.

This section is mostly written for logically and proof-theoretically, rather than algebraically, minded readers. So, we do not presume much knowledge of universal algebra and try to be somewhat economical with the use of algebraic terminology.

7.1 Lindenbaum algebras

Let an elementary presented theory T containing EA be given. The *Lindenbaum boolean algebra* of T, denoted $\mathcal{B}_T$, has as its universe the set of all T-sentences modulo the equivalence relation

$$\varphi \sim_T \psi \iff T \vdash \varphi \leftrightarrow \psi.$$

Officially we denote by $[\varphi]_T$ the equivalence class $\{\psi : \psi \sim_T \varphi\}$ of φ, but in practice we shall often identify the equivalence classes and formulas. The implication $\to$ induces an operation on the set of equivalence classes: $[\varphi]_T \to [\psi]_T = [\varphi \to \psi]_T$. Together with obviously defined constants $\bot$ and $\top$ it gives the set $\mathcal{B}_T$ a structure of a boolean algebra $(\mathcal{B}_T, \to, \bot, \top)$. As before, we regard $\wedge$, $\vee$, $\neg$, $\leftrightarrow$ as defined operations. The relation

$$[\varphi]_T \leq [\psi]_T \iff [\varphi]_T \to [\psi]_T = \top \iff T \vdash \varphi \to \psi$$

is the standard partial ordering on $\mathcal{B}_T$.

The structure $\mathcal{B}_T$ provides an algebraic view of some proof-theoretic objects: *schemata* over T correspond to subsets of $\mathcal{B}_T$; extra-logical *inference rules* correspond to *operators* acting on $\mathcal{B}_T$; deductively closed sets of formulas, usually called *extensions of* T, correspond to *filters* of $\mathcal{B}_T$, that is, subsets of $\mathcal{B}_T$ upwards closed w.r.t. $\leq$ and closed under $\wedge$. If U is an extension of T, then $\mathcal{B}_U$ can be identified with the corresponding quotient algebra of $\mathcal{B}_T$.[11]

The notion of Lindenbaum algebra makes sense for any formal system containing propositional logic. If T is a consistent elementary presented extension of EA, the boolean algebra $\mathcal{B}_T$ turns out to be uniquely defined and well understood.

[11] See Section 7.4 below for a definition of a quotient algebra.

PROPOSITION 64. *If T is consistent and contains* EA, *$\mathcal{B}_T$ is a countable dense boolean algebra.*

Proof. Countability of $\mathcal{B}_T$ is clear, since we assume that the language of T is countable. Density means that if $\varphi < \psi$ in $\mathcal{B}_T$, then there is a θ such that $\varphi < \theta < \psi$.

Consider a theory $T_1 = T + \psi + \neg\varphi$. Since $T \nvdash \psi \to \varphi$, T_1 is consistent. So, by Rosser's theorem, there is a sentence ρ such that both $T_1 + \rho$ and $T_1 + \neg\rho$ are consistent. Take $\theta = (\rho \wedge \psi) \vee \varphi$. Obviously, $\varphi \leq \theta \leq \psi$. On the other hand, if $T \vdash \theta \to \varphi$, then $T_1 \vdash \neg\rho$. If $T \vdash \psi \to \theta$, then $T_1 \vdash \rho$. Both statements contradict the choice of ρ. ■

We mention without proof the following simple fact from basic boolean algebra theory (see also [Goncharov, 1997] for an in-depth monograph on countable boolean algebras).

PROPOSITION 65. *Any two countable dense boolean algebras are isomorphic.*

So, the Lindenbaum algebras of all interesting theories, such as PA, EA, ZF, etc., are isomorphic. Moreover, by [Pour-El and Kripke, 1967] they are even recursively isomorphic, considered as numerated structures. This indicates that the structure of the Lindenbaum boolean algebra is too poor to capture essential proof-theoretic information on any particular system. We want to enrich this structure.

7.2 Provability algebras

The provability predicate $\mathsf{Prov}_T(x)$, by Löb's derivability conditions, correctly defines an operator

$$\Box_T : [\varphi]_T \longmapsto [\mathsf{Prov}_T(\ulcorner\varphi\urcorner)]_T$$

acting on the Lindenbaum algebra $\mathcal{B}_T$. Indeed, if $T \vdash \varphi \leftrightarrow \psi$, then $T \vdash \mathsf{Prov}_T(\ulcorner\varphi\urcorner) \leftrightarrow \mathsf{Prov}_T(\ulcorner\psi\urcorner)$. The enriched structure $\mathcal{M}_T = (\mathcal{B}_T, \Box_T)$ is called the *provability algebra of T*.

One of the first questions one usually asks about a newly defined algebra $\mathcal{A}$ is: what identities does it satisfy? Recall that an *identity* of $\mathcal{A}$ is a valid in $\mathcal{A}$ formula of the form $\forall\vec{x}\,(t_1(\vec{x}) = t_2(\vec{x}))$, where t_1, t_2 are terms in the language of $\mathcal{A}$.

Terms in the language of provability algebras are built up from variables and $\top, \bot$ by the operations $\to, \Box$ and can be identified with propositional modal formulas. From now on we shall denote provability algebra terms and propositional formulas by the same letters φ, ψ, etc. Any equation in a boolean algebra can be written in a simplified form, where the second term is just a constant:

$$\mathcal{A} \vDash \varphi_1 = \varphi_2 \iff \mathcal{A} \vDash (\varphi_1 \leftrightarrow \varphi_2) = \top.$$

The following proposition shows that such simplified identities of $\mathcal{M}_T$ are described by the provability logic of T.

PROPOSITION 66. *For any modal formula/term $\varphi(\vec{p})$,*

$$\mathcal{M}_T \vDash \forall \vec{p}\,(\varphi(\vec{p}) = \top) \iff \varphi \in \boldsymbol{PL}_T(T).$$

Proof. Obviously, $\mathcal{M}_T \vDash \forall\vec{p}(\varphi(\vec{p}) = \top)$ iff $f_T(\varphi)$ is provable in T, for every arithmetical realization f. ■

Solovay's first completeness theorem now translates to the following

COROLLARY 67. *If $ch(T) = \infty$, then for any modal formula/term $\varphi(\vec{p})$,*

$$\mathsf{GL} \vdash \varphi(\vec{p}) \iff \mathcal{M}_T \vDash \forall \vec{p}\,(\varphi(\vec{p}) = \top).$$

7.3 Magari algebras and duality

R. Magari introduced a general notion of algebra satisfying all the identities of provability algebras. A *Magari algebra* $\mathcal{M}$ is a boolean algebra equipped with an additional operator $\Box$ satisfying the identities:

(i) $\Box(\varphi \to \psi) \to (\Box\varphi \to \Box\psi) = \top$;

(ii) $\Box(\Box\varphi \to \varphi) \to \Box\varphi = \top$;

(iii) $\Box\top = \top$.

It is easy to verify that any theorem $\varphi(\vec{p})$ of GL represents an identity $\varphi(\vec{p}) = \top$ that holds in all Magari algebras. Identity (iii) ensures the closure under the necessitation rule. Vice versa, if $\varphi(\vec{p}) = \top$ holds in all Magari algebras, in particular, it must hold in $\mathcal{M}_{\mathsf{PA}}$, therefore $\mathsf{GL} \vdash \varphi$.

Natural Magari algebras can be constructed from Kripke frames for GL. Consider a frame $\mathcal{K} = (K, \prec)$ and let $\mathcal{M}$ be the algebra of all subsets of K with the standard boolean operations and the following operation $\Box$: for any $X \subseteq K$,

$$\Box X := \{x \in K : \forall y \in K\ (x \prec y \Rightarrow y \in X)\}.$$

Then it is easy to verify that $\mathcal{M}$ is a Magari algebra satisfying those identities $\varphi(\vec{p}) = \top$ such that $\mathcal{K} \vDash \varphi(\vec{p})$.

This method is especially useful for constructing finite Magari algebras. For example, from the finite model property for GL (Corollary 4) we obtain the corresponding finite model property for Magari algebras.

PROPOSITION 68. *If an identity holds in every finite Magari algebra, then it holds in all Magari algebras.*

A natural generalization of the above construction and its inverse lead to the Stone duality theory for Magari algebras. It is not our intention here to go into this topic. See [Magari, 1975b] and [Bull and Segerberg, 2001; van Benthem, 2001] for the details.

7.4 Subalgebras, filters, free algebras

Here we describe the method of constructing Magari algebras as quotient algebras of free algebras. First, we recall some standard terminology applicable to Magari algebras.

A *homomorphism* between Magari algebras $\mathcal{A}$ and $\mathcal{B}$ is a mapping $f : \mathcal{A} \to \mathcal{B}$ preserving all the operations, that is,

- $f(\top) = \top$, $f(\bot) = \bot$;
- $f(\varphi \to \psi) = (f(\varphi) \to f(\psi))$;
- $f(\Box\varphi) = \Box(f(\varphi))$.

An *embedding* is a one-to-one homomorphism; an *epimorphism* is an onto homomorphism. An *isomorphism* is both an embedding and an epimorphism.

Let a subset X of a Magari algebra $\mathcal{A}$ be given. X generates a *subalgebra* $\langle X \rangle$ of $\mathcal{A}$, that is, the smallest subset of $\mathcal{A}$ containing $X \cup \{\top, \bot\}$ and closed under all functions of $\mathcal{A}$. It can also be described as the set of values of all terms in the language of $\mathcal{A}$ on arguments coming from X. We say that X *generates* $\mathcal{A}$ if $\langle X \rangle = \mathcal{A}$. Among various subalgebras of $\mathcal{A}$ there always is the minimal one, $\langle \varnothing \rangle$, which is called the *prime* subalgebra of $\mathcal{A}$.

A filter P of (the boolean part of) a Magari algebra $\mathcal{A}$ is called a $\Box$-*filter*, if $x \in P$ implies $\Box x \in P$, for all $x \in \mathcal{A}$. If P is a $\Box$-filter, then the corresponding *quotient algebra* $\mathcal{A}/P$ is defined as the set of equivalence classes of $\mathcal{A}$ modulo the relation $x \sim_P y \iff (x \leftrightarrow y) \in P$, with the inherited operations $\top$, $\bot$, $\to$ and $\Box$. Clearly, $\mathcal{A}/P$ will also be a Magari algebra. The mapping $\pi_P : x \mapsto [x]_P$ is called the *canonical epimorphism* from $\mathcal{A}$ to $\mathcal{A}/P$.

Every Magari algebra generated by X is isomorphic to a suitable quotient algebra of a *free* algebra on X. The latter is defined as follows.

Let X be a set of propositional variables, and let $\mathcal{L}(X)$ be the language of GL with the variables from X. Provable equivalence in GL induces an equivalence relation on the set of $\mathcal{L}(X)$-formulas, and the resulting Lindenbaum algebra obviously bears the structure of a Magari algebra:

$$\Box([\varphi]_{\mathsf{GL}}) := [\Box\varphi]_{\mathsf{GL}}.$$

We call this algebra *free* and denote it by $\mathbf{Fr}(X)$. $\mathbf{Fr}(n)$ and $\mathbf{Fr}(\omega)$ denote, respectively, free Magari algebras $\mathbf{Fr}(X)$ for $X = \{p_0, \dots, p_{n-1}\}$ and $X = \{p_i : i \in \omega\}$. $\mathbf{Fr}(0)$ is the Lindenbaum algebra of the letterless fragment of GL.

Notice that $\Box$-filters on $\mathbf{Fr}(X)$ can be identified with *propositional modal theories* containing GL, that is, with sets of modal formulas containing all theorems of GL and closed under the modus ponens and necessitation rules. The rule of substitution is generally not admissible (otherwise a logician would speak about a *propositional logic*, not a theory). In other words, in this situation the elements of X are treated as propositional *constants*, not as variables.

Thus, the quotient algebra of $\mathbf{Fr}(X)$ w.r.t. a $\Box$-filter P is just the Lindenbaum Magari algebra of the propositional theory axiomatized by formulas from P over GL. If $\mathcal{A}$ is generated by X, then there is a natural epimorphism $\pi : \mathbf{Fr}(X) \to \mathcal{A}$ which maps every formula from $\mathbf{Fr}(X)$ to the value of the corresponding term in $\mathcal{A}$. Then $\mathcal{A}$ will be isomorphic to the quotient algebra $\mathbf{Fr}(X)/P_\pi$, where $P_\pi = \{\varphi : \mathcal{A} \vDash \pi(\varphi) = \top\}$. We formulate this simple but important fact as a separate proposition.

PROPOSITION 69. *Every Magari algebra $\mathcal{A}$ generated by X is isomorphic to the quotient algebra $\mathbf{Fr}(X)/P$ for a suitable $\Box$-filter P. Equivalently, it is isomorphic to the Lindenbaum Magari algebra of a propositional modal theory P in $\mathcal{L}(X)$.*

In this sense, speaking about Magari algebras is equivalent to speaking about propositional theories. It is important to realize, however, that provability algebras coming from arithmetic lack *natural* systems of generators. The only such system we can think of is just the set of all arithmetical sentences. Thus, a transparent description of provability algebras as the quotient algebras of free algebras is missing.

7.5 Subalgebras of free algebras

It is well known that any subgroup of a free group is free. Does the same fact hold for Magari algebras? The answer is a definite NO. How can we characterize such structures?

Here we consider finitely generated subalgebras of free Magari algebras.

Let $\mathcal{A}$ be a subalgebra of $\mathbf{Fr}(\omega)$ generated by elements $B_1(\vec{q}), \dots, B_n(\vec{q})$, where we assume that $q_1,\ q_2, \dots$ are free generators of $\mathbf{Fr}(\omega)$. We can look at these formulas as defining a substitution $\sigma : Fm(\vec{p}) \to Fm(\vec{q})$ so that $\sigma(p_i) = B_i$, for $i = 1, \dots, n$. Vice versa, any substitution of this kind defines a subalgebra of $\mathbf{Fr}(\omega)$.

The following theorem comes from [de Jongh and Visser, 1996] where the corresponding fact was stated for subalgebras of free Heyting algebras. The argument is based on the uniform interpolation theorem for GL [Shavrukov, 1993b].

THEOREM 70. *For every σ we can effectively find a formula $T_\sigma \in Fm(\vec{p})$ such that $\mathcal{A}$ is isomorphic to $\mathbf{Fr}(\vec{p})/T_\sigma$, that is, the quotient algebra of the free algebra by the principal filter generated by T_σ.*

Proof. Let $C_\sigma(\vec{p}, \vec{q})$ denote the formula $\bigwedge_{i=1}^n \Box(p_i \leftrightarrow B_i(\vec{q}))$. By the substitution theorem, we have

$$\mathsf{GL} \vdash C_\sigma(\vec{p}, \vec{q}) \to (\varphi(B_1, \ldots, B_n) \leftrightarrow \varphi(p_1, \ldots, p_n)).$$

Therefore, we obtain

$$\mathsf{GL} \vdash \varphi(B_1, \ldots, B_n) \iff \mathsf{GL} \vdash C_\sigma(\vec{p}, \vec{q}) \to \varphi(\vec{p}).$$

By the uniform interpolation theorem, from C_σ we can effectively construct a formula $T_\sigma(\vec{p})$ such that $\mathsf{GL} \vdash C_\sigma(\vec{p}, \vec{q}) \to T_\sigma(\vec{p})$ and for any $\varphi \in Fm(\vec{p})$,

$$\mathsf{GL} \vdash C_\sigma(\vec{p}, \vec{q}) \to \varphi(\vec{p}) \Rightarrow \mathsf{GL} \vdash T_\sigma \to \varphi(\vec{p}).$$

Hence,

$$\mathsf{GL} \vdash \varphi(B_1, \ldots, B_n) \iff \mathsf{GL} \vdash T_\sigma \to \varphi(\vec{p}),$$

as required. ■

COROLLARY 71. *The propositional theory of any finitely generated subalgebra of a free Magari algebra is finitely axiomatizable.*

Notice that for each σ the formula T_σ is defined uniquely modulo provable equivalence in GL. Formulas of the form T_σ are called *exact* in [de Jongh and Visser, 1996]. This notion turns out to be equivalent to the notion of *projectivity* introduced by [Ghilardi, 2000]. The following characterization can be inferred from Ghilardi's results (A. Visser, unpublished).

THEOREM 72. *The following statements are equivalent, for any formula φ:*

(i) *φ is exact;*

(ii) *φ has the* extension property*, that is, for every (non-rooted) model $\mathcal{K}$ such that $\mathcal{K} \vDash \varphi$, there is a model of φ obtained by just attaching a new root to $\mathcal{K}$;*

(iii) *φ satisfies the extension property for finite models.*

S. Ghilardi also established that the exactness/projectivity property is decidable.

Our next goal is the study of subalgebras of provability algebras. For this we need yet another important general notion.

7.6 Numerated and positive algebras

Provability algebras, consisting of (equivalence classes of) arithmetical formulas, bear a natural Gödel numbering that allows to speak about their computational properties. This numbering is not one-to-one because every element of $\mathcal{M}_T$ corresponds to an infinite r.e. set of sentences. Yet, we can call a subalgebra of $\mathcal{M}_T$ r.e. if so is the set of all Gödel numbers of its elements.

The notion of *numerated algebra* expresses on a more abstract level the idea of an algebra endowed with a Gödel numbering. Numerated Magari algebras can be defined as follows.

Assume $\mathcal{A} = \langle X \rangle$ and $|X| \leq \omega$. Then $\mathbf{Fr}(X)$ is isomorphic to $\mathbf{Fr}(\alpha)$ for some $\alpha \leq \omega$. An epimorphism $\pi : \mathbf{Fr}(\alpha) \to \mathcal{A}$ is called a *numeration of* $\mathcal{A}$. We look at the elements of $\mathbf{Fr}(\alpha)$ as the codes of the elements of $\mathcal{A}$. Magari algebras equipped with a numeration are called *numerated.* A numerated algebra $\mathcal{A}$ is *positive*, if the associated theory $P_\pi = \{\varphi : \pi(\varphi) = \top\}$ is r.e..

Notice that any provability algebra $\mathcal{M}_T$ is a numerated Magari algebra in the sense of the above definition: one considers all sentences as generators of $\mathcal{M}_T$ and maps p_n to the arithmetical sentence with the Gödel number n. Since we assume T to be r.e., $\mathcal{M}_T$ is a positive algebra. Any finitely generated subalgebra of a positive algebra is also positive. Any r.e. set of sentences generates a positive subalgebra of $\mathcal{M}_T$.

7.7 Subalgebras of provability algebras

Shortly after the Solovay theorems were published several authors independently found an improvement, which has become known as the *uniform Solovay theorem* [Montagna, 1979; Artemov, 1979; Visser, 1980; Boolos, 1982; Avron, 1984].

THEOREM 73. *Suppose T has infinite characteristic. Then there is an arithmetical realization f such that, for any modal formula φ,*

$$\mathsf{GL} \vdash \varphi \iff T \vdash f_T(\varphi).$$

Proof. The usual proof of this theorem applies the Solovay construction to the Kripke model obtained by a disjoint union of all finite treelike models and adding a new root 0 below them all. This model is infinite but converse well-founded and can be elementarily represented in arithmetic. Any modal formula not provable in GL is refuted at some node of this model. Therefore, the associated Solovay function provides the required uniform realization f.

There is only one detail to be taken care of: T must prove that the Solovay function has a limit. Since our model is infinite, this only works for T containing $I\Sigma_1$. D. Zambella [Zambella, 1994] found (even in a more general situation) a modification of the Solovay construction which shows that the theorem also holds for any T containing EA. ■

Theorem 73 appears to be very natural from the algebraic point of view. It is essentially equivalent to the following fact.

COROLLARY 74. *If T has infinite characteristic, then $\mathbf{Fr}(\omega)$ is embeddable into $\mathcal{M}_T$, the provability algebra of T.*

This leads one to a more general question, what kind of Magari algebras are embeddable into $\mathcal{M}_T$. In view of Proposition 69 this problem is equivalent to the one about propositional theories P *realizable in* T. We say that P is realizable in T, if for some arithmetical realization f,

$$\varphi \in P \iff T \vdash f_T(\varphi).$$

An almost complete characterization of subalgebras of provability algebras was obtained by V. Shavrukov [Shavrukov, 1993b]. D. Zambella (unpublished) later filled in the last remaining gap. He also extended Shavrukov's results to arbitrary extensions of EA [Zambella, 1994]. The most important part of the above problem concerns positive Magari algebras or, equivalently, r.e. propositional theories P.

A Magari algebra $\mathcal{A}$ has *characteristic* n, where $0 \leq n < \infty$, if n is the minimal natural number such that $\mathcal{A} \vDash \Box^n \bot = \top$. If such an n does not exist, we say that the characteristic of $\mathcal{A}$ is infinite. Notice that the characteristic of a provability algebra $\mathcal{M}_T$ equals $ch(T)$. Also, all subalgebras of any algebra have the same characteristic.

$\mathcal{A}$ has the *strong disjunction property* (s.d.p.), if $\top \neq \bot$ and

$$\mathcal{A} \vDash \Box\varphi \vee \Box\psi = \top \quad \Rightarrow \quad \mathcal{A} \vDash \varphi = \top \text{ or } \mathcal{A} \vDash \psi = \top.$$

If T is Σ_1-sound, $\mathcal{M}_T$ and all of its subalgebras obviously satisfy s.d.p.

The same terminology applies to propositional theories, that is, we say that a theory P has one of the above properties if the corresponding algebra $\mathbf{Fr}(X)/P$ does.

V. Shavrukov [Shavrukov, 1993b] proved the following two theorems, which together characterize r.e. subalgebras of any provability algebra.

THEOREM 75. *Suppose T is Σ_1-sound. $\mathcal{A}$ is isomorphic to an r.e. subalgebra of $\mathcal{M}_T$ iff $\mathcal{A}$ is positive and satisfies s.d.p.*

THEOREM 76. *Suppose T is not Σ_1-sound. $\mathcal{A}$ is isomorphic to an r.e. subalgebra of $\mathcal{M}_T$ iff $\mathcal{A}$ is positive and the characteristic of $\mathcal{A}$ equals $ch(T)$.*

We sketch some ideas of the proofs of these theorems in the following subsections. So far, only the "only if" parts of both theorems are clear. First, we develop some understanding of the strong disjunction property.

7.8 Strong disjunction property

Let P be a set of modal formulas. We write $P \vdash \varphi$ if a formula φ is provable from P and axioms of GL using *modus ponens* and necessitation rules. This is another form of saying that φ belongs to the $\Box$-filter generated by P in $\mathbf{Fr}(\omega)$. We say that $\mathcal{K}$ is a model of P if $\mathcal{K} \vDash \varphi$ for all $\varphi \in P$.

Let C be a finite set of formulas. A propositional theory P satisfies *s.d.p. for C* if $P \nvdash \bot$ and

$$P \vdash \Box\varphi \vee \Box\psi \;\Rightarrow\; (P \vdash \varphi \text{ or } P \vdash \psi),$$

for all formulas $\Box\varphi, \Box\psi \in C$.

THEOREM 77. *Let P be a finite propositional theory. The following statements are equivalent:*

(i) P has s.d.p.;

(ii) P has s.d.p. for the set of subformulas of P;

(iii) any two finite Kripke models of P are embeddable into a model of P as proper submodels.

Proof. The implications (i)⇒(ii) and (iii)⇒(i) are easy. We prove (ii)⇒(iii).

Let a pair of models $\mathcal{K}_1$, $\mathcal{K}_2$ be given, and let C be the set of all subformulas of formulas in P. Consider the set

$$Q := \{\neg\Box\psi : \mathcal{K}_1 \nvDash \psi \text{ or } \mathcal{K}_2 \nvDash \psi; \Box\psi \in C\}.$$

By s.d.p. of P, $P \nvdash \bigwedge Q \to \bot$. Hence, by Theorem 2, there is a finite Kripke model $\mathcal{W}$ such that $\mathcal{W} \vDash P$ and $\mathcal{W} \Vdash \bigwedge Q$.

Let $\mathcal{W}'$ be obtained by adding to $\mathcal{W}$ the models $\mathcal{K}_1$ and $\mathcal{K}_2$ immediately above the root. By induction on φ one easily verifies that

$$\mathcal{W} \Vdash \varphi \iff \mathcal{W}' \Vdash \varphi,$$

for any formula $\varphi \in C$.

Consider the case $\varphi = \Box\psi$. If $\mathcal{W} \nVdash \Box\psi$, then clearly $\mathcal{W}' \nVdash \Box\psi$. If $\mathcal{W} \Vdash \Box\psi$, then $\neg\Box\psi \notin Q$ because $\mathcal{W} \Vdash \bigwedge Q$. Hence, by definition of Q, we have $\mathcal{K}_1, \mathcal{K}_2 \vDash \psi$. This implies $\mathcal{W}' \Vdash \Box\psi$.

Thus, $\mathcal{W}' \Vdash P$ and $\mathcal{W}' \vDash P$. ■

COROLLARY 78. *The s.d.p. of a finitely axiomatized propositional theory is decidable.*

For infinite theories P we have the following characterization.

THEOREM 79. *P satisfies s.d.p. iff for every φ such that $P \vdash \varphi$ there is a finite subtheory A of P such that A has s.d.p. and $A \vdash \varphi$.*

Proof. ($\Leftarrow$) Assume $P \vdash \Box\varphi \vee \Box\psi$. Then for a suitable A we have $A \vdash \Box\varphi \vee \Box\psi$. Hence, $A \vdash \varphi$ or $A \vdash \psi$ and the same holds for P, since $P \vdash A$.

($\Rightarrow$) Assume P enjoys s.d.p. and $P \vdash \varphi$. Let C be the set of all subformulas of φ. Consider the set $P_C := \{\psi \in C : P \vdash \psi\}$. We claim that P_C has s.d.p. for C.

Indeed, if $\Box\psi_1$ and $\Box\psi_2$ are in C and $P_C \vdash \Box\psi_1 \vee \Box\psi_2$, then $P \vdash \Box\psi_1 \vee \Box\psi_2$, hence $P \vdash \psi_1$ or $P \vdash \psi_2$. Since $\psi_1, \psi_2 \in C$, we obtain $\psi_1 \in P_C$ or $\psi_2 \in P_C$, respectively.

From the previous theorem we conclude that P_C has s.d.p. and satisfies all the requirements. ■

As an illustration we give an interesting example from [Shavrukov, 1993b].

Consider a set of propositional letters $X = \{p_\alpha : \alpha \in \mathbb{Q}\}$, where $\mathbb{Q}$ is the linearly ordered set of rational numbers. Let a propositional theory Q be given by the axioms

$$\{\Diamond p_\alpha \to \Diamond\Diamond p_\beta : \alpha, \beta \in \mathbb{Q},\ \alpha > \beta\}.$$

One can show by a simple Kripke model argument that Q has s.d.p., in fact, every fragment of Q in finitely many variables does. It follows that $\mathbf{Fr}(X)/Q$ is embeddable into $\mathcal{M}_T$ for any Σ_1-sound T. This means that there is a family of Π_1-sentences — the interpretations of the formulas $\Diamond p_\alpha$ — ordered as $\mathbb{Q}$ by the relation 'φ proves the consistency of ψ over T', a nontrivial fact earlier proved in [Simmons, 1988] by purely arithmetical means.

This example shows that Shavrukov's theorems are essentially about simultaneous arithmetical realization of infinite families of modal formulas. In a sense, these results provide a generalization of Solovay's theorems from finite to infinite r.e. sets of formulas.

7.9 Proofs of Shavrukov's theorems

Here we sketch main ideas of the proofs of Theorems 75 and 76. These proofs have been greatly simplified by D. Zambella and we follow his presentation very closely [Zambella, 1994]. To simplify things yet further and concentrate on the essentials, we shall assume throughout this section that $ch(T) = \infty$ and T contains $I\Sigma_1$.

We fix some natural Gödel numbering of modal formulas. To simplify the notation, we shall essentially identify modal formulas and their codes and incorporate the variables φ, ψ ranging over modal formulas into the language of arithmetic. We shall also adopt the convention that these variables, unless explicitly said otherwise or bound by quantifiers, denote the standard formulas, that is, the numerals of their codes.

Let P be an r.e. set of formulas. We call an *elementary presentation of* P a Δ_0-formula "$\varphi \in P_n$", with the free variables n and φ, satisfying the following conditions:

- $\varphi \in P \iff \mathbb{N} \vDash \exists n$ "$\varphi \in P_n$";
- $T \vdash \forall \varphi, n\,($"$\varphi \in P_n$" $\to \varphi < n)$.

Thus, P_n denotes a finite part of P such that $P = \bigcup_{n \geq 0} P_n$.

Formalizing in T the notion of derivability in GL we construct from an elementary presentation of P a Δ_0-formula "$P_n \vdash \varphi$" naturally expressing the statement $P_n \vdash \varphi$. "$P \vdash \varphi$" then stands for $\exists x$ "$P_x \vdash \varphi$" within T.

Once an elementary presentation of P is fixed, we say that P satisfies s.d.p. *provably in* T if T proves

$$\neg\text{“}P \vdash \bot\text{”} \wedge \forall \varphi, \psi\,(\text{“}P \vdash \Box\varphi \vee \Box\psi\text{”} \;\to\; \text{“}P \vdash \varphi\text{”} \vee \text{“}P \vdash \psi\text{”}).$$

The central part of the proof of both theorems is the following lemma.

LEMMA 80. *Assume that $ch(T) = \infty$. If P is elementary presented and satisfies s.d.p. provably in T, then P is realizable in T.*

From this lemma one obtains Theorems 75, 76 (for the case $ch(T) = \infty$) by modifying any given elementary presentation of an r.e. set P in such a way that it becomes provably strongly disjunctive.

LEMMA 81. *Assume that $ch(T) = \infty$. Assume further that either P has s.d.p., or T is not Σ_1-sound. Then there is an elementary presentation of P for which P satisfies s.d.p. provably in T.*

Notice that, if T is Σ_1-unsound, T can be made to *think* that P has s.d.p., whereas in reality P need not satisfy this property. For the case of P satisfying s.d.p. one can easily define such an elementary presentation using an effective (elementary) version of Theorem 79: just enumerate P in such an order that the finite subtheories P_n at all stages have s.d.p. We shall omit a formal proof of Lemma 81, which is mainly technical, and concentrate our attention on Lemma 80.

Proof. Assume a propositional theory P provably satisfying s.d.p. is given. We shall define a Solovay-like function whose value $h(n)$ is either 0 or the code of a finite treelike model of P_m, for some $m \leq n$. We identify now the models and their codes and assume that submodels of a model have smaller codes than the model itself. We write $k_1 \prec k_2$ if k_2 is a proper submodel of the model k_1.

Fix a natural Δ_0-formula "$k \Vdash \varphi$", with the free variables k and φ, expressing the validity of φ in a model (coded by) k. Assume that 0 is not a code of a model and that "$0 \Vdash \varphi$" never holds. For each φ define an arithmetical sentence S_φ as follows:

$$S_\varphi = \exists k, m \forall n \geq m\ (h(n) = k \wedge \text{“}k \Vdash \varphi\text{”}).$$

Thus, the sentence S_φ asserts that φ holds at the limit of h, essentially as for the standard Solovay realization of φ. We also let $S_0 = (\forall n\ h(n) = 0)$. The elementary function h will be defined self-referentially by formalizing the following definition.

Let $h(0) = 0$. If n codes a proof of $S_0 \vee S_\varphi$ for some formula φ, then $h(n+1)$ is defined by the clauses:

(a) If $h(n) = 0$ and $P_n \nvDash \varphi$, then choose the minimal model k such that $k \vDash P_n$ and $k \nVdash \varphi$ and put $h(n+1) = k$.

(b) If $h(n) = k \neq 0$ and the root of some submodel of k forces $\neg\varphi$, then let k_1 be the minimal such submodel and put $h(n+1) = k_1$.

(c) In all other cases let $h(n+1) = h(n)$.

One way of looking at the above function is to have in mind the Kripke model consisting of all finite treelike models ordered by $\prec$ (and an attached root 0). Notice that h is now defined in terms of the sentences S_φ for arbitrary formulas φ, not just for those corresponding to isolated nodes of the model. Also, unlike in the original Solovay construction, h always tries to jump to the highest possible node of the model falsifying a certain modal formula.

With this in mind one observes that the behavior of h is similar to that of the usual Solovay function above the root.

LEMMA 82. *The following statements are provable in T:*

(i) $\exists z, m \forall n > m\, h(n) = z;$

(ii) $\forall n, k\ (h(n) = k \neq 0 \to \mathsf{Prov}_T(\ulcorner \exists m\ \dot{k} \prec h(m) \urcorner))$.

Proof. For (i) reason in $I\Sigma_1$ as follows: either h stays in 0 for all n, or it jumps. Using the Σ_1-least element principle pick the smallest model in the range of h. By the upwards monotonicity, this model can only be its limit.

For (ii) notice that h jumps to a node k above 0 only in the case there is a T-proof of $S_0 \vee S_\varphi$, for some formula φ that is false at k. By Σ_1-completeness, $h(n) \neq 0$ implies $\neg S_0$ and $\mathsf{Prov}_T(\ulcorner \neg S_0 \urcorner)$. Hence, $\mathsf{Prov}_T(\ulcorner S_\varphi \urcorner)$. But S_φ means that the limit of h forces φ, so it cannot be the same node as k and is therefore a proper submodel of k. ■

COROLLARY 83. *S_0 holds in the standard model.*

Proof. We note that for any $k \neq 0$ there is an m (the height of k) such that

$$T \vdash \exists n\, h(n) = \bar{k} \to (\Box^{m+1}\bot)^T.$$

This is easy to see from Lemma 82(ii) by a subsidiary induction on m. So, if $h(n) = k \neq 0$, then $T \vdash h(\bar{n}) = \bar{k}$ and $T \vdash (\Box^{m+1}\bot)^T$ contradicting the assumption that $ch(T) = \infty$. ■

Although from the point of view of the standard model h never leaves 0, T needs not see it. T only knows that the longer h stays at 0, the less possible worlds of the model remain where it can jump to because such nodes must validate ever larger fragments of P. So, in a sense, our model shrinks with time as long as $h(n) = 0$.

LEMMA 84. $T \vdash \forall n\, (h(n) = 0 \wedge \text{“}P_n \vdash \varphi\text{”} \rightarrow S_\varphi)$.

Proof. By (a) of the definition of h, if $h(n) = 0$ and $P_n \vdash \varphi$, then h jumps at stage $n+1$ and $h(m)$ will be a model of P_n for all $m \geq n+1$. Therefore, for all such m, $h(m) \Vdash \varphi$. ■

To define the required realization f we first have to specify the validity of modal formulas at the root 0 of the model in a suitable way. This will be achieved by constructing an arithmetical formula $\Phi(x)$ satisfying the requirements of the following lemma.

LEMMA 85. *There is an arithmetical formula Φ such that for any (standard) modal formulas φ, ψ the following conditions hold provably in T:*

(i) $\neg\Phi(\bot)$, $\Phi(\top)$;

(ii) $\Phi(\varphi \rightarrow \psi) \leftrightarrow (\Phi(\varphi) \rightarrow \Phi(\psi))$;

(iii) “$P \vdash \varphi$” $\rightarrow \Phi(\varphi)$;

(iv) $\Phi(\Box\varphi) \rightarrow$ “$P \vdash \varphi$”.

For a moment we postpone the proof of this lemma. Having such a formula Φ, we define the required arithmetical realization f as follows:

$$f(\varphi) := S_\varphi \vee (S_0 \wedge \Phi(\varphi)).$$

Here φ is any modal formula, in particular, a variable. To show that f behaves like an arithmetical interpretation we prove

LEMMA 86. *For all formulas φ, ψ the theory T proves:*

(i) $f(\bot) \leftrightarrow \bot$, $f(\top) \leftrightarrow \top$;

(ii) $f(\varphi \rightarrow \psi) \leftrightarrow (f(\varphi) \rightarrow f(\psi))$;

(iii) $f(\Box\varphi) \leftrightarrow \mathsf{Prov}_T(\ulcorner f(\varphi) \urcorner)$.

Proof. The cases (i) and (ii) are easy. We prove (iii).

($\rightarrow$) Reasoning in T assume $f(\Box\varphi)$. We shall consider two subcases depending on whether S_0 or $\neg S_0$ holds.

Assume S_0. Then $f(\Box\varphi)$ is equivalent to $\Phi(\Box\varphi)$. By (iv) of the previous lemma, we obtain $P \vdash \varphi$, therefore for some n, $P_n \vdash \varphi$. Since we assumed S_0, $h(n) = 0$ and, by the provable Σ_1-completeness,

$$\mathsf{Prov}_T(\ulcorner h(\dot{n}) = 0 \wedge \text{“}P_{\dot{n}} \vdash \varphi\text{”} \urcorner).$$

From Lemma 82 we conclude that $\mathsf{Prov}_T(\ulcorner S_\varphi \urcorner)$ and $\mathsf{Prov}_T(\ulcorner f(\varphi)\urcorner)$.

Assume $\neg S_0$. Then $f(\Box\varphi)$ implies $S_{\Box\varphi}$, hence for some n, $h(n) \Vdash \Box\varphi$. By our agreement on $\Vdash$ this means, in particular, that $h(n) \neq 0$. By the provable Σ_1-completeness, we obtain $\mathsf{Prov}_T(\ulcorner h(\dot{n}) \Vdash \Box\varphi\urcorner)$. Since $h(n) \neq 0$, by Lemma 82, T proves that the limit of h is a proper submodel of $h(n)$. This implies that φ holds in the limit of h, that is, $\mathsf{Prov}_T(\ulcorner S_\varphi \urcorner)$ and $\mathsf{Prov}_T(\ulcorner f(\varphi)\urcorner)$.

($\leftarrow$) Assume $\mathsf{Prov}_T(\ulcorner f(\varphi)\urcorner)$ and S_0. We have $\mathsf{Prov}_T(\ulcorner S_0 \vee S_\varphi \urcorner)$. Let n be the code of a T-proof of $S_0 \vee S_\varphi$. Since we assumed S_0, there holds $h(n) = 0$. Then $P_n \vdash \varphi$, otherwise h would make a jump at stage $n+1$ contradicting S_0. Thus, $P \vdash \varphi$ and, by Lemma 85(iii), $\Phi(\varphi)$, which implies $f(\Box\varphi)$, as required.

Assume now $\neg S_0$. Again, let n be the code of a T-proof of $S_0 \vee S_\varphi$ large enough to have $h(n) \neq 0$. If $h(n) \Vdash \Box\varphi$, then $h(n+1) = h(n)$, otherwise $h(n+1)$ will be the least submodel of $h(n)$ forcing $\neg\varphi$. In both cases we have $h(n+1) \Vdash \Box\varphi$. By monotonicity, this also implies that $h(m) \Vdash \Box\varphi$, for all $m \geq n+1$, that is, $S_{\Box\varphi}$. Hence, $f(\Box\varphi)$ holds. ■

COROLLARY 87. *For any modal formula φ,*

$$T \vdash f(\varphi) \leftrightarrow f_T(\varphi).$$

Now we can easily prove that f realizes the propositional theory P.

LEMMA 88. *For any modal formula φ, $T \vdash f_T(\varphi)$ iff $P \vdash \varphi$.*

Proof. Assume $P \vdash \varphi$, then for some n, $P_n \vdash \varphi$. By Σ_1-completeness, we obtain

$$T \vdash h(\bar{n}) = 0 \wedge \text{“}P_{\bar{n}} \vdash \varphi\text{”}.$$

By Lemma 84, conclude that $T \vdash S_\varphi$ and hence $T \vdash f(\varphi)$.

Vice versa, assume that $T \vdash f(\varphi)$ and $P \nvdash \varphi$. Let n be the code of a T-proof of $S_0 \vee S_\varphi$. We obviously have $P_n \nvdash \varphi$, hence $h(n+1) \neq 0$, which is impossible in the standard model. ■

Proof of Lemma 85. The proof is essentially a construction within T of a maximal consistent set containing $P \cup \{\neg\Box\varphi : P \nvdash \varphi\}$. It is easy to see by the s.d.p. of P that such a set exists (externally).

We shall deal with finite binary strings σ, τ, etc.; $|\sigma|$ denotes the length of σ; $\sigma(i)$ is the i-the element of σ; $\sigma <_{\text{lex}} \tau$ means that σ precedes τ

lexicographically and $|\sigma| = |\tau|$. We shall use the same notation within arithmetic, assuming some natural elementary encoding.

For any string σ such that $|\sigma| = n$ let Φ_σ denote the conjunction of $\{\varphi : \varphi < n \wedge \sigma(\varphi) = 1\}$ and $\{\neg\varphi : \varphi < n \wedge \sigma(\varphi) = 0\}$. We define the formulas $V(\sigma)$ and $U(\sigma)$ as follows:

$$\begin{aligned} V(\sigma) &= \forall\psi\,(\text{“}P \vdash \Phi_\sigma \to \Box\psi\text{”} \to \text{“}P \vdash \psi\text{”}) \\ U(\sigma) &= V(\sigma) \wedge \forall\tau\,(\tau <_{\text{lex}} \sigma \to \neg V(\tau)). \end{aligned}$$

The formula $\Phi(\psi)$ is then defined by $\exists\sigma\,(\psi < |\sigma| \wedge U(\sigma) \wedge \sigma(\psi) = 1)$.

Claim: for each n, $T \vdash \exists\sigma\,(|\sigma| = \bar{n} \wedge V(\sigma))$. Reasoning within T assume that, for each string σ of length n, $\neg V(\sigma)$. Then for any such σ there is a formula ψ_σ such that $P \vdash \Phi_\sigma \to \Box\psi_\sigma$ and $P \nvdash \psi_\sigma$. Since $\bigvee_{|\sigma|=n} \Phi_\sigma$ is a tautology, we infer $P \vdash \bigvee_{|\sigma|=n} \Box\psi_\sigma$. By the (provable) s.d.p. of P, $P \vdash \psi_\sigma$ for some σ, a contradiction.

From the claim we also conclude that, for any n,

$$T \vdash \exists!\sigma\,(|\sigma| = \bar{n} \wedge U(\sigma)). \tag{2}$$

Indeed, since n is standard, the least element principle up to n reduces to a tautology and is provable in T. The minimal string is obviously unique.

Now we check the properties (i)–(iv) of Φ.

(i) Assume $\Phi(\bot)$, then for some σ such that $\bot < |\sigma|$ we have $\sigma(\bot) = 1$ and $V(\sigma)$. But if $\sigma(\bot) = 1$, then $\Phi_\sigma \leftrightarrow \bot$, therefore $P \vdash \Phi_\sigma \to \Box\bot$, so $V(\sigma)$ implies $P \vdash \bot$, which is impossible by the s.d.p. The case $\Phi(\top)$ is treated similarly.

(ii) Assume $\Phi(\varphi)$ and $\Phi(\varphi \to \psi)$. Let σ be a string longer than all of φ, ψ and $\varphi \to \psi$ such that $U(\sigma)$ holds. By the uniqueness of σ in (2), we must have $\sigma(\varphi) = \sigma(\varphi \to \psi) = 1$. But then necessarily $\sigma(\psi) = 1$, for otherwise $\Phi_\sigma \leftrightarrow \bot$ and we get a contradiction in P as in (i).

Conversely, assume either $\neg\Phi(\varphi)$ or $\Phi(\psi)$. Again, let σ be long enough so that $\sigma(\varphi) = 0$ or $\sigma(\psi) = 1$. In each case $\sigma(\varphi \to \psi) = 1$, lest P should be inconsistent.

(iii) Assume $P \vdash \varphi$. Let σ be longer than φ and $U(\sigma)$. We show that $\sigma(\varphi) = 1$. Suppose the contrary, then $P \vdash \Phi_\sigma \to \neg\varphi$ and hence $P \vdash \Phi_\sigma \to \bot$. From $V(\sigma)$ it follows that $P \vdash \bot$, which is impossible. Therefore, $\Phi(\varphi)$.

(iv) Assume $\Phi(\varphi)$ and let σ be a sufficiently long string such that $U(\sigma)$ and $\sigma(\Box\varphi) = 1$. Then tautologically $P \vdash \Phi_\sigma \to \Box\varphi$. By the definition of $V(\sigma)$, $P \vdash \varphi$. ■

This completes the proof of Lemma 80 and thereby the proof of Shavrukov's theorems. ■

7.10 Arbitrary subalgebras

Shavrukov's theorems characterize r.e. subalgebras of provability algebras. Here we briefly consider the case of arbitrary subalgebras.

A countable algebra is called *locally positive* if it is numerated and each one of its finitely generated subalgebras, with the induced numeration, is positive. Dually speaking, a propositional theory P is called locally r.e. if the fragment of P in the first n variables is r.e., for any n. Obviously, any subalgebra of a positive algebra and thus, any subalgebra of a provability algebra, is locally positive. V. Shavrukov [Shavrukov, 1993b] gave an example of a locally positive Magari algebra that is not positive. For Σ_1-unsound theories local positivity, together with the obvious consideration of the characteristic, turns out to be sufficient for the embeddability.

THEOREM 89. *Suppose T is not Σ_1-sound. A countable Magari algebra $\mathcal{A}$ is embeddable into $\mathcal{M}_T$ iff $\mathcal{A}$ is locally positive and the characteristic of $\mathcal{A}$ equals $ch(T)$.*

A more important characterization of arbitrary subalgebras of Σ_1-sound theories was given by D. Zambella (unpublished), but it is slightly less elegant. We call a propositional theory P (and the corresponding quotient algebra $\mathbf{Fr}(\omega)/P$) *uniformly s.d.*, if for every $n \in \omega$ there is a recursive sequence of formulas $P_n(m)$, $m = 0, 1, \ldots$ in the variables $p_0, \ldots, p_n$ such that

- $\{P_n(m) : n, m \in \omega\}$ axiomatize P;
- $P_n(m)$ has s.d.p.;
- $P_n(m+1) \vdash P_n(m)$;
- $P_{n+1}(m) \vdash P_n(m)$.

THEOREM 90 (Zambella). *Let T be a Σ_1-sound theory. $\mathcal{A}$ is embeddable into $\mathcal{M}_T$ iff $\mathcal{A}$ is locally positive and uniformly s.d.*

7.11 Isomorphisms of provability algebras

Having isolated an interesting class of algebraic structures mathematicians often initiate the program of classifying all such structures modulo isomorphism: from a purely algebraic point of view isomorphic structures are the same. Even when this task in its full extent happens to be too difficult, the study can provide partial answers and useful insights into the structure of individual algebras. In the case of provability algebras there is really not much hope for any such classification. In this section we review what little we know. To get some feeling of the matter we first discuss the relationships between provability algebras and free Magari algebras.

EXAMPLE 91. Let $\mathcal{A}$ be the free algebra on 0 generators, $\mathbf{Fr}(0)$, and let $\mathcal{B}$ be the Magari algebra of all finite and cofinite subsets of natural numbers ω with the standard boolean operations and the operation $\Box$ defined as follows:

$$\Box X = \{x \in \omega : \forall y \in \omega\, (y < x \Rightarrow y \in X)\}.$$

Recall that $\mathcal{A}$ is the Lindenbaum algebra of the letterless fragment of GL.

The notion of *trace* of a modal formula introduced in Section 2.7 defines an isomorphism between $\mathcal{A}$ and $\mathcal{B}$:

$$f : \varphi \longmapsto \omega \setminus tr(\varphi).$$

Indeed, the normal form theorem for letterless formulas (Theorem 15) shows that f is a one-to-one and onto homomorphism.

From this example we can conclude that $\mathbf{Fr}(0)$ is not isomorphic to any provability algebra $\mathcal{M}_T$, for T containing EA. One reason is that $\mathbf{Fr}(0)$ is not dense (the formulas $\neg F_n$ are the minimal elements above 0), whereas, by Proposition 64, $\mathcal{M}_T$ is. Of course, $\mathbf{Fr}(0)$ is naturally embeddable into $\mathcal{M}_T$, for any T of infinite characteristic, as its prime subalgebra.

Free algebras on more than 0 generators do not have such a clear representation as in Example 91. There is a representation based on a universal Kripke model as described, e.g., in [Rybakov, 1989; Artemov and Beklemishev, 1993; Grigolia, 1987]. However, this model is not as nice as the one for $\mathbf{Fr}(0)$. It is well-known (see e.g. [Artemov and Beklemishev, 1993]) that $\mathbf{Fr}(\alpha)$ are not dense for $\alpha < \omega$, but $\mathbf{Fr}(\omega)$ is.

PROPOSITION 92. *None of the provability algebras $\mathcal{M}_T$ is embeddable into (let alone isomorphic to) a free algebra.*

Proof. If $ch(T) < \infty$, the algebra $\mathcal{M}_T$ is clearly not embeddable into a free one. If $ch(T) = \infty$, we use the following observation: in $\mathcal{M}_T$ there is an element φ such that $\varphi \neq \bot$ and $\varphi \leq \Diamond^n \top$, for all n.

If T is Σ_1-sound, one can simply take $\mathsf{RFN}_{\Sigma_1}(T)$ for φ. A little arithmetical fixed point argument works more generally. However, we would like to apply the recently acquired heavy weaponry. Consider the propositional theory P axiomatized by the formulas $\{p \to \Diamond^n \top : n < \omega\}$. It is easy to verify by a Kripke model argument that P (provably) satisfies s.d.p.. Hence, by Shavrukov's theorems, it is realizable in T. Let φ be the realization of p.

Now we observe that in a free algebra such a φ cannot exist by the following

Claim: if $\mathsf{GL} \vdash \varphi \to \Diamond^n \top$, for each n, then $\mathsf{GL} \vdash \neg\varphi$.

Indeed, if $\mathsf{GL} \nvdash \neg\varphi$, then there is a finite model $\mathcal{K}$ such that $\mathcal{K} \Vdash \varphi$. Let n be the height of $\mathcal{K}$. Then $\mathcal{K} \Vdash \varphi \wedge \neg\Diamond^{n+1}\top$, which contradicts our assumption. ■

In contrast with the previous proposition, recall that by the uniform Solovay theorem the free algebra $\mathbf{Fr}(\omega)$ is embeddable into $\mathcal{M}_T$, if $ch(T) = \infty$. The following interesting statement is a direct corollary of Shavrukov's theorems.

COROLLARY 93. *If T and U are Σ_1-sound extensions of EA, then $\mathcal{M}_T$ is embeddable into $\mathcal{M}_U$.*

Proof. Both algebras are positive and satisfy s.d.p.. ■

Thus, for example, $\mathcal{M}_{\mathsf{ZF}}$ is isomorphic to a subalgebra of $\mathcal{M}_{\mathsf{PA}}$, which may seem rather puzzling at first sight. All provability algebras of Σ_1-sound theories have the same subalgebras (modulo isomorphism).

V. Shavrukov [Shavrukov, 1993a] obtained the following beautiful result that we regretfully must leave without a proof.

THEOREM 94. *Let T and U be Σ_1-sound theories containing EA such that $U \vdash \mathsf{RFN}_{\Sigma_1}(T)$. Then $\mathcal{M}_T$ and $\mathcal{M}_U$ are not isomorphic.*

It follows, for example, that the provability algebras of EA, $I\Sigma_1$, PA, ZF, etc., are all pairwise nonisomorphic. This theorem shows, once again, that the notion of provability algebra is much richer than that of the Lindenbaum boolean algebra. Its proof relies on the difference in the rate of growth of provably total computable functions of the theories T and U and borrows some ideas from the Blum complexity theory.

As a complement to Theorem 94, V. Shavrukov [Shavrukov, 1997a] also obtained a positive result on the existence of isomorphisms between certain provability algebras. We formulate it in a somewhat simplified form.

THEOREM 95. *Assume that T and U contain $I\Sigma_1$ and prove the same boolean combinations of Σ_1-sentences. Assume further that this fact is provable both in T and in U. Then $\mathcal{M}_T$ is (recursively) isomorphic to $\mathcal{M}_U$.*

From this result one infers that the provability algebras for the following pairs of theories are isomorphic: $(\mathsf{PA}, \mathsf{ACA}_0)$, $(\mathsf{ZF}, \mathsf{GB})$. The result also holds, e.g., for the pair $(\mathsf{PRA}, I\Sigma_1)$, although it is not covered by Theorem 95 as we formulated it. De facto, Theorem 95 holds for all extensions of EA closed under the so-called Σ_1-*collection rule*, which includes all the 'usual' extensions of EA.

Finally, here is another relevant result of V. Shavrukov [Shavrukov, 1997a] that will be useful in the next section.

THEOREM 96. *Assume that T contains EA and φ is a sentence not equivalent to any boolean combination of Σ_1-sentences in T. Then there is an isomorphism $f : \mathcal{M}_T \to \mathcal{M}_T$ such that $f(\varphi) \neq \varphi$.*

7.12 First order theories of provability algebras

Universal theory

Recall that the identities of provability algebras are described by the provability logics $\boldsymbol{PL}_T(T)$. More generally, we also have a nice characterization of the set of universal formulas valid in $\mathcal{M}_T$, denoted $\mathbf{Th}_\forall(\mathcal{M}_T)$.

Any such formula has the form $\forall\vec{p}\,\chi(\vec{p})$, where $\chi(\vec{p})$ is a boolean combination of term equalities of the form $\varphi_i(\vec{p}) = \psi_j(\vec{p})$. Given a realization f of the variables $\vec{p}$,

$$\mathcal{M}_T \vDash \varphi_i(\vec{p}) = \psi_j(\vec{p}) \iff \mathbb{N} \vDash f_T(\Box(\varphi_i(\vec{p}) \leftrightarrow \psi_j(\vec{p}))).$$

Therefore, $\mathcal{M}_T \vDash \forall\vec{p}\,\chi(\vec{p})$ iff $\mathbb{N} \vDash f_T(\tilde{\chi}(\vec{p}))$, for all realizations f, where $\tilde{\chi}$ denotes the result of replacing in χ every equality $\varphi_i(\vec{p}) = \psi_j(\vec{p})$ by $\Box(\varphi_i(\vec{p}) \leftrightarrow \psi_j(\vec{p}))$. Hence, we obtain the following result.

THEOREM 97. *$\mathcal{M}_T \vDash \forall\vec{p}\,\chi(\vec{p})$ iff $\tilde{\chi} \in \boldsymbol{PL}_T(\mathsf{TA})$.*

Since all the truth provability logics are decidable, we obtain the following corollary.

COROLLARY 98. *If T is an elementary presented extension of EA, then $\mathbf{Th}_\forall(\mathcal{M}_T)$ is decidable.*

An example: admissible rules in PA

In perfect analogy with the definition of admissible rules in a propositional logic L we now define propositional admissible rules in an arithmetical theory T. A inference rule

$$\frac{\varphi_1, \ldots, \varphi_n}{\psi}\ (R)$$

is *admissible* in a T, if whenever $T \vdash f_T(\varphi_i)$, for $i = 1, \ldots, n$, there holds $T \vdash f_T(\psi)$, for any realization f.

It is easy to see that any admissible rule in PA is also admissible in GL. The converse is actually not true. We have the following straightforward characterization.

PROPOSITION 99. *A rule (R) is admissible in T iff*

$$\mathcal{M}_T \vDash \forall\vec{p}\,(\varphi_1(\vec{p}) \wedge \ldots \wedge \varphi_n(\vec{p}) = \top \to \psi(\vec{p}) = \top).$$

Thus, admissibility of a rule in T is expressible by a particular universal formula in $\mathcal{M}_T$. From the previous corollary we conclude that the admissibility of propositional inference rules in T is decidable and reducible to the derivability in S, if T is sound.

COROLLARY 100. *(R) is admissible in* PA *iff* $\mathsf{S} \vdash \Box\varphi_1 \wedge \ldots \wedge \Box\varphi_n \to \Box\psi$.

Consider the following example (suggested by J. Joosten):

$$\frac{\Box p \to \Box\bot \quad \Box\neg p \to \Box\bot}{\bot}.$$

This rule is admissible in GL because the two-element linear Kripke frame shows that for no formula φ do we have

$$\mathsf{GL} \vdash \Diamond\top \to (\Diamond\varphi \wedge \Diamond\neg\varphi).$$

On the other hand, this rule is not admissible in PA. Substitution of the well-known *Rosser sentence* for p makes both premises of the rule provable. Alternatively, one can just check by a Kripke model argument that

$$\mathsf{S} \nvdash [\Box(\Box p \to \Box\bot) \wedge \Box(\Box\neg p \to \Box\bot)] \to \Box\bot.$$

Full first order theory

For several years one of the major questions in the area was the problem of the decidability of the full first order theory of the provability algebra of PA. This theory can be perceived in two, obviously equivalent, ways.

First of all, it is the set of all first order formulas in the language of Magari algebras that are valid in $\mathcal{M}_{\mathsf{PA}}$. We can also look at it as at the set of all valid second order propositional modal formulas with the following restrictions on the occurrences of propositional quantifiers and modalities:

- no quantifier occurs within the scope of $\Box$;
- no variable occurs outside the scope of $\Box$.

Of course, the quantifiers range over arbitrary arithmetical sentences. For example, the formula

$$\forall p_1, p_2 \exists q\, \Box(\Box q \leftrightarrow (\Box p_1 \vee \Box p_2))$$

is valid in $\mathcal{M}_T$, for any T containing EA (this follows from the so-called FGH-principle [Smoryński, 1981; Visser, 2002a]). Similarly, formula

$$\forall p\, (\Diamond p \to \exists q\, (\neg\Box(p \to q) \wedge \neg\Box(\neg p \to q)))$$

represents Rosser's theorem for the theory $T + p$. We denote the first order theory of $\mathcal{M}_T$ by $\mathbf{Th}(\mathcal{M}_T)$.

V. Shavrukov answered the main problem by the following remarkable result [Shavrukov, 1997b] that also yielded solutions to a number of related questions.

THEOREM 101. *If $ch(T) = \infty$ and T contains EA, then neither $\mathbf{Th}(\mathcal{M}_T)$, nor any of its subtheories is decidable.*

For Σ_1-sound theories T one has a considerable strengthening.

THEOREM 102. *If T is a Σ_1-sound theory containing* EA*, then* $\mathbf{Th}(\mathcal{M}_T)$ *is mutually interpretable with* TA.

Since the interpretation constructed by Shavrukov does not actually depend on T, this yields the following corollary.

COROLLARY 103. *For T a Σ_1-sound theory,* $\mathbf{Th}(\mathcal{M}_T)$ *is not arithmetic. The same holds for the first order theory of any collection of provability algebras of such theories.*

In particular, it would be hopeless to look for an r.e. axiomatization of $\mathbf{Th}(\mathcal{M}_{\mathsf{PA}})$. The proofs of these results are ingenious and technically complicated. We will not go into them, but remark that the main idea and difficulty in the proof was to find a first order definition of the set of elements in $\mathcal{M}_T$ that would allow to model the structure of natural numbers (recall that the ordering of $\mathcal{M}_T$ is dense!). This has been achieved by the following lemma, which is interesting in its own right.

LEMMA 104. *Suppose T is Σ_1-sound. There is a first order formula $N(x)$ in the language of Magari algebras such that, for any $x \in \mathcal{M}_T$,*

$$\mathcal{M}_T \vDash N(x) \iff x \in \{\Diamond_T^n \top : n \in \omega\}.$$

With this lemma at hand, it is then relatively easy to show the undecidability of $\mathbf{Th}(\mathcal{M}_T)$, e.g., by interpreting the theory of finite partial orderings in it (see [Artemov and Beklemishev, 1993]). To interpret the full true arithmetic in $\mathbf{Th}(\mathcal{M}_T)$, however, some more powerful machinery seems to be necessary. Here, Shavrukov's theorems on the subalgebras of $\mathcal{M}_T$ provided a useful tool that allowed to give a short construction.

The proof of Theorem 101 in full generality bears on the same technical ideas, but is more involved. V. Shavrukov showed how to directly simulate Minsky's *monogenic normal canonical systems* within $\mathcal{M}_T$. One important corollary of Lemma 104 is a strengthening of Theorem 94 stating that certain provability algebras are not elementary equivalent.

THEOREM 105. *Let T and U be Σ_1-sound theories containing* EA *such that $U \vdash \mathsf{RFN}_{\Sigma_1}(T)$. Then there is a formula φ such that $\varphi \in \mathbf{Th}(\mathcal{M}_T)$ and $\varphi \notin \mathbf{Th}(\mathcal{M}_U)$.*

For the record we also mention some results on first order theories of Magari algebras that have been obtained prior to Shavrukov's.

[Montagna, 1980] and [Smoryński, 1982] showed that the first order theory of the class of *all* Magari algebras (and any subtheory of this theory) is undecidable.

[Artemov and Beklemishev, 1993] investigated first order theories of free algebras. It turns out that $\mathbf{Th}(\mathbf{Fr}(0))$ is decidable and, in fact, is mutually interpretable with the *weak monadic second order theory* of the linear ordering of ω. The interpretation is essentially provided by the isomorphism of $\mathbf{Fr}(0)$ and the algebra of finite and cofinite subsets of ω. It is known by A. Meyer [Meyer, 1975] that the decision procedure for the theory in question cannot be elementary. The theories $\mathbf{Th}(\mathbf{Fr}(\alpha))$ for $\alpha > 0$ are all undecidable.

There are many questions about first order theories of provability algebras that are still open. For example, we do not know if the $\forall\exists$- and $\forall\exists\forall$-fragments of $\mathbf{Th}(\mathcal{M}_{\mathsf{PA}})$ are decidable. From Shavrukov's proof of Theorem 101 it only follows that the $\forall\exists\forall\exists$-fragment is undecidable. We also know very little about definable elements of $\mathcal{M}_{\mathsf{PA}}$. Of course, the elements of its prime subalgebra are definable. Since definable elements cannot be moved by automorphisms, from Theorem 96 we immediately conclude that sentences not equivalent to boolean combinations of Σ_1-sentences cannot be definable. The situation with the rest of the elements of $\mathcal{M}_{\mathsf{PA}}$ is unknown. It is also open, whether the provability algebras of PA and $\mathsf{PA} + \mathsf{Con}(\mathsf{PA})$ are isomorphic.

8 BIMODAL AND POLYMODAL PROVABILITY LOGIC

An obvious way to increase the expressive power of modal language is to consider several interacting provability operators, which naturally leads to *bi-* and *polymodal provability logic*. In this section we overview this large and non-uniform area.

Perhaps, the most natural provability interpretation of polymodal language is the understanding of modalities as provability predicates in some elementary presented theories containing EA. A modal description of two such provability predicates is, in general, already a considerably more difficult task than a characterization of each one's provability logic. There is no single system that can be justifiably called *the* bimodal provability logic — rather, we know particular systems for different natural pairs of theories, and none of those systems occupies any privileged place among the others. Numerous isolated results accumulated in this area, so far, give us little information as to a possible general classification of bimodal provability logics for pairs of (sound) r.e. theories. This problem is one of the major remaining open problems in provability logic.

On a par with the interpretation of modalities as the *usual* provability predicates we consider some interpretations, where modalities correspond to more general provability-like concepts of essentially semantical nature. For example, we consider the predicates of *n-provability* expressing the provability from the set of all true Π_n-sentences. The corresponding polymodal logic, formulated by G. Japaridze, appears to be particularly useful for the applications in proof theory (see Section 10). It seems at present that further applications will require more work to be done on the polymodal analysis of the other *strong*, that is, not r.e. provability concepts, such as ω-provability in the second-order arithmetic and validity in particular classes of models of set theory.

In addition to the studies of the usual and strong provability predicates, a considerable amount of work has been done on studying interaction of the provability operator with some exotic provability concepts, such as Rosser's provability predicate, Feferman's provability predicate, and some others. These results can also be coached in terms of bimodal logic. Below we overview some of these developments as well.

8.1 Bimodal logics for pairs of r.e. theories

The language $\mathcal{L}(\Box, \triangle)$ of bimodal provability logic is obtained from that of propositional calculus by adding two unary modal operators $\Box$ and $\triangle$. Let (T, U) be a pair of elementary presented theories and let f be an arithmetical realization (cf. Section 3.3). An *arithmetical interpretation* $f_{T,U}(\varphi)$ *of a formula* φ *w.r.t.* (T, U) translates $\Box$ as provability in T and $\triangle$ as that in U:

$$f_{T,U}(\Box\varphi) = \mathsf{Prov}_T(\ulcorner f_{T,U}(\varphi)\urcorner), \qquad f_{T,U}(\triangle\varphi) = \mathsf{Prov}_U(\ulcorner f_{T,U}(\varphi)\urcorner).$$

The provability logic for (T, U) is the collection of all $\mathcal{L}(\Box, \triangle)$-formulas φ such that $T \cap U \vdash f_{T,U}(\varphi)$, for every arithmetical realization f. It is denoted $\boldsymbol{PL}_{T,U}$. In general, similarly to the unimodal case, one can consider bimodal provability logics for (T, U) relative to an arbitrary metatheory V. $\boldsymbol{PL}_{T,U}(V)$ is the set of all formulas φ such that $V \vdash f_{T,U}(\varphi)$, for every arithmetical realization f. Thus, $\boldsymbol{PL}_{T,U}$ corresponds to $V = T \cap U$.

Basic bimodal logic

Basic work on bimodal provability logic has been done by C. Smoryński and T. Carlson [Smoryński, 1985; Carlson, 1986]. Not too much can a priori be said about $\boldsymbol{PL}_{T,U}$, for arbitrary T and U. Firstly, $\boldsymbol{PL}_{T,U}$ is closed under *modus ponens*, substitution, $\Box$- and $\triangle$-necessitation rules.[12] Secondly, $\boldsymbol{PL}_{T,U}$ has to be an extension of the following bimodal logic CS:

Axioms: (i) schemes of GL for $\Box$ and for $\triangle$;

(ii) $\Box\varphi \to \triangle\Box\varphi$;

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

Rules: *modus ponens*, $\varphi/\Box\varphi$, $\varphi/\triangle\varphi$.

[12] Notice that neither $\boldsymbol{PL}_{T,U}(T)$ nor $\boldsymbol{PL}_{T,U}(U)$ will in general be closed under both necessitation rules.

Arithmetical soundness of CS is expressed by

PROPOSITION 106. *For any theories T, U one has* $\mathsf{CS} \subseteq \boldsymbol{PL}_{T,U}(\mathsf{EA})$.

Proof. The validity of (ii) and (iii) is just a consequence of provable Σ_1-completeness and the fact that both provability predicates are Σ_1. ■

Carlson models

Suitable Kripke-style models for CS are introduced by the following definition.

DEFINITION 107. A *Carlson model* $\mathcal{K} = (K, M_0, M_1, \prec, \Vdash)$ is a usual Kripke model $(K, \prec, \Vdash)$ for GL equipped with two distinguished subsets $M_0,\ M_1 \subseteq K$. The forcing of bimodal formulas on $\mathcal{K}$ is defined according to the following clauses:

1. $x \nVdash \bot$, $x \Vdash \top$;
2. $x \Vdash \varphi \to \psi \iff (x \nVdash \varphi \text{ or } x \Vdash \psi)$;
3. $x \Vdash \Box\varphi \iff \forall y \in M_0 (x \prec y \Rightarrow y \Vdash \varphi)$;
4. $x \Vdash \triangle\varphi \iff \forall y \in M_1 (x \prec y \Rightarrow y \Vdash \varphi)$.

As usual, we write $\mathcal{K} \vDash \varphi$, if $x \Vdash \varphi$, for each $x \in K$.

Obviously, CS is sound with respect to Carlson models. The standard canonical model construction followed by a filtration and unravelling argument yields the following completeness result (see [Smoryński, 1985]).

THEOREM 108. *The following statements are equivalent:*

(i) $\mathsf{CS} \vdash \varphi$;

(ii) $\mathcal{K} \vDash \varphi$, *for each Carlson model* $\mathcal{K}$;

(iii) $\mathcal{K} \vDash \varphi$, *for each finite Carlson model* $\mathcal{K}$.

COROLLARY 109. CS *is decidable.*

Arithmetical completeness of CS

[Smoryński, 1985] showed that CS is, indeed, the minimal bimodal provability logic, that is, coincides with $\boldsymbol{PL}_{T,U}$ for a certain pair of finite extensions T, U of Peano arithmetic. [Beklemishev, 1992] proved that there is a pair of provability predicates for Peano arithmetic itself such that the corresponding bimodal provability logic coincides with CS. Such predicates can be called *independent* in the sense that they ‘know’ as little about each other as is possible in principle. However, neither the theories in Smoryński's example, nor the independent provability predicates are natural — they are constructed by tricky diagonalization. Thus, we are in the curious situation where the bimodal logic CS, which structurally occupies a privileged place among the provability logics, does not correspond to any (known) *natural* pair of theories.

THEOREM 110 (Smoryński). *Let V be a Σ_1-sound theory. There is a pair (T,U) of finite extensions of V such that*

$$\boldsymbol{PL}_{T,U} = \boldsymbol{PL}_{T,U}(V) = \mathsf{CS}.$$

Proof. Consider a translation $(\cdot)^*$ of the language $\mathcal{L}(\Box,\triangle)$ into $\mathcal{L}(\Box,c_0,c_1)$ that preserves variables and boolean connectives and satisfies

$$(\Box\varphi)^* = \Box(c_0 \to \varphi^*); \quad (\triangle\varphi)^* = \Box(c_1 \to \varphi^*).$$

Here c_0 and c_1 are propositional variables that do not occur in the language $\mathcal{L}(\Box,\triangle)$.

We claim: $\mathsf{CS} \vdash \varphi \iff \mathsf{GL} \vdash \varphi^*$, for any formula $\varphi \in \mathcal{L}(\Box,\triangle)$.

Indeed, the implication $(\Rightarrow)$ is easy to check by induction on the length of a derivation in CS. For a proof of $(\Leftarrow)$ one argues by contraposition.

If $\mathsf{CS} \nvdash \varphi$ then there is a Carlson model $\mathcal{K}$ such that $\mathcal{K} \nvDash \varphi$. Define $x \Vdash c_0 \Leftrightarrow x \in M_0$ and $x \Vdash c_1 \Leftrightarrow x \in M_1$. This makes $\mathcal{K}$ a Kripke model for $\mathcal{L}(\Box,c_0,c_1)$ falsifying φ^* ($\Box$ corresponds to the accessibility relation on $\mathcal{K}$) and proves our claim.

By the uniform arithmetical completeness theorem (Theorem 73) we obtain a realization f such that for any formula $\varphi \in \mathcal{L}(\Box,c_0,c_1)$,

$$\mathsf{GL} \vdash \varphi \iff V \vdash f_V(\varphi).$$

Let $T := V + f_V(c_0)$ and $U := V + f_V(c_1)$. This agrees with the translation $(\cdot)^*$, that is,

$$V \vdash f_{T,U}(\varphi) \leftrightarrow f_V(\varphi^*),$$

so we obtain

$$\mathsf{CS} \nvdash \varphi \Rightarrow \mathsf{GL} \nvdash \varphi^* \Rightarrow V \nvdash f_V(\varphi^*) \Rightarrow V \nvdash f_{T,U}(\varphi).$$

This proves $\boldsymbol{PL}_{T,U}(V) = \mathsf{CS}$.

For the same theories (T,U) we also have $\boldsymbol{PL}_{T,U} = \mathsf{CS}$ because

$$\begin{aligned} \varphi \in \boldsymbol{PL}_{T,U} &\iff (\Box\varphi \wedge \triangle\varphi) \in \boldsymbol{PL}_{T,U}(V) \\ &\iff \mathsf{CS} \vdash \Box\varphi \wedge \triangle\varphi \\ &\iff \mathsf{CS} \vdash \varphi. \end{aligned}$$

■

Types of bimodal provability logics

Deeper structural information on bimodal provability logics is provided by the Classification theorem for arithmetically complete unimodal logics. With every (normal) bimodal logic L containing CS we can associate its *type*:

$$(L)^{\Box} := \{\varphi \in \mathcal{L}(\Box) : L \vdash \triangle\varphi\}.$$

It is obvious that $(L)^{\Box}$ contains GL and is closed under modus ponens and substitution rules. Under the assumption of Σ_1-soundness of V we obviously have:

$$\boldsymbol{PL}_T(U) = (\boldsymbol{PL}_{T,U}(V))^{\Box}.$$

The Classification theorem shows that not every extension of GL is materialized as the type of a bimodal provability logic and gives us a description of all such possible types: GL_α, GL_β^-, S_β, D_β, $\alpha, \beta \subseteq \omega$, $\omega \setminus \beta$ finite. The set α also has to be r.e. because we assume U to be an r.e. theory. Indeed, the set of all formulas F_n such that $U \vdash F_n^T$ is r.e. and therefore so is α.

Natural bimodal provability logics

Apart from the above general observations, a number of particular bimodal provability logics for natural pairs of theories is known. These logics cover most of the examples of pairs of arithmetical theories that come to mind, but, unfortunately, are far from being an exhaustive list of all bimodal provability logics.

The best known system is the logic $\boldsymbol{PL}_{\mathsf{PA},\mathsf{ZF}}$ characterized by [Carlson, 1986] and independently (with a different interpretation in mind) by [Montagna, 1987b].[13] This logic can be axiomatized over CS by the principle of *essential reflexivity*

$$\mathsf{ER}: \quad \triangle(\Box\varphi \to \varphi).$$

A pair (T, U) is called an *essentially reflexive extension*, if U proves the local reflection principle for T. (In this case U also necessarily contains T.) Further examples of essentially reflexive extensions are: $(I\Sigma_1, \mathsf{PA})$, $(\mathsf{ACA}_0, \mathsf{ACA})$, etc. Here ACA denotes the second order arithmetic with arithmetical comprehension and full induction and ACA_0 is ACA with the induction formulated as a single axiom.

In the following we shall denote by $L \oplus A$ the closure of a normal (bimodal) logic L and an axiom schema A by modus ponens, substitution and both necessitation rules.

THEOREM 111 (Carlson). $\boldsymbol{PL}_{T,U} = \mathsf{CS} \oplus \mathsf{ER}$*, whenever the theories* T, U *are sound and* (T, U) *is an essentially reflexive extension.*

[13] It does not really matter here that PA and ZF are officially formulated in different languages. We may just assume that PA is interpreted into the language of ZF.

Thus, $\mathsf{CS} \oplus \mathsf{ER}$ is the *only* bimodal provability logic of type S and a maximal among the bimodal logics for pairs of sound elementary presented theories.

An extended treatment of $\boldsymbol{PL}_{\mathsf{PA},\mathsf{ZF}}$ is given in [Smoryński, 1985]. Suitable Kripke models for this logic are developed in [Visser, 1995].

Below we deal with more general *extensions*, that is, pairs of theories (T, U) such that U provably contains T. If (T, U) is an extension, then $\boldsymbol{PL}_{T,U}(\mathsf{EA})$ satisfies the additional *monotonicity* principle

$$\mathsf{M}: \quad \Box\varphi \to \triangle\varphi.$$

CS together with the monotonicity axiom will be denoted CSM.[14] Notice that M follows from ER over CS. CSM is sound and complete with respect to (finite) Carlson models $\mathcal{K}$ for which $M_0 = K$.

Now we describe two natural bimodal provability logics of type D, introduced in [Beklemishev, 1996]. The first one corresponds to pairs of theories (T, U) such that U is a finite extension of T and proves the local Σ_1-reflection principle for T. Typical examples are the pairs $(\mathsf{EA}, \mathsf{EA}^+)$, $(I\Sigma_m, I\Sigma_n)$, for $n > m \geq 1$, $(\mathsf{PA}, \mathsf{PA} + \mathsf{RFN}_{\Sigma_1}(\mathsf{PA}))$, etc. The logic can be axiomatized over CSM by the schema

$$\mathsf{EC}_\Sigma: \quad \triangle(\Box\sigma \to \sigma), \quad \sigma \in \Sigma,$$

where Σ denotes the set of all (possibly empty) disjunctions of formulas of the form $\Box\psi$ and $\triangle\psi$.

THEOREM 112. *If $U \vdash \mathsf{Rfn}_{\Sigma_1}(T)$ and U is a sound finite extension of T, then $\boldsymbol{PL}_{T,U} = \mathsf{CSM} \oplus \mathsf{EC}_\Sigma$.*

$\mathsf{CSM} \oplus \mathsf{EC}_\Sigma$ is the minimal bimodal provability logic of type D containing M. Indeed, if $\boldsymbol{PL}_T(U) = \mathsf{D}$, then U proves the local Σ_1-reflection principle for T, by Corollary 52. Hence, $\boldsymbol{PL}_{T,U}(\mathsf{EA})$ must also satisfy EC_Σ.

Another bimodal logic of type D corresponds to Π_1-essentially reflexive extensions of theories of bounded arithmetical complexity. An extension (T, U) is called Π_1-*essentially reflexive*, if $U \vdash \mathsf{Rfn}_{\Sigma_1}(T + \varphi)$ whenever $U \vdash \varphi$. Thus, a Π_1-essentially reflexive extension is never finite.

Examples of such extensions among fragments of PA are: $(\mathsf{EA}, \mathsf{PRA})$, $(I\Sigma_n, I\Sigma_{n+1}^R)$, for $n \geq 1$, $(\mathsf{EA}, I\Sigma_1^-)$, etc. The logic can be axiomatized over CSM by the schema

$$\mathsf{ER}_\Sigma: \quad \triangle\varphi \to \triangle(\Box(\varphi \to \sigma) \to \sigma), \quad \sigma \in \Sigma.$$

THEOREM 113. *If U is a sound and (provably) Π_1-essentially reflexive extension of T of bounded arithmetical complexity, then $\boldsymbol{PL}_{T,U} = \mathsf{CSM} \oplus \mathsf{ER}_\Sigma$.*

[14] CSM also stands for Carlson–Smoryński–Montagna as suggested by A. Visser.

Obviously, ER_Σ implies EC_Σ. We do not know whether $\mathsf{CSM}\oplus\mathsf{ER}_\Sigma$ is the maximal among all bimodal provability logics of type D containing CSM. Nor do we have examples of provability logics strictly between $\mathsf{CSM}\oplus\mathsf{EC}_\Sigma$ and $\mathsf{CSM}\oplus\mathsf{ER}_\Sigma$.

Let us now turn to the provability logics of type GL_ω. Two such logics for natural pairs of theories were described in [Beklemishev, 1994]. They both concern Π_1-*axiomatized* extensions of theories, that is, when U is obtained from T by adding Π_1-axioms only (and this fact is verifiable). Then $\boldsymbol{PL}_{T,U}$ contains an additional principle

$$\mathsf{P}: \quad \triangle\varphi\to\Box(\triangle\bot\vee\varphi)$$

which follows from provable Σ_1-completeness.

Similarly to the proof of Smoryński's theorem one can establish that $\mathsf{CSM}\oplus\mathsf{P}$ is the bimodal provability logic for a suitably general Π_1-axiomatized extension. $\mathsf{CSM}\oplus\mathsf{P}$ is complete with respect to Carlson models satisfying additional requirements that $M_0=K$ and M_1 is downwards closed [Beklemishev, 1994].

Let us call an extension (T,U) *infinitely confident*, if U proves k-times iterated consistency for T, for each $k\geq 0$. This essentially means that U believes that T has infinite characteristic. For such extensions $\boldsymbol{PL}_T(U)$ contains GL_ω and $\boldsymbol{PL}_{T,U}(\mathsf{EA})$ satisfies the additional principle

$$\mathsf{IC}: \quad \triangle\neg\Box^n\bot, \quad \text{for all } n\geq 1.$$

THEOREM 114. *If U is sound, infinitely confident, Π_1-axiomatized and finite extension of T, then $\boldsymbol{PL}_{T,U}=\mathsf{CSM}\oplus\mathsf{P}\oplus\mathsf{IC}$.*

Here are some typical examples of such pairs of theories: $(\mathsf{PA},\mathsf{PA}+\mathsf{Con}(\mathsf{ZF}))$, $(I\Sigma_1, I\Sigma_1+\mathsf{Con}(I\Sigma_2))$, etc.

The second system corresponds to (provably) reflexive Π_1-axiomatizable extensions, such as $(\mathsf{PA},\mathsf{PA}_\omega)$ or $(I\Sigma_1, I\Sigma_1+\{\mathsf{Con}(I\Sigma_n):n\geq 1\})$. An extension (T,U) is called *reflexive*, if $U\vdash\mathsf{Con}(T+\varphi)$ whenever $U\vdash\varphi$. Notice that every reflexive extension is infinitely confident and cannot be finite.

This logic can be axiomatized over $\mathsf{CSM}\oplus\mathsf{P}$ by the *reflexivity* axiom

$$\mathsf{R}: \quad \triangle\varphi\to\triangle\Diamond\varphi.$$

THEOREM 115. *If U is a sound, Π_1-axiomatized and provably reflexive extension of T, then $\boldsymbol{PL}_{T,U}=\mathsf{CSM}\oplus\mathsf{P}\oplus\mathsf{R}$.*

A remarkable property of the last logic is that it is the supremum of all provability logics for infinitely confident Π_1-axiomatized extensions of theories. Thus, all such logics lie between $\mathsf{CSM}\oplus\mathsf{P}\oplus\mathsf{IC}$ and $\mathsf{CSM}\oplus\mathsf{P}\oplus\mathsf{R}$.

Name	Axiom	Examples
M	$\Box\varphi \to \triangle\varphi$	general extensions
ER	$\triangle(\Box\varphi \to \varphi)$	$(\mathsf{PA}, \mathsf{ZF})$; $(I\Sigma_n, \mathsf{PA})$
EC_Σ	$\triangle(\Box\sigma \to \sigma)$, where $\sigma \in \Sigma$	$(I\Sigma_n, I\Sigma_{n+1})$
ER_Σ	$\triangle\varphi \to \triangle(\Box(\varphi \to \sigma) \to \sigma)$, where $\sigma \in \Sigma$	$(\mathsf{EA}, \mathsf{PRA})$; $(I\Sigma_n, I\Sigma^R_{n+1})$
P	$\triangle\varphi \to \Box(\triangle\bot \vee \varphi)$	Π_1-axiomatized extensions
IC	$\{\triangle\neg\Box^k\bot : k < \omega\}$	$(\mathsf{EA}, \mathsf{EA} + \mathsf{Con}(I\Sigma_1))$
R	$\triangle\varphi \to \triangle\Diamond\varphi$	$(I\Sigma_1, I\Sigma_1 + \{\mathsf{Con}(I\Sigma_n) : n < \omega\})$
$\mathsf{Cons}_{\mathcal{B}}$	$\triangle\beta \to \Box\beta$, where $\beta \in \mathcal{B}$	$(\mathsf{PRA}, I\Sigma_1)$, $(I\Sigma^R_n, I\Sigma_n)$

Table 1. Bimodal provability logic axioms

We also know that there really are some provability logics between these two [Beklemishev, 1994].

Finally, we describe yet another natural system of type GL formulated in [Beklemishev, 1996] that corresponds to finite extensions of theories of the form $(T, T+\varphi)$, where both $T+\varphi$ and $T+\neg\varphi$ are (provably) conservative over T w.r.t. boolean combinations of Σ_1-sentences. Examples of such pairs are $(\mathsf{PRA}, I\Sigma_1)$, $(I\Sigma^R_n, I\Sigma_n)$, for $n \geq 1$, and many others.[15]

This logic is axiomatized over CSM by the schema

$$\mathsf{Cons}_{\mathcal{B}}: \quad \triangle\beta \to \Box\beta, \quad \beta \in \mathcal{B},$$

where $\mathcal{B}$ denotes the set of boolean combinations of formulas of the form $\Box\psi$ and $\triangle\psi$.

THEOREM 116. *Assume both $T+\varphi$ and $T+\neg\varphi$ are provably conservative over T for boolean combinations of Σ_1-sentences and $U = T+\varphi$ is sound. Then $\boldsymbol{PL}_{T,U} = \mathsf{CSM} \oplus \mathsf{Cons}_{\mathcal{B}}$.*

The six bimodal logics described above essentially exhaust all nontrivial cases, currently known, for which natural provability logics are explicitly characterized.

Remarks

1. It is worth mentioning that all these systems are decidable, and a suitable Kripke-style semantics is known for each of them.

The logics CS, CSM and $\mathsf{CSM} \oplus \mathsf{P}$ are complete w.r.t. simple classes of Carlson frames and enjoy the finite model property. All the other considered logics do not behave as nicely.

[15] These conservation results are known to be provable in EA^+ but not in EA (see Section 10.1). However, EA^+ is much weaker than T for most of such examples.

$\mathsf{CSM} \oplus \mathsf{P} \oplus \mathsf{IC}$ is complete for a natural class of infinite Carlson frames (satisfying the condition that M_1 is downward closed and every point in M_1 has infinite depth). All such frames are infinite and the logic, obviously, does not satisfy the finite model property.

The other logics, such as $\mathsf{CSM} \oplus \mathsf{ER}$ and $\mathsf{CSM} \oplus \mathsf{EC}_\Sigma$, are not complete for any class of standard Kripke frames. This might have been a serious obstacle. Nonetheless, there are translations of these logics to CSM similar to those of the systems S and D to GL. This provides for all such systems a decision procedure and allows for an arithmetical completeness proof in the style of Solovay.

On the other hand, [Visser, 1995] devised a nice generalized Kripke semantics for $\mathsf{CSM} \oplus \mathsf{ER}$ which is sufficiently well-behaved though the models are infinite.

[Wolter, 1998] studied extensions of CSM using the notion of *subframe logic*. In particular, he showed that every finitely axiomatizable subframe logic containing CSM is decidable (see also [Chagrov *et al.*, 2001]). These methods may become useful for a solution of the general Classification problem for bimodal provability logics.

2. The arithmetical completeness proofs in each case are obtained by suitable modifications of the Solovay construction. Essentially, every theorem of this kind requires a new modification. In such constructions the techniques of D. Guaspari and P. Lindström of constructing partially conservative sentences are very useful [Guaspari, 1979; Lindström, 1984].

3. In all the considered cases of natural bimodal logics we also know the truth provability logic $\boldsymbol{PL}_{T,U}(\mathsf{TA})$. It is axiomatized by closing $\boldsymbol{PL}_{T,U}$ and the soundness schema $\triangle\varphi \to \varphi$ under modus ponens and substitution.

This relationship cannot, however, hold in general. Consider, e.g., a Π_1-axiomatized extension of theories that is reflexive but not provably so. Then the reflexivity axiom belongs to $\boldsymbol{PL}_{T,U}(\mathsf{TA})$ but not to $\boldsymbol{PL}_{T,U}$.

4. In general, the picture of bimodal provability logics for pairs of r.e. theories still has many white spots. For example, we do not know the axiomatizations of the logics for natural Π_1-conservative extensions such as $(\mathsf{EA}_\omega, \mathsf{EA}^+)$ or $(I\Sigma_1 + \{\mathsf{Con}(I\Sigma_n) : n < \omega\}, \mathsf{PA})$. The logics of pairs of incomparable fragments of PA such as $(I\Sigma_1, I\Pi_2^-)$ or $(I\Sigma_2^R, I\Pi_2^-)$ have never been investigated.

An important subproblem of the general classification problem for bimodal provability logics is to characterize all such logics for the most common types, such as D and GL_ω.

Polymodal provability logic

Most of the results in bimodal provability logic can be generalized to *polymodal logic*. Such a generalization is particularly natural in the modal-logical study of progressions of theories — a topic in proof theory that goes as far back as the work [Turing, 1939]. From the modal logical point of view, however, such a generalization, in all known cases, does not lead to essentially new phenomena compared to the bimodal logics, therefore we shall not go into any details here.

Polymodal analogues are known for natural provability logics due to Carlson and Beklemishev. Here, the modal operators correspond to theories of the original Turing–Feferman progressions of transfinitely iterated reflection principles and, thus, are indexed by ordinals of some constructive system of ordinal notation, say, the natural one up to ϵ_0. Iterating full reflection leads to the polymodal analogue of $\boldsymbol{PL}_{\mathsf{PA},\mathsf{ZF}}$, and transfinitely iterated consistency leads to a natural polymodal analogue of provability logics of type GL_ω. Successor ordinals correspond to finitely axiomatized extensions and limit ordinals to reflexive extensions (see [Beklemishev, 1991; Beklemishev, 1994]).

8.2 Logics with propositional constants

Some of the results on bimodal logics described in the previous section can be extended to the so-called *provability logics with propositional constants.*

Let $\mathcal{L}(\Box,\vec{c})$ be the language of GL equipped with a tuple $\vec{c} = (c_0,\dots,c_n)$ of new propositional constants. Fix an interpretation of $\vec{c}$ by choosing a tuple of arithmetical sentences $\vec{A} = (A_0,\dots,A_n)$. Given a realization f, the *arithmetical interpretation* $f_T(\varphi)$ of a formula $\varphi \in \mathcal{L}(\Box,\vec{c})$ is defined as usual, except that we stipulate that $f_T(c_i) = A_i$ for each $i \leq n$. The *provability logic* $\boldsymbol{PL}_{T,\vec{A}}(U)$ is defined as the set of all $\mathcal{L}(\Box,\vec{c})$-formulas that are provable in U under every realization f. We let $\boldsymbol{PL}_{T,\vec{A}}$ be $\boldsymbol{PL}_{T,\vec{A}}(T)$.

Obviously, the propositions $\vec{c}$ in $\boldsymbol{PL}_{T,\vec{A}}(U)$ really behave as constants: the logic is, in general, not closed under the rule of substitution of formulas for $\vec{c}$. However, it is closed under the substitution rule for the other propositional variables of the language.

Provability algebraic view

Describing propositional logics with constants is very close to describing *universal types* in the provability algebras. $\boldsymbol{PL}_{T,\vec{A}}$ represents the set of terms $\varphi(\vec{c},\vec{p})$ such that, in the provability algebra of T, there holds

$$\mathcal{M}_T \vDash \forall\vec{p}\,\varphi(\vec{A},\vec{p}) = \top.$$

The *universal type* of a tuple $\vec{A}$ in the provability algebra $\mathcal{M}_T$ is the set of all universal formulas, in the language of $\mathcal{M}_T$ enriched by the constants $\vec{c}$, that are true under the interpretation of $\vec{c}$ as $\vec{A}$. In particular, this set contains $\boldsymbol{PL}_{T,\vec{A}}$. The universal types can be exactly characterized in terms of provability logics as follows (compare with Theorem 97).

Let $\chi(\vec{c},\vec{p})$ be a quantifier-free formula in the language of $\mathcal{M}_T$ with distinguished parameters $\vec{c}$. Let $\tilde{\chi}$ be defined as in Theorem 97.

PROPOSITION 117. *$\mathcal{M}_T \vDash \forall\vec{p}\,\chi(\vec{A},\vec{p})$ iff $\tilde{\chi} \in \boldsymbol{PL}_{T,\vec{A}}(\mathsf{TA})$.*

We say that a universal formula $\forall\vec{p}\,\chi(\vec{c},\vec{p})$ is *realizable* in $\mathcal{M}_T$, if $\mathcal{M}_T \vDash \exists\vec{c}\forall\vec{p}\,\chi(\vec{c},\vec{p})$. Hence, effectively describing realizable universal formulas is equivalent to deciding the $\exists\forall$-fragment of $\mathcal{M}_T$. If one obtains a sufficiently effective classification of provability logics with constants, that would presumably imply a decision procedure for $\mathbf{Th}_{\exists\forall}(\mathcal{M}_T)$.

Natural logics with constants

We know the characterizations of several natural provability logics with constants. These cases mainly correspond to finite extensions for which we already know the bimodal provability logics. Thus, we know the axiomatizations (and decision procedures), in particular, for the following logics: the logic of $I\Sigma_n$ with a constant for $I\Sigma_{n+1}$, the logic of $I\Sigma_1$ with a constant for $\mathsf{Con}(\mathsf{PA})$, the logic of PRA with a constant for $I\Sigma_1$. These are representative examples of the broader classes of pairs (theory, constant) with the same logics, corresponding to Theorems 112, 114, and 116. We refrain from presenting the axiomatizations here, but refer the reader to [Beklemishev, 1994; Beklemishev, 1996].

Another interesting case has been considered by [Visser, 1992], who characterized the letterless (or, rather, the variable-free) fragment of $\boldsymbol{PL}_{\mathsf{S}_2,\mathrm{exp}}$, where S_2 is Buss' bounded arithmetic [Buss, 1986; Buss, 1998] and exp is the axiom stating the totality of the exponentiation function. This example is interesting because we do not know whether the usual provability logic $\boldsymbol{PL}_{\mathsf{S}_2}(\mathsf{S}_2)$ coincides with GL.

Translating polymodal logics

The bimodal logics $\boldsymbol{PL}_{T,U}$ and $\boldsymbol{PL}_{T,U}(V)$, where T and U are finite extensions of a given theory V, can be considered as fragments of $\boldsymbol{PL}_{V,A_0,A_1}$ formulated in the language with two additional constants c_0, c_1 for the axioms A_0 of T and A_1 of U. As in the proof of Theorem 110 one can define a translation $(\cdot)^*$ by specifying $(\Box\varphi)^* := \Box(c_0 \to \varphi^*)$ and $(\triangle\varphi)^* := \Box(c_1 \to \varphi^*)$.

PROPOSITION 118. *$\varphi \in \boldsymbol{PL}_{T,U}(V)$ iff $\varphi^* \in \boldsymbol{PL}_{V,A_0,A_1}$.*

A strong similarity between Kripke models for $\mathsf{GL}(c_0, c_1)$ and Carlson models suggests that, in fact, not too much information is being lost by going from a provability logic with constants to its bimodal fragment. Thus, the problem of classifying polymodal logics and the one of classifying provability logics with constants are very close to each other. However, technically speaking, they seem to be incomparable questions. Polymodal logic allows to speak about non-finitely axiomatizable extensions of theories, whereas the language with constants is more expressive in the finitely axiomatizable case.

Variable-free fragments

More information on bimodal provability logics can be extracted from the description of subalgebras of provability algebras. Assume for simplicity that T is a Σ_1-sound theory. Shavrukov's theorem characterizes all possible variable-free fragments of the logics $\boldsymbol{PL}_{T,\vec{A}}$ as those propositional theories in the language with constants $\vec{c}$ that are r.e. and satisfy the strong disjunction property. Thus, any propositional theory satisfying these broad conditions corresponds to some choice of sentences $\vec{A}$. (Notice that in this case we only deal with finitely generated subalgebras.)

The variable-free fragments of bimodal provability logics are, therefore, the fragments of such propositional theories obtained via the translation $(\cdot)^*$.

THEOREM 119. *Let P be a variable-free theory in the language $\mathcal{L}(\Box, \triangle)$. P is the variable-free fragment of $\boldsymbol{PL}_{T,U}(V)$ for some (finite) extensions (T, U) of a Σ_1-sound theory V iff P is r.e. and satisfies the following disjunction property: for any finite set S of formulas of the form $\Box\varphi$ or $\triangle\psi$,*

$$P \vDash \bigvee S \quad \Rightarrow \quad \exists \sigma \in S \ P \vDash \sigma.$$

This observation is due to the second author jointly with A. Visser and has not been published.

8.3 Strong provability predicates

Apart from describing the joint behavior of two 'usual' provability predicates, each of which alone being well enough understood, bimodal logic has been successfully used for the analysis of some *strong*, that is, non-r.e. concepts of provability. The notion of *n-provability* will be especially useful for us in Section 10, where we discuss applications in proof theory. Therefore, we shall slow down and present a few more details on it here.

n-Provability and n-Consistency

Let $Th_{\Pi_n}(\mathbb{N})$ denote the set of all true arithmetical Π_n-sentences. A theory T is called *n-consistent* if $T + Th_{\Pi_n}(\mathbb{N})$ is consistent. If T is elementary presented, the theory $U \equiv T + Th_{\Pi_n}(\mathbb{N})$ will generally not be r.e., but it can be presented by the Π_n-formula

$$\mathsf{Ax}_U(x) := (\mathsf{Ax}_T(x) \vee \mathsf{True}_{\Pi_n}(x)),$$

where $\mathsf{True}_{\Pi_n}(x)$ is a truth-definition for Π_n-sentences in EA. The corresponding *n-provability predicate* and *n-consistency assertion*,

$$n\text{-}\mathsf{Prov}_T(x) := \mathsf{Prov}_U(x) \quad \text{and} \quad n\text{-}\mathsf{Con}(T) := \mathsf{Con}(U),$$

will have arithmetical complexity Σ_{n+1} and Π_{n+1}, respectively.

For $n = 0$ these concepts coincide with the usual ones for T. For brevity, we write $[n]_T\varphi$ for n-$\mathsf{Prov}_T(\ulcorner\varphi\urcorner)$ and $\langle n\rangle_T\varphi$ for $\neg[n]_T\neg\varphi$ or, equivalently, n-$\mathsf{Con}(T+\varphi)$. Thus, $[n]_T\varphi$ asserts that φ is provable from the axioms of T and some true Π_n-sentences.

Many properties of n-provability and n-consistency are very similar to those of the usual provability predicate.

PROPOSITION 120 (provable Σ_{n+1}-completeness). *For any Σ_{n+1}-formula $\sigma(x_1,\dots,x_n)$ with exactly the variables $x_1,\dots,x_n$ free*

$$\mathsf{EA} \vdash \sigma(x_1,\dots,x_n) \to [n]_T\sigma(\dot{x}_1,\dots,\dot{x}_n).$$

PROPOSITION 121. *The n-provability predicate $[n]_T$ satisfies Bernays–Löb derivability conditions:*

L1. $T + Th_{\Pi_n}(\mathbb{N}) \vdash \varphi \iff \mathsf{EA} + Th_{\Pi_n}(\mathbb{N}) \vdash [n]_T\varphi$*;*

L2. $\mathsf{EA} \vdash [n]_T(\varphi\to\psi) \to ([n]_T\varphi \to [n]_T\psi)$*;*

L3. $\mathsf{EA} \vdash [n]_T\varphi \to [n]_T[n]_T\varphi$*.*

The following useful lemma shows that n-consistency assertions are equivalent to uniform reflection principles for T (see Section 4.2).

PROPOSITION 122 (Reflection). *Over* EA*,*

$$n\text{-}\mathsf{Con}(T) \equiv \mathsf{RFN}_{\Pi_{n+1}}(T).$$

Proof. Recall that $\mathsf{RFN}_{\Pi_{n+1}}(T)$ is the schema

$$\{\forall x(\Box_T\varphi(\dot{x}) \to \varphi(x)) : \varphi \in \Pi_{n+1}\}.$$

($\Rightarrow$) If $\varphi(x) \in \Pi_{n+1}$, then $\neg\varphi(x)$ implies $[n]_T\neg\varphi(\dot{x})$, by Σ_{n+1}-completeness. Therefore, $\Box_T\varphi(\dot{x})$ implies $[n]_T(\varphi(\dot{x}) \wedge \neg\varphi(\dot{x}))$, that is, $[n]_T\bot$.

($\Leftarrow$) If $[n]_T\bot$, then for some true $\pi \in \Pi_n$, $\Box_T\neg\pi$, by formalized Deduction theorem. Take $\varphi(x) := \neg\mathsf{True}_{\Pi_n}(x)$ so that

$$\mathsf{EA} \vdash \pi \leftrightarrow \mathsf{True}_{\Pi_n}(\ulcorner\pi\urcorner).$$

We have $\Box_T\varphi(\ulcorner\pi\urcorner)$ but $\neg\varphi(\ulcorner\pi\urcorner)$. ■

Japaridze logic

[Smoryński, 1985] observed that the logic of the n-provability predicate coincides with GL. The proof literally follows the one of Solovay's theorem.

Japaridze logic is the polymodal logic of n-provability predicates for all n's taken together. Consider the propositional language with the modalities [0], [1], [2], etc. Let f be an arithmetical realization. The arithmetical interpretation $f_T(\varphi)$ of a formula φ in this language under the realization f is defined as usual, except that now we require, for each $n \in \omega$, that

$$f_T([n]\psi) = n\text{-}\mathsf{Prov}_T(\ulcorner\varphi\urcorner).$$

The system GLP introduced in [Japaridze, 1988; Japaridze, 1986] is given by the following axioms and rules of inference.

Axioms: (i) Axioms of GL for each operator $[n]$;

(ii) $[m]\varphi \to [n]\varphi$, for $m \leq n$;

(iii) $\langle m\rangle\varphi \to [n]\langle m\rangle\varphi$, for $m < n$.

Rules: modus ponens, $\varphi \vdash [n]\varphi$.

THEOREM 123 (Japaridze). *For any sound theory T containing* EA *and any polymodal formula φ,*

$$\mathsf{GLP} \vdash \varphi \iff T \vdash f_T(\varphi), \quad \textit{for any realization } f.$$

Originally, G. Japaridze formulated this result for a somewhat different interpretation of modalities $[n]$. The history is as follows.

[Boolos, 1980] undertook a modal investigation of the concept of ω-*provability*, the notion dual to the Gödel's notion of ω-consistency, and observed that its logic coincides with GL. ω-provability can be described as the provability in arithmetic by one application of ω-rule, that is, provability in the theory

$$T' := T + \{\forall x\varphi(x) : \forall n\ T \vdash \varphi(\bar{n})\}.$$

[Japaridze, 1986] made a great step forward by characterizing the bimodal logic of provability and ω-provability for PA. In fact, he formulated the polymodal logic GLP with the interpretation of [1], [2], etc., as provability in PA closed under 1, 2, etc. nested applications of the ω-rule, that is, provability in PA', PA'', etc.

Although provability in T' is rather similar to 1-provability, it is not the same notion. [Smoryński, 1977a] showed that ω-consistency of T is equivalent to the statement $\mathsf{RFN}_{\Pi_3}(T + \mathsf{RFN}(T))$, which is much stronger than 1-consistency of T. In fact, by Proposition 122, 1-$\mathsf{Con}(T)$ is equivalent to $\mathsf{RFN}_{\Pi_2}(T)$. The quantifier complexity of the ω-provability predicate is Σ_3.

Later [Ignatiev, 1993a] simplified Japaridze's work and thoroughly investigated modal logical properties of GLP. He observed that Japaridze's theorem holds under more general assumptions than originally stated. In particular, GLP was proved to be arithmetically complete for the n-provability interpretation and, more generally, for the interpretation of $[n]$ as arithmetical predicates satisfying some sufficiently broad assumptions. (We refer the reader to [Ignatiev, 1993a] for an accurate formulation of these assumptions.)

Yet another interpretation of GLP was considered in [Boolos, 1993], who proved that the bimodal fragment of GLP is complete with respect to the interpretation of [1] as the Π^1_1-complete predicate of *provability under the ω-rule in second-order arithmetic.* The proof essentially followed Ignatiev's one.

Japaridze's logic is decidable and enjoys a reasonable Kripke semantics. The situation here is similar to that with the other bimodal provability logics such as $\mathsf{CSM} \oplus \mathsf{ER}$. GLP is not, per se, Kripke complete. However, it has a simple translation into a weaker logic GLP^-, obtained from GLP by replacing axioms (ii) by the weaker principle

$$[m]\varphi \to [n][m]\varphi, \quad \text{for } m \leq n.$$

GLP^- is already sound and complete with respect to a nice class of (finite) Kripke frames.

GLP enjoys the Craig interpolation property and the fixed point property [Ignatiev, 1993a]. More importantly, Ignatiev also found normal forms for letterless formulas in GLP which play a significant role in our Section 10. A very readable treatment of Japaridze's logic is given in [Boolos, 1993], so we omit any further details here. For the mentioned applications in proof theory only the soundness part of Theorem 123 will be essential. The soundness of GLP directly follows from Σ_{n+1}-completeness of $[n]_T$ (Axiom (iii)) and the derivability conditions (Axioms (i) and (ii)).

Other strong provability concepts

The extension of methods introduced in Section 10 to the proof-theoretic analysis of theories stronger than PA may require the study of yet stronger provability-like concepts.

Provability of φ in T can also be understood as the statement that φ is valid in all models of T. If T is expressive enough to formalize the notion of a model within its own language, we can look at $\Box$ in this model-theoretic way. This approach is especially useful in set theory, where one also considers various specific classes of models and the corresponding reflection principles. Set-theoretic reflection principles also play a significant role in modern proof-theoretic ordinal analysis, see [Pohlers, 1998; Rathjen, 1994; Rathjen, 1999].

By and large, this area of potential interest is still unexplored from the viewpoint of provability logic. A few first steps, however, have been taken as early as in 1975 by (guess whom?) R. Solovay. He has characterized the logics resulting from the interpretation of $\Box\varphi$ as φ *is valid in all transitive models of* ZF and φ *is valid in all universes* V_α*, for* α *inaccessible.* The proofs have appeared for the first time in [Boolos, 1993].

Both logics happen to be normal extensions of GL. The first one is axiomatized over GL by the principle

$$\mathbf{I}: \qquad \Box(\Box\varphi \to \Box\psi) \vee \Box(\Box\psi \to \boxdot\varphi).$$

This logic is characterized by the (finite) Kripke frames $(K, \prec)$ for GL that are converse *prewellorders* in the sense that

$$\forall x, y, z \in K\ (z \prec x \Rightarrow z \prec y \text{ or } y \prec x).$$

The second logic is axiomatized over GL by the *linearity* principle

$$\mathbf{J}: \qquad \Box(\Box\varphi \to \psi) \vee \Box(\boxdot\psi \to \varphi),$$

which is characterized by Kripke frames $(K, \prec)$ that are finite strict linear orders (or, more generally, converse well-orders).

We refer the reader to [Boolos, 1993] for further details.

8.4 Unusual provability concepts

Along with the bimodal study of the natural provability predicates, several researchers undertook a bimodal analysis of some unusual, or even pathological, provability concepts. Rosser and Feferman provability predicates are well-known technical tools in the study of incompleteness in arithmetic. Studies of these and similar notions by means of bimodal logic were mainly motivated by their curious, somewhat human-like, self-correcting behavior. There were also some modest technical uses, related to the properties of interpretability, which were later essentially overshadowed by the interpretability logic.

[Visser, 1989] was an influential paper that brought to life a host of these *'smart children of Peano'* and stimulated further work [Shavrukov, 1991; Shavrukov, 1994]. There were some precursors to that paper, though, most notably [Montagna, 1978; Guaspari and Solovay, 1979; Montagna, 1987b].

Genuine arithmetical completeness results in this area are rare, mostly because nearly all of the unusual provability concepts suffer from the lack of *robustness*. In other words, the modal properties of these concepts are dependent on minor details of the Gödel numbering or ordering of proofs. (Some examples to that effect are given below.) Therefore, the authors mainly concentrated on partial systems and purely syntactic uses of modality [Visser, 1989; Smoryński, 1985]. Yet, there are a few successes that are described below.

Rosser's provability predicate

Using the work [Guaspari and Solovay, 1979], V. Shavrukov [Shavrukov, 1991] found a complete axiomatization of the bimodal logic of the usual and *Rosser's provability predicate* for Peano arithmetic. We say that a sentence φ is *Rosser provable* if there is a proof of φ such that there is no proof of $\neg\varphi$ with a smaller Gödel number. Formally,

$$\Box^R\varphi := \exists y\,(\mathsf{Prf}_{\mathsf{PA}}(y, \ulcorner\varphi\urcorner) \wedge \forall z < y\, \neg\mathsf{Prf}_{\mathsf{PA}}(z, \ulcorner\neg\varphi\urcorner)).$$

Rosser provability has been invented in a classical paper [Rosser, 1936] in order to strengthen Gödel's first incompleteness theorem to arbitrary consistent theories containing PA. Externally, since PA is consistent, a sentence is provable in PA iff it is Rosser provable. However, provable properties of the Rosser provability predicate are very much different from those of the natural provability predicate. For example, Rosser's consistency of PA is a provable fact and Rosser's provability predicate is not, in general, provably closed under modus ponens.

The following principles axiomatize the joint logic of the usual $\Box$ and Rosser's $\Box^R$ provability predicates, which is called GR by V. Shavrukov.

Axioms: (i) axiom schemes of GL for $\Box$;

(ii) $\Box^R\varphi \to \Box\varphi$;

(iii) $\Box\varphi \to \Box\Box^R\varphi$;

(iv) $\Box\varphi \to (\Box\bot \vee \Box^R\varphi)$;

(v) $\Diamond\Box^R\varphi \to \Diamond\varphi$.

Rules: *modus ponens*, $\varphi/\Box\varphi$, $\Box\varphi/\varphi$.

It was already mentioned above that nonstandard concepts of provability are, as a rule, very unstable. Slight variations of, say, the Gödel numbering of proofs may result in great changes of the Rosser provability logic principles. For example, one can construct a provability predicate for PA, provably equivalent to the usual one, such that the corresponding Rosser's provability satisfies the GR-unprovable principle

$$\Box^R(\varphi \to \psi) \to (\Box^R\varphi \to \Box^R\psi).$$

Thus, GR actually axiomatizes the minimal set of principles shared by all Rosser's provability predicates satisfying some reasonable assumptions. V. Shavrukov proves an analog of the uniform arithmetical completeness theorem for GR showing that there is a particular Rosser's provability predicate whose logic is GR.

Guaspari–Solovay logic

The logic of Rosser provability can be seen as a fragment of the provability logic with *witness comparison* earlier introduced in [Guaspari and Solovay, 1979]. In their very insightful paper, Guaspari and Solovay enriched the language of GL by new connectives $\prec$ and $\preceq$ to allow formulas of the form $\Box\varphi \prec \Box\psi$ and $\Box\varphi \preceq \Box\psi$. The intended arithmetical interpretation of these formulas are statements *there is a proof of* φ *such that no proof of* ψ *has a smaller or equal (resp., smaller) Gödel number.* Thus, $\Box^R\varphi$ can be expressed by $\Box\varphi \preceq \Box\neg\varphi$.

[Guaspari and Solovay, 1979] characterized the minimal set R of principles of $\Box, \prec, \preceq$ shared by all reasonable provability predicates. The system R is extensively treated in [Smoryński, 1985], therefore we do not present any further details here. One remark, however, is in order.

The conditions on the class of provability predicates considered in [Guaspari and Solovay, 1979] are less restrictive than those from [Shavrukov, 1991]. Indeed, any proof predicate satisfies either $\Box(\top\wedge\top) \preceq \Box(\top\vee\top)$ or $\Box(\top\vee\top) \preceq \Box(\top\wedge\top)$. Neither principle, of course, belongs to R. Therefore R, unlike GR, cannot be the logic of *any single* proof predicate. Moreover, the arithmetical completeness proof of [Guaspari and Solovay, 1979] requires the use of *multi-conclusion* proof predicates, a property that is not shared by the usual proof predicate.

In this sense, the arithmetical completeness result for GR is stronger than the one we have for the richer language of R. On the other hand, if one is mainly interested in purely syntactical uses, R appears to be more convenient. It allows to formalize a number of standard arithmetical arguments involving fixed points (see [Smoryński, 1985]).

Feferman's provability predicate

In order to examine conditions necessary for the validity of Gödel's second incompleteness theorem, [Feferman, 1960] introduced another pathological provability predicate for PA. It also turned out to be a very useful technical tool in the study of interpretability.

Let $\mathsf{PA}\restriction n$ denote a sequence of finite subtheories of PA such that $\mathsf{PA} \equiv \bigcup_{n\geq 0}\mathsf{PA}\restriction n$. We say that a sentence φ is *Feferman provable*, if for some n such that $\mathsf{PA}\restriction n$ is consistent, $\mathsf{PA}\restriction n \vdash \varphi$. Feferman's provability predicate $\Box^F$ is just a formalization of this statement.

It is obvious that, externally, a sentence is Feferman provable iff it is provable in PA. This is not obvious from the point of view of PA, though, because PA easily proves its own Feferman consistency.

For most of the known technical applications of the Feferman provability the choice of a specific sequence of finite subtheories $\mathsf{PA}\restriction n$ is immaterial. However, the logic and certain results on the number of fixed points heavily depend on such a choice [Smoryński, 1989].

[Shavrukov, 1994] considered the sequence $\mathsf{PA} \upharpoonright n := I\Sigma_n$ and showed that this sequence allows for a nice axiomatization of the bimodal logic of the natural and the Feferman provability predicates. The logic turns out to be decidable. The arithmetical completeness proof in this case is based on a modification of the Solovay construction similar to the one used in interpretability logic [Berarducci, 1990; Visser, 1991; Japaridze, 1994].

Some other neglected children

Shavrukov's work on Feferman predicate was preceded by [Visser, 1995], a paper that appeared in 1987 in the form of a preprint. Apart from the development of Kripke semantics for bimodal logics, in that paper the concept of *provability in* PA *from 'non-standardly finitely many' axioms* was bimodally characterized. The resulting system can be obtained from CSM by adding the axiom schema

$$\triangle\psi \land \neg\Box\psi \to \triangle(\Box\varphi \to \varphi)$$

and, thus, is akin to $\mathsf{CSM} \oplus \mathsf{ER}$.

Another interesting proof predicate was considered by [Lindström, 1994]. Say that a sentence φ is *Parikh provable*, if it is provable in PA together with the inference rule $\Box\psi/\psi$, where $\Box$ is the usual provability in PA. Clearly, the Parikh rule is conservative over PA, so externally Parikh provability coincides with the usual one. Moreover, it is r.e. and satisfies Bernays–Löb derivability conditions. However, [Parikh, 1971] showed that this rule shortens some proofs in a non-provably recursive manner. Therefore, Parikh provability is not provably equivalent to the usual one. [Lindström, 1994] showed that the bimodal logic of the usual $\Box$ and Parikh provability $\triangle$ is axiomatized over CSM by

$$\triangle\varphi \leftrightarrow \triangle\Box\varphi.$$

Additional early results in bimodal logic, e.g., a bimodal analysis of the so-called Mostowski operator, can be found in [Smoryński, 1985].

9 PROVABILITY LOGIC IN INTUITIONISTIC ARITHMETIC

A challenging remaining problem in provability logic is the characterization of the propositional provability logic for Heyting arithmetic, HA. This problem is one of the main concerns for the Dutch school of provability logic from the end of the 70's (see, e.g., [Visser, 1981]). Indeed, the provability logic properties of intuitionistic and constructive provability turn out to be more complicated than those of the classical provability.

Although a solution of this problem, so far, has proved to be rather elusive, a significant amount of effort has been invested to this area and interesting partial results have been found there, in particular, in the recent years. Here we quickly review main developments in this fascinating field. We presuppose some familiarity with the intuitionistic logic and its Kripke models.

9.1 Intuitionistic arithmetic: background

Heyting arithmetic HA is the intuitionistic counterpart of PA. In other words, it can be axiomatized exactly as PA over the intuitionistic predicate calculus IQC. Intuitionistic versions of the other arithmetical theories can be formulated similarly. The axiomatization should, however, be chosen carefully: e.g., the least element principle intuitionistically implies the law of excluded middle. Also, we do not have prenex normal form theorem for IQC. The prenex formula classes Π_n and Σ_n, in general, are intuitionistically too restrictive. See [Burr, 2000] for an attempt to define proper intuitionistic analogues of these classes. Yet, the theories such as $i\mathsf{EA}$ and $iI\Sigma_1$ are well-behaved and roughly relate to their classical counterparts as HA to PA. Here, 'i' indicates that the underlying logic is the intuitionistic one, whereas the nonlogical axioms of these systems are the same as those in the classical case.

See [Troelstra, 1973; Troelstra and van Dalen, 1988] for standard sources on intuitionistic metamathematics.

The usual process of arithmetization of syntax is constructive and therefore can be carried out in $i\mathsf{EA}$. In particular, the provability predicate for HA or, for that matter, for any other elementary presented theory T, can be formulated as a Σ_1-formula. Moreover, this formula satisfies the usual Löb's derivability conditions within $i\mathsf{EA}$.

The definitions of provability interpretation and of provability logic of a theory w.r.t. a metatheory carry over without any change. $\boldsymbol{PL}_{\mathsf{HA}}(\mathsf{HA})$ will denote the provability logic of Heyting arithmetic that we are particularly interested in.

9.2 Some valid principles

It is not difficult to convince oneself that, once we have the derivability conditions, the proof of the fixed-point lemma, and therefore that of Löb's theorem, can be carried out in $i\mathsf{EA}$. Consequently, the logic $\boldsymbol{PL}_T(i\mathsf{EA})$, and hence also $\boldsymbol{PL}_T(\mathsf{HA})$, contains the axioms and rules of GL formulated over the intuitionistic propositional logic IPC. We denote this basic system by $i\mathsf{GL}$.

It was immediately clear that $\boldsymbol{PL}_{\mathsf{HA}}(\mathsf{HA})$ satisfies some additional principles. A number of such independent principles were found by [Visser, 1981].

EXAMPLE 124. HA is closed under the so-called *Markov's rule* (see [Troelstra, 1973]):

$$\mathsf{HA} \vdash \neg\neg\pi \Rightarrow \mathsf{HA} \vdash \pi,$$

where π is a Π_2-formula. This can be proved constructively using the so-called *Friedman–Dragalin translation.* Thus, a proof of this fact can be formalized in HA itself, therefore $\boldsymbol{PL}_{\mathsf{HA}}(\mathsf{HA})$ contains the principle

$$\Box\neg\neg\Box\varphi \to \Box\Box\varphi.$$

A more general provable form of the same principle is as follows:

$$\mathrm{Ma}: \quad \Box\neg\neg(\Box\psi \to \bigvee_{i=1}^{n} \Box\varphi_i) \to \Box(\Box\psi \to \bigvee_{i=1}^{n} \Box\varphi_i).$$

EXAMPLE 125. The *disjunction property* for HA is the statement that, whenever $\mathsf{HA} \vdash \varphi \vee \psi$, one has $\mathsf{HA} \vdash \varphi$ or $\mathsf{HA} \vdash \psi$. This can be written down as

$$\mathrm{DP}: \quad \Box(\varphi \vee \psi) \to \Box\varphi \vee \Box\psi.$$

However, [Friedman, 1975b] has shown that the proof of disjunction property cannot be formalized in HA itself. In fact, this property is equivalent over $i\mathsf{EA}$, assuming $\mathsf{Con}(\mathsf{HA})$, to $\mathsf{RFN}_{\Sigma_1}(\mathsf{HA})$. Even if one restricts the attention to the local (or sentential) disjunction property, it is not formalizable in HA.

Let ρ be the Rosser sentence for HA, that is,

$$i\mathsf{EA} \vdash \rho \leftrightarrow \exists x\, (\mathsf{Prf}_{\mathsf{HA}}(x, \ulcorner\rho\urcorner) \wedge \forall y \leq x\, \neg\mathsf{Prf}_{\mathsf{HA}}(x, \ulcorner\neg\rho\urcorner)).$$

We also let $\rho^{\perp}$ denote

$$\exists x\, (\mathsf{Prf}_{\mathsf{HA}}(x, \ulcorner\neg\rho\urcorner) \wedge \forall y < x\, \neg\mathsf{Prf}_{\mathsf{HA}}(x, \ulcorner\rho\urcorner)).$$

Then

$$\begin{array}{rcll} \mathsf{HA} \vdash \Box_{\mathsf{HA}}\Box_{\mathsf{HA}}\bot & \to & \Box_{\mathsf{HA}}(\rho \vee \rho^{\perp}) & \\ & \to & \Box_{\mathsf{HA}}\rho \vee \Box_{\mathsf{HA}}\rho^{\perp}, & \text{assuming DP} \\ & \to & \Box_{\mathsf{HA}}\bot, & \end{array}$$

which contradicts Löb's Theorem.

Hence, the formula DP does not belong to $\boldsymbol{PL}_{\mathsf{HA}}(\mathsf{HA})$, but it does belong to $\boldsymbol{PL}_{\mathsf{HA}}(\mathsf{HA} + \mathsf{RFN}_{\Sigma_1}(\mathsf{HA}))$.

EXAMPLE 126. The previous example has been repaired by D. Leivant who found a weakening of the disjunction property that was already provable in HA:

$$\mathrm{Le}: \quad \Box(\varphi \vee \psi) \to \Box(\Box\varphi \vee \psi).$$

This principle was formulated by D. Leivant in his Ph.D. thesis, for a proof of this fact see [Visser, 2002b].

9.3 Partial completeness results

As it was mentioned above, the analog of Solovay's theorem for HA is unknown. For all we know, the logic $\boldsymbol{PL}_{\mathsf{HA}}(\mathsf{HA})$ may even be Π_2-complete. However, partial arithmetical completeness results for some, rather weak, fragments of the modal logic language are known. Basically, there are three meaningful fragments of $\boldsymbol{PL}_{\mathsf{HA}}(\mathsf{HA})$ that have been characterized: the box-free fragment, the 'admissible rules' fragment and the letterless fragment.

De Jongh theorem

In contrast with the classical provability, already the characterization of the $\Box$-free fragment of $\boldsymbol{PL}_{\mathsf{HA}}(\mathsf{HA})$ constitutes an important nontrivial result known as *de Jongh's theorem* [de Jongh, 1970; Smoryński, 1973].

THEOREM 127 (de Jongh). *For any formula φ of* IPC*,*

$$\mathsf{IPC} \vdash \varphi \iff \varphi \in \boldsymbol{PL}_{\mathsf{HA}}(\mathsf{HA}).$$

In fact, D. de Jongh proved a much stronger result for the predicate intuitionistic logic IQC, not just for IPC. A number of different proofs and strengthenings of this theorem have been found since, for an overview see [Visser, 1999].

In particular, [Friedman, 1975c] obtained a result analogous to the uniform Solovay theorem. He showed that the free Heyting algebra on countably many generators is embeddable into the Lindenbaum Heyting algebra of HA. Later A. Visser [Visser, 1985; de Jongh and Visser, 1996] optimized the logical complexity of the embedding by showing that the generators can be chosen to be Σ_1-sentences. We call an arithmetical realization f a Σ_1-*realization* if $f(p) \in \Sigma_1$ for each propositional letter p.

THEOREM 128 (Friedman, Visser). *There is a Σ_1-realization f such that*

$$\mathsf{IPC} \vdash \varphi \iff \mathsf{HA} \vdash f(\varphi),$$

for any formula φ of IPC.

Analogs of de Jongh's theorem also hold for the provability logics of HA plus the *extended Church thesis* ECT_0 and some other systems [Gavrilenko, 1981; Visser, 1981].

Admissible rules

Recall that a propositional inference rule φ/ψ is admissible in a logic L, if for every substitution σ of formulas of L for propositional variables, we have

$$L \vdash \sigma(\varphi) \Rightarrow L \vdash \sigma(\psi).$$

Similarly, the rule is admissible in an arithmetical theory T if, for every realization f,

$$T \vdash f(\varphi) \Rightarrow T \vdash f(\psi).$$

The simplest example of a (nontrivial) admissible rule in IPC is the *independence of premise* rule:

$$\text{IP:} \quad \mathsf{IPC} \vdash \neg\varphi \to \psi \vee \theta \Rightarrow \mathsf{IPC} \vdash (\neg\varphi \to \psi) \vee (\neg\varphi \to \theta).$$

A well-known result obtained in [Rybakov, 1984; Rybakov, 1997] is that the property of a rule being admissible in IPC is decidable. [Visser, 1999] showed that the propositional admissible rules for HA are the same as those for IPC. (In contrast, recall that in Section 7.12 we showed that the modal admissible rules of PA are *not* the same as those of GL.)

It is clear that any admissible propositional inference rule φ/ψ in HA (equivalently, IPC) delivers a principle of the provability logic $\boldsymbol{PL}_{\mathsf{HA}}(\mathsf{TA})$ of the form

$$\Box\varphi \to \Box\psi. \tag{$*$}$$

The question is whether such principles also belong to $\boldsymbol{PL}_{\mathsf{HA}}(\mathsf{HA})$. Recently this question has been answered affirmatively; here is the story.

Although V. Rybakov proved that the set of admissible rules for IPC does not have a finite basis, A. Visser and D. de Jongh suggested an infinite (elementary) set of specific provably admissible rules and conjectured that it essentially constitutes an axiomatization of the set of all admissible rules.

Building on the work [Ghilardi, 1999], R. Iemhoff [Iemhoff, 2001b] proved the conjecture of A. Visser and D. de Jongh, thus characterizing the set of all admissible rules of IPC. From the characterization by R. Iemhoff and the results by A. Visser it also follows that all admissible rules in IPC are HA-*provably* admissible in HA. Therefore, any principle of the form $(*)$, where the rule φ/ψ is admissible in IPC, belongs to $\boldsymbol{PL}_{\mathsf{HA}}(\mathsf{HA})$ and vice versa, if a formula of the form $\Box\varphi \to \Box\psi$ where φ and ψ are box-free belongs to $\boldsymbol{PL}_{\mathsf{HA}}(\mathsf{HA})$, then the rule φ/ψ is (provably) admissible in IPC.

Letterless fragment

A. Visser [Visser, 1985; Visser, 2002b] proved that the letterless fragment of $\boldsymbol{PL}_{\mathsf{HA}}(\mathsf{HA})$ is decidable. Essentially, he proved a (weak) normal form result for letterless formulas in $\boldsymbol{PL}_{\mathsf{HA}}(\mathsf{HA})$. Define $\Box^{\infty}\bot := \top$.

THEOREM 129 (Visser). *For any letterless formula φ, one can effectively find an $\alpha \in \omega \cup \{\infty\}$ such that*

$$\mathsf{HA} \vdash \Box_{\mathsf{HA}}\varphi^{\mathsf{HA}} \leftrightarrow \Box^{\alpha}_{\mathsf{HA}}\bot.$$

For any letterless formula φ we have

$$\mathsf{HA} \vdash \varphi^{\mathsf{HA}} \iff \mathsf{HA} \vdash \Box_{\mathsf{HA}}\varphi^{\mathsf{HA}}.$$

Therefore, we can decide if $\varphi \in \boldsymbol{PL}_{\mathsf{HA}}(\mathsf{HA})$ by bringing $\Box\varphi$ to the form $\Box^{\alpha}\bot$ and checking if $\alpha = \infty$ (the formulas $\Box^{n}\bot$ for $n < \infty$ are never provable because they are false).

COROLLARY 130. *The letterless fragment of* $\boldsymbol{PL}_{\mathsf{HA}}(\mathsf{HA})$ *is decidable.*

The proof of Visser's theorem contains two essential ingredients. The first one is an algorithm of bringing φ to a formula in a special *no-nested-implications-on-the-left* (NNIL) form. The role of such formulas is best to be understood in terms of admissible rules for Σ_1-realizations, as explained below. The second ingredient is a special Gödel-style translation that we call *Beeson–Visser translation.* It will be dealt with in the section devoted to the proof of Visser's theorem. Further, we survey the results on general admissible rules for HA and IPC and the corresponding fragment of $\boldsymbol{PL}_{\mathsf{HA}}(\mathsf{HA})$ in Section 9.8.

9.4 *Admissible rules for* Σ_1*-realizations*

[Visser, 1985; Visser, 2002b] studied the provability logic of HA under the arithmetical realizations f such that $f(p)$ is a Σ_1-formula, for any propositional variable p. We call such realizations Σ_1*-realizations.* This restriction is sufficiently natural because Σ_1-sentences are 'constructive'. It also turned out to be technically useful, in particular, in the study of the letterless fragment of $\boldsymbol{PL}_{\mathsf{HA}}(\mathsf{HA})$.

The notion of Σ_1-realization is intrinsically linked with the notion of a NNIL-formula. NNIL-*formulas* are those formulas of IPC that have no nestings of implications on the left. Formally, NNIL is the minimal class of formulas containing propositional letters, $\bot$, $\top$, and closed under $\wedge$, $\vee$ and the following formation rule:

$$\varphi \text{ is implication-free and } \psi \in \text{NNIL} \Rightarrow (\varphi \to \psi) \in \text{NNIL}.$$

As usual, $\neg\varphi$ is understood as an abbreviation for $\varphi \to \bot$.

It is not difficult to verify that there are at most finitely many non-equivalent NNIL-formulas in n variables. A natural semantic characterization of NNIL-formulas was obtained by A. Visser and with a different proof by J. van Benthem (see [Visser, 1994; Visser, 2002b; Visser *et al.*, 1995]).

THEOREM 131. *Let* φ *be an* IPC*-formula. The following statements are equivalent:*

(i) *φ is equivalent to a NNIL-formula.*

(ii) *For every Kripke model* $\mathcal{K} \Vdash \varphi$ *and every subset* $M \subseteq K$*, if* $\mathcal{M}$ *is the restriction of* $\mathcal{K}$ *to* M*, then* $\mathcal{M} \Vdash \varphi$.

The next theorem is a central result on NNIL that has several applications in arithmetic. In particular, it gives a description of propositional admissible rules in HA for Σ_1-realizations. The theorem was proved by A. Visser as early as in 1985, but the journal publication has only recently appeared in [Visser, 2002b]. A proof consists of an algorithm of bringing a given formula φ to a NNIL-form, eventually decreasing the number of nested implications on the left, while preserving the admissible consequence relation.

THEOREM 132. *For every* IPC*-formula* φ *we can effectively find a NNIL-formula* $\varphi^\sharp$ *such that*

$$\varphi^\sharp \vdash \psi \iff \varphi/\psi \textit{ is an admissible rule in } \mathsf{HA} \textit{ for } \Sigma_1\textit{-realizations},$$

in other words, $\varphi^\sharp \vdash \psi$ *iff for every* Σ_1*-realization* f*,*

$$\mathsf{HA} \vdash f(\varphi) \Rightarrow \mathsf{HA} \vdash f(\psi).$$

Notice that the rule $\varphi/\varphi^\sharp$ is admissible for Σ_1-realizations and $\varphi^\sharp$ is uniquely defined up to logical equivalence.

COROLLARY 133. *Admissibility of a rule under* Σ_1*-realizations in* HA *is decidable.*

From the proof of Theorem 132 one can conclude that any admissible rule for Σ_1-realizations is also provably admissible. (See also Section 9.8 below.) So, one can infer some additional principles for the provability logic of HA. Indeed, if φ is any formula of IPC and φ° denotes the result of replacing all variables p_i occurring in φ by $\Box p_i$, then the formula

$$\Box\varphi^\circ \to \Box(\varphi^\sharp)^\circ$$

belongs to $\boldsymbol{PL}_{\mathsf{HA}}(\mathsf{HA})$.

EXAMPLE 134 ([Visser, 1981]). Using the algorithm from the proof of Theorem 132 one observes that $(\neg\neg p \to p)^\sharp$ is $p \vee \neg p$. (One could also independently conclude this using the Friedman–Dragalin translation.) Therefore, we have within HA:

$$\begin{array}{lll} \Box(\neg\neg\Box\varphi \to \Box\varphi) & \to & \Box(\neg\Box\varphi \vee \Box\varphi) \\ & \to & \Box(\Box\neg\Box\varphi \vee \Box\varphi) \quad \text{by Leivant's principle} \\ & \to & \Box(\Box\bot \vee \Box\varphi) \\ & \to & \Box\Box\varphi. \end{array}$$

So, the following is a principle of $\boldsymbol{PL}_{\mathsf{HA}}(\mathsf{HA})$:

$$\Box(\neg\neg\Box\varphi \to \Box\varphi) \to \Box\Box\varphi.$$

Now we turn to the Beeson–Visser translation.

9.5 HA^* *and Beeson–Visser translation*

In many results on the provability logic of HA another intuitionistic theory, called HA^*, plays a role. HA^* is simpler than HA in many respects. On the other hand, it is sufficiently conservative over HA, so the results obtained for HA^* can sometimes be transferred to HA.

Behind HA^* hides a rather natural translation, akin to Gödel's, that was suggested in [Beeson, 1975] and further simplified in [Visser, 1982]. For any formula φ of HA, let $\varphi^\Box$ be defined as follows.

(i) $\varphi^\Box = \varphi$, if φ is an atomic formula of HA;

(ii) $(\cdot)^\Box$ commutes with $\wedge, \vee, \exists$;

(iii) $(\varphi \to \psi)^\Box = \Box_{\mathsf{HA}}(\varphi^\Box \to \psi^\Box) \wedge (\varphi^\Box \to \psi^\Box)$;

(iv) $(\forall x\, \varphi(x))^\Box = \Box_{\mathsf{HA}}(\forall x\, \varphi^\Box(x)) \wedge \forall x\, \varphi^\Box(x)$.

To have some feeling about working of this translation we note the following property.

LEMMA 135. *For any* HA*-formula* φ*,* $\mathsf{HA} \vdash \varphi^\Box \to \Box_{\mathsf{HA}}\varphi^\Box$.

Proof. This is an easy induction on the build-up of φ.

For atomic formulas the claim is obvious. If φ is an implication or begins with a universal quantifier, the statement follows from the clauses (iii) and (iv). In all other cases, an application of the induction hypothesis is sufficient. For example, $(\varphi \vee \psi)^\Box$ implies $\varphi^\Box \vee \psi^\Box$, hence $\Box\varphi^\Box \vee \Box\psi^\Box$ and $\Box(\varphi^\Box \vee \psi^\Box)$. ■

LEMMA 136. *For any formula* $\sigma \in \Sigma_1$*,* $\mathsf{HA} \vdash \sigma \leftrightarrow \sigma^\Box$.

Proof. It is clearly sufficient to prove the claim for Δ_0-formulas σ. The nontrivial cases are when σ is an implication or begins with a (bounded) universal quantifier.

Let $\sigma = (\varphi \to \psi)$. Then $\sigma^\Box$ is equivalent to $\boxdot_{\mathsf{HA}}(\varphi^\Box \to \psi^\Box)$, and hence it implies $\varphi \to \psi$, by the induction hypothesis. Vice versa, $\varphi \to \psi$ implies $\boxdot_{\mathsf{HA}}(\varphi \to \psi)$, by provable Σ_1-completeness of HA, therefore also $\boxdot_{\mathsf{HA}}(\varphi^\Box \to \psi^\Box)$, by the induction hypothesis.

Bounded universal quantifiers are treated similarly. ■

LEMMA 137. *For any formula* φ*, if* $\mathsf{HA} \vdash \varphi$ *then* $\mathsf{HA} \vdash \varphi^\Box$.

Proof. This is straightforward for all the logical axioms and inference rules. The quantifier-free axioms of HA are preserved, by the previous lemma. The translation of the induction schema looks essentially as follows:

$$\varphi^\Box(0) \wedge \boxdot_{\mathsf{HA}}\forall x\, \boxdot_{\mathsf{HA}}\, (\varphi^\Box(x) \to \varphi^\Box(x+1)) \to \boxdot_{\mathsf{HA}}\forall x \varphi^\Box(x),$$

where we omitted some outer boxes. Now, the conclusion $\forall x \varphi^{\Box}(x)$ follows from the usual induction schema in HA. The formula $\Box_{\mathsf{HA}} \forall x \varphi^{\Box}(x)$ can be inferred from $\Box_{\mathsf{HA}} \forall x\, (\varphi^{\Box}(x) \to \varphi^{\Box}(x+1))$ and $\Box_{\mathsf{HA}} \varphi^{\Box}(0)$, where one uses Lemma 135 to obtain $\Box_{\mathsf{HA}} \varphi^{\Box}(0)$ from $\varphi^{\Box}(0)$. ■

We let HA^* be the set of all φ such that $\mathsf{HA} \vdash \varphi^{\Box}$. HA^* is obviously deductively closed and, by the previous lemma, contains HA. The corresponding provability predicate can be defined by

$$\Box_{\mathsf{HA}^*} \varphi := \Box_{\mathsf{HA}} \varphi^{\Box}.$$

LEMMA 138. HA^* *proves its own* completeness principle*:*

$$\mathsf{HA}^* \vdash \varphi \to \Box_{\mathsf{HA}^*} \varphi.$$

Proof. We prove that the translation of the completeness principle is provable in HA. By Lemma 135, $\varphi^{\Box}$ implies $\Box_{\mathsf{HA}} \varphi^{\Box}$. This formula is Σ_1, so by Lemma 136, it implies $(\Box_{\mathsf{HA}} \varphi^{\Box})^{\Box}$. ■

[Visser, 1982] showed that, under some natural assumptions, HA^* can be axiomatized over HA by its own completeness principle.

The role of the completeness principle is similar to that of Church thesis w.r.t. Kleene realizability translation. The completeness principle is classically false, therefore HA^* is not sound, but this does not make HA^* inconsistent. In fact, it is sufficiently conservative over HA. We will need the following conservation result, which is not optimal but will do for a proof of Visser's theorem.

LEMMA 139. *Let $\varphi \in NNIL$, and let f be a Σ_1-realization. Then*

$$\mathsf{HA} \vdash f(\varphi)^{\Box} \to f(\varphi).$$

Proof. Induction on the build-up of φ. The only nontrivial case is when $\varphi = (\psi \to \theta)$. Since ψ is implication-free, $f(\psi)$ is HA-equivalent to a Σ_1-formula. Then $f(\psi \to \theta)^{\Box}$ is equivalent to $\boxdot_{\mathsf{HA}}(f(\psi)^{\Box} \to f(\theta)^{\Box})$. We show that $\mathsf{HA} \vdash f(\psi) \to f(\theta)$.

Assume $f(\psi)$. By Lemma 136, we obtain $f(\psi)^{\Box}$, hence $f(\theta)^{\Box}$. By the induction hypothesis $f(\theta)^{\Box}$ implies $f(\theta)$, as required. ■

The previous lemma is formalizable in HA. Together with Lemma 137 this yields the following corollary.

COROLLARY 140. *Let $\varphi \in NNIL$, and let f be a Σ_1-realization. Then*

$$\mathsf{HA} \vdash \Box_{\mathsf{HA}^*} f(\varphi) \leftrightarrow \Box_{\mathsf{HA}} f(\varphi)^{\Box} \leftrightarrow \Box_{\mathsf{HA}} f(\varphi).$$

9.6 Proof of Visser's theorem

Here we prove Theorem 129. Let φ be a given letterless modal formula. We show by induction on φ that $\Box\varphi$ is equivalent to $\Box^{\alpha}\bot$, for a suitable α. Note that φ can be seen as a boolean combination of formulas of the form $\Box\psi_i$. So, by induction hypothesis, it is sufficient to show the statement of the theorem for any boolean combination of formulas of the form $\Box^{\alpha_i}\bot$. So, let $\varphi = \psi(\Box^{\alpha_1}\bot, \dots, \Box^{\alpha_n}\bot)$, where $\psi(p_1, \dots, p_n)$ is box-free. Notice that the arithmetical interpretations of formulas $\Box^{\alpha_i}\bot$ are Σ_1. Let f be the realization f mapping p_i to $\Box^{\alpha_i}_{\mathsf{HA}}\bot$. Applying Theorem 132 to f, we obtain

$$\begin{array}{rcll} \mathsf{HA} \vdash \Box_{\mathsf{HA}}\varphi^{\mathsf{HA}} & \leftrightarrow & \Box_{\mathsf{HA}} f(\psi^{\sharp}) & \\ & \to & \Box_{\mathsf{HA}} f(\psi^{\sharp})^{\Box}, & \text{by Corollary 140.} \end{array} \tag{3}$$

Now we prove by induction on the length of an IPC-formula θ the following lemma.

LEMMA 141. *For any* IPC*-formula* θ *there is an* α *such that*

$$\mathsf{HA} \vdash f(\theta)^{\Box} \leftrightarrow \Box^{\alpha}_{\mathsf{HA}}\bot.$$

Proof. The basis of induction is clear. Further, notice that for any α, β,

$$\mathsf{HA} \vdash \Box^{\alpha}_{\mathsf{HA}}\bot \to \Box^{\beta}_{\mathsf{HA}}\bot \iff \alpha \leq \beta.$$

This implies that formulas of the form $\Box^{\alpha}_{\mathsf{HA}}\bot$ are closed under conjunction and disjunction modulo equivalence in HA, so we only have to treat the case $\theta = (A \to B)$.

By the induction hypothesis, we may assume that $f(A)^{\Box}$ is equivalent to a formula $\Box^{\alpha}_{\mathsf{HA}}\bot$ for some α, and $\mathsf{HA} \vdash f(B)^{\Box} \leftrightarrow \Box^{\beta}_{\mathsf{HA}}\bot$, for some β. We have

$$\begin{array}{rcl} \mathsf{HA} \vdash f(A \to B)^{\Box} & \leftrightarrow & \boxdot_{\mathsf{HA}}(f(A)^{\Box} \to f(B)^{\Box}) \\ & \leftrightarrow & \boxdot_{\mathsf{HA}}(\Box^{\alpha}_{\mathsf{HA}}\bot \to \Box^{\beta}_{\mathsf{HA}}\bot). \end{array}$$

The latter formula is equivalent, by Löb's theorem, to $\Box^{\beta}_{\mathsf{HA}}\bot$ if $\beta < \alpha$, and to $\top$, otherwise. ■

Visser's theorem now follows from this lemma and (3).

9.7 Subalgebras of the Lindenbaum Heyting algebras

Another important application of HA^* is the Visser–de Jongh characterization of its subalgebras that bears consequences on the subalgebras of HA. [de Jongh and Visser, 1996] proved that positive Heyting algebras satisfying the disjunction property are precisely the algebras embeddable in the Lindenbaum algebra of HA^*. We formulate their result in terms of intuitionistic propositional theories.

A propositional theory P in the language of IPC satisfies the *disjunction property*, if $P \vdash \varphi \vee \psi$ implies $P \vdash \varphi$ or $P \vdash \psi$, for any formulas φ, ψ. As before, we say that P is *realizable* in a theory T, if there is an arithmetical realization f such that

$$P \vdash \varphi \iff T \vdash f(\varphi).$$

P is Σ_1-*realizable*, if f can be chosen to be a Σ_1-realization.

Obviously, any propositional theory realizable in HA^* satisfies the disjunction property because HA^* does so:

$$\mathsf{HA} \vdash (\varphi \vee \psi)^\Box \;\Rightarrow\; \mathsf{HA} \vdash \varphi^\Box \vee \psi^\Box \;\Rightarrow\; \mathsf{HA} \vdash \varphi^\Box \text{ or } \mathsf{HA} \vdash \psi^\Box.$$

The disjunction property turns out to be sufficient for the realizability of r.e. propositional theories in HA^*.

THEOREM 142 (de Jongh, Visser). *Any r.e. propositional theory P satisfying the disjunction property is Σ_1-realizable in* HA^*.

This theorem is analogous to Shavrukov's characterization of r.e. subalgebras of the provability algebra of PA. In fact, it is proved by an adaptation of a corresponding method of V. Shavrukov and D. Zambella, which in this situation even becomes technically simpler. However, a suitable characterization of subalgebras of the Lindenbaum Heyting algebra of HA itself, and of realizable propositional theories in HA, remains an open problem. A. Visser proved the following corollary about Σ_1-realizable theories.

COROLLARY 143. *Let P be a propositional theory in the language of* IPC. *Then P is Σ_1-realizable in* HA *iff P can be axiomatized by NNIL-formulas and has the disjunction property.*

Proof. Assume that P is Σ_1-realizable by f. It is sufficient to show that P is closed under the operation $(\cdot)^\sharp$.

If $P \vdash \varphi$, then $\mathsf{HA} \vdash f(\varphi)$. However, by Theorem 132, we have, for any Σ_1-realization g,

$$\mathsf{HA} \vdash g(\varphi) \;\Rightarrow\; \mathsf{HA} \vdash g(\varphi^\sharp).$$

It follows that $\mathsf{HA} \vdash f(\varphi^\sharp)$ and $P \vdash \varphi^\sharp$.

Suppose P is axiomatized by NNIL-formulas and has the disjunction property. Then P is realizable in HA^* by a Σ_1-realization f, that is,

$$P \vdash \varphi \iff \mathsf{HA}^* \vdash f(\varphi).$$

In particular, $\mathsf{HA}^* \vdash f(A)$, for every $A \in P$. By Corollary 140, we have

$$\mathsf{HA}^* \vdash f(A) \iff \mathsf{HA} \vdash f(A),$$

therefore, $\mathsf{HA} \vdash f(A)$. Hence, $P \vdash \varphi$ implies $\mathsf{HA} \vdash f(\varphi)$, for any formula φ.

On the other hand, $P \nvdash \varphi$ implies $\mathsf{HA}^* \nvdash f(\varphi)$, and hence $\mathsf{HA} \nvdash f(\varphi)$ because HA^* contains HA. Hence, f is a Σ_1-realization of P in HA. ■

We also remark that a Σ_1-realizable propositional theory in finitely many variables is axiomatizable by a *single* NNIL-formula because there are no more than finitely many non-equivalent NNIL-formulas in those variables.

Another corollary of the previous result is the above mentioned uniform version of de Jongh's theorem (Theorem 128) stating that the empty propositional theory is Σ_1-realizable in HA.

9.8 Admissible rules

Here we give a number of characterizations of admissible rules in IPC due to S. Ghilardi and R. Iemhoff.

Visser–Iemhoff calculus

A. Visser and D. de Jongh (unpublished) suggested an infinite series of admissible rules in IPC and conjectured that they form a basis of admissible rules. [Iemhoff, 2001b] later proved their conjecture. The form of the rules is sufficiently intricate. First, we introduce an abbreviation.

Define

$$(A)(B_1, \dots, B_n) := (A \to B_1) \vee \dots \vee (A \to B_n).$$

Visser's rule (V_n) is as follows:

$$\frac{(A \to (B \vee C)) \vee D}{(A)(E_1, \dots, E_n, B, C) \vee D,}$$

where $A = \bigwedge_{i=1}^{n} (E_i \to F_i)$.

A formula D is hanging around for purely technical reasons of generality: by the disjunction property, the rule with D is admissible iff the one without D is. Yet, since the disjunction property is not an inference rule, we have to keep D around to include these trivial variants as derived rules.

PROPOSITION 144 (de Jongh, Visser). *For each n, the rule (V_n) is admissible in* IPC.

Proof. Assume the premise is derivable and the conclusion is not. Then none of the formulas $(A \to E_i)$, $(A \to B)$ and $(A \to C)$ is derivable. Take the disjoint union of the countermodels for these formulas and attach a new root b to it. Since the premise is derivable, it is true at b, but none of the formulas B, C and E_i can be true at b. Hence, A is false at b. But then, since A is true everywhere except for b, one of the formulas E_i must be true at b. A contradiction.[16] ■

We will write $\varphi \vdash \psi$, if $\mathsf{IPC} \vdash \varphi \to \psi$, and $\varphi \vdash_{\mathsf{VI}} \psi$, if ψ is provable from φ using intuitionistic logic and the rules (V_n). (VI stands for Visser and Iemhoff.) For obvious reasons we obtain

COROLLARY 145. *$\varphi \vdash_{\mathsf{VI}} \psi$ implies that the rule φ/ψ is admissible in* IPC.

Iemhoff models

R. Iemhoff introduced an appropriate notion of Kripke model for the consequence relation $\vdash_{\mathsf{VI}}$. We call a Kripke model $\mathcal{K}$ for IPC an *Iemhoff model*, if every finite set of nodes $\{u_1, \ldots, u_n\}$ in $\mathcal{K}$ has a *tight predecessor*, that is, a node u such that

$$u \preceq u_1, \ldots, u_n \wedge \forall y \succ u\, \exists i \leq n\, (u_i \preceq y).$$

An Iemhoff model is *locally finite* if every of its generated submodels is finite.

The following important theorem [Iemhoff, 2001b] is a combination of the results by S. Ghilardi and R. Iemhoff.

THEOREM 146. *The following statements are equivalent:*

(i) A rule φ/ψ is admissible in IPC*;*

(ii) $\varphi \vdash_{\mathsf{VI}} \psi$;

(iii) ψ is valid in all (locally finite) Iemhoff-models, where φ is valid.

We omit the proof, but note that the implication (ii)⇒(i) is Corollary 145. The implication (iii)⇒(ii) was proved by R. Iemhoff using a canonical model construction. The implication (iii)⇒(i) is, essentially, a reformulation of a result of [Ghilardi, 1999].

We also mention without proof the following result from [Ghilardi, 1999] that parallels Theorem 132.

THEOREM 147 (Ghilardi). *For every formula φ of* IPC *one can effectively construct a formula φ^* such that a rule φ/ψ is admissible in* IPC *iff $\varphi^* \vdash \psi$.*

S. Ghilardi calls such a formula φ^* *projective approximation* of φ. As a corollary one obtains another proof of Rybakov's theorem.

COROLLARY 148 (Rybakov). *Admissibility of an inference rule in* IPC *is decidable.*

[16] Essentially the same argument, but now with Kripke models for HA, shows that the rules (V_n) are admissible in HA. This proof is not, as it stands, formalizable in HA, though.

Σ_1*-preservativity*

A. Visser noted that the rules (V_n) are also valid for a certain arithmetical interpretation. Say that an arithmetical formula φ Σ_1*-preserves* ψ if for every Σ_1-sentence C, $C \vdash_{\mathsf{HA}} \varphi$ implies $C \vdash_{\mathsf{HA}} \psi$. A propositional (modal) formula φ Σ_1*-preserves* ψ, if $f_{\mathsf{HA}}(\varphi)$ Σ_1-preserves $f_{\mathsf{HA}}(\psi)$, for every arithmetical realization f. We say that φ Σ_1-preserves ψ *provably in* HA if for every realization f,

$$\mathsf{HA} \vdash \text{“}f_{\mathsf{HA}}(\varphi)\ \Sigma_1\text{-preserves } f_{\mathsf{HA}}(\psi)\text{”}.$$

[Visser, 2002b] obtained the following result.

THEOREM 149. *For any* IPC*-formulas* φ, ψ *the following statements are equivalent:*

(i) $\varphi \vdash_{\mathsf{VI}} \psi$;

(ii) φ Σ_1*-preserves* ψ *provably in* HA;

(iii) φ Σ_1*-preserves* ψ;

(iv) φ/ψ *is (provably) admissible in* HA;

(v) φ/ψ *is admissible in* IPC.

Proof. The implication (i)⇒(ii) is proved by the so-called de Jongh translation, we omit the proof. The implications (ii)⇒(iii), (iii)⇒(iv) and (iv)⇒(v) are easy. The implication (v)⇒(i) follows from Theorem 146. ■

The following corollary was earlier obtained with a different proof in [Visser, 1999].

COROLLARY 150. *Admissible rules for* HA *and* IPC *are the same.*

By virtue of (ii) this theorem, in particular, characterizes the ‘admissible rules’ fragment of $\boldsymbol{PL}_{\mathsf{HA}}(\mathsf{HA})$.

COROLLARY 151. *For any box-free formulas* φ, ψ,

$$(\Box\varphi \to \Box\psi) \in \boldsymbol{PL}_{\mathsf{HA}}(\mathsf{HA}) \iff \varphi \vdash_{\mathsf{VI}} \psi.$$

Σ_1-preservativity can be understood as a modality that is similar (and classically equivalent) to the dual Π_1*-conservativity* modality of interpretability logic. The provability logic of HA can then be viewed as a fragment of the preservativity logic. Despite the more complicated language, using preservativity logic is technically advantageous in the study of the provability logic of HA, for the system allows to more naturally express certain principles that are built into $\boldsymbol{PL}_{\mathsf{HA}}(\mathsf{HA})$. Here we formulate the axioms of the preservativity logic of HA and formulate a current conjecture about $\boldsymbol{PL}_{\mathsf{HA}}(\mathsf{HA})$.

The language of the preservativity logic is obtained from that of IPC by adding a binary modality $\rhd$. $\Box\varphi$ goes as an abbreviation for $\top \rhd \varphi$. Preservativity logic is given by the following axioms and inference rules:

Axioms:

1. Tautologies of IPC
2. $\varphi \rhd \psi \wedge \psi \rhd \theta \to \varphi \rhd \theta$
3. $\theta \rhd \varphi \wedge \theta \rhd \psi \to \theta \rhd (\varphi \wedge \psi)$
4. $\varphi \rhd \theta \wedge \psi \rhd \theta \to (\varphi \vee \psi) \rhd \theta$
5. $\varphi \rhd \Box\varphi$
6. $(\Box\varphi \to \varphi) \rhd \varphi$
7. $\varphi \rhd \psi \to (\Box\theta \to \varphi) \rhd (\Box\theta \to \psi)$
8. (Visser's scheme) $(\alpha \to (\beta \vee \gamma)) \rhd (\alpha)(\varphi_1, \ldots, \varphi_n, \beta, \gamma)$, where $\alpha = \bigwedge_{i=1}^{n}(\varphi_i \to \psi_i)$ and the operation $(\cdot)(\cdots)$ is defined as follows:

$$\begin{array}{rcl}
(\alpha)(\beta_1, \ldots, \beta_n) & = & \displaystyle\bigvee_{i=1}^{n}(\alpha)(\beta_i) \\
(\alpha)(\bot) & = & \bot \\
(\alpha)(\gamma_1 \wedge \gamma_2) & = & (\alpha)(\gamma_1) \wedge (\alpha)(\gamma_2) \\
(\alpha)(\Box\gamma) & = & \Box\gamma \\
(\alpha)(\beta) & = & (\alpha \to \beta), \\
 & & \text{if } \beta \text{ is not of the form } \bot,\ \Box\gamma \text{ or } \gamma_1 \wedge \gamma_2.
\end{array}$$

Rules of inference: *modus ponens*; $\varphi \to \psi / \varphi \rhd \psi$ (*preservation rule*).

The *arithmetical interpretation* of the language of preservativity logic is defined as usual except that now $f_{\mathsf{HA}}(\varphi \rhd \psi)$ denotes the arithmetical formula expressing that $f_{\mathsf{HA}}(\varphi)$ Σ_1-preserves $f_{\mathsf{HA}}(\psi)$.

Notice that Visser's scheme is now more general than the one considered before because of the richer language we are working in. It has been shown by A. Visser that all the axioms and rules of preservativity logic are sound w.r.t. the intended preservativity interpretation. In particular, the validity of Axiom 4 follows from the disjunction property. However, unlike the disjunction property, this schema is also verifiable in HA.

From Axiom 2 one concludes

$$\varphi \rhd \psi \to (\Box\varphi \to \Box\psi).$$

It follows that Axioms 5 and 6 strengthen the transitivity axiom and Löb's axiom of provability logic, respectively.

It is open, whether the above preservativity logic is arithmetically complete. R. Iemhoff and A. Visser conjecture that it is. For one thing, we know that this system derives all the principles of $\boldsymbol{PL}_{\mathsf{HA}}(\mathsf{HA})$ known so far. In particular, Leivant's principle follows from Axioms 4 and 5, and Markov's principle is derivable from Visser's principle. Thus, the current conjecture is that $\boldsymbol{PL}_{\mathsf{HA}}(\mathsf{HA})$ is the $\Box$-fragment of the preservativity logic given by the above principles.

[Iemhoff, 2001a; Iemhoff, 2001c] developed suitable Kripke semantics for the preservativity logic. We refrain from formulating it here, but refer the interested reader to the original publications.

10 APPLICATIONS IN PROOF THEORY

The aim of this section is to present some applications of provability logic in proof theory and arithmetic. Provability logic was designed as a system to reason about formal provability. Yet, there is an obstacle: the properties of provability operators expressed by the logic $\boldsymbol{PL}_T(T)$ happen to be the same for all reasonable theories T. How can provability logic then say anything useful about a concrete formal system T? However, recently several genuine applications of provability logic in proof theory have been found. The idea is to use the provability logic for T not to investigate T itself, but rather to study some specific extensions of T. The universality of the provability logic then turns to an advantage: it allows to apply the same argument to various theories and languages of completely different power.

The plan of this section is as follows. First, we get some more background in proof theory and formal arithmetic. Necessarily, our exposition of this area is very fragmentary. We emphasize the notions of provably total computable function and program. Secondly, we present additional material on reflection principles that was instrumental in recent applications of provability logic. Finally, we present three applications [Beklemishev, 1999b; Beklemishev, 2003b; Beklemishev, 2004].

The first one is the result that the class of provably total computable functions of the fragment of PA with the induction schema restricted to Π_2-formulas without parameters coincides with the class of primitive recursive functions. We also obtain some other related results on parameter-free induction schemata.

The second application we give here is a new proof of the famous result by G. Gentzen [Gentzen, 1936]: the consistency of Peano arithmetic is provable (over EA) by transfinite induction up to the ordinal ϵ_0.

Finally, we present a simple combinatorial independent principle with a provability logic interpretation,"the Worm principle". These results are obtained using the notion of *graded provability algebra* generalizing the ordinary provability algebras studied in Chapter 7 and the polymodal provability logic introduced by G. Japaridze [Japaridze, 1986].

10.1 Fragments of arithmetic

We consider main fragments of arithmetic obtained by restricting, in one way or another, the axiom schema of induction axiomatizing PA. A common restriction is that of the arithmetical complexity of the induction formulas. Secondly, one can restrict or disallow the use of parameters in the induction schema. Thirdly, induction is sometimes applied in the form of an inference rule rather than a schema. (Those working in a Gentzen-style proof system would speak in this case about restricting the complexity of the *side formulas* of the induction rule.) For the very weak systems further kinds of restrictions of induction make sense, but we shall not consider them here.

There are alternative schemata axiomatizing PA, such as the collection schema or the pigeon-hole principle, that can also be restricted in similar ways and give rise to different families of fragments of PA.

Thus, in the theory of fragments, rather than investigating the whole continuum of possible subtheories of PA, one concentrates on the study of reasonably few "canonical" fragments. This allows, for example, for a rough analysis of proofs of mathematical statements in PA: in every such proof only specific instances of induction are used. Their quantifier complexity, as well as the presence of parameters and whether the induction is applied as a rule, can usually be easily checked. This allows, for example, to roughly estimate the rate of growth of functions involved in the proof.

Figure 1 shows the relationships between fragments of PA defined by restricted induction over EA. Here, $I\Sigma_n$ is axiomatized over EA by the induction schema for Σ_n-formulas with parameters. $I\Sigma_n$ is equivalent to $I\Pi_n$ by [Parsons, 1972]. $I\Sigma_n^-$ and $I\Pi_n^-$ denote the corresponding parameter-free schemata. $I\Sigma_n^R$ is the closure of EA under the Σ_n-induction rule:

$$\frac{\varphi(0), \quad \forall x\,(\varphi(x) \to \varphi(x+1))}{\forall x \varphi(x)}.$$

$I\Sigma_n^R$ is known to be equivalent to $I\Pi_{n+1}^R$ and to their parameter-free versions [Parsons, 1972; Beklemishev, 1998a]. $I\Sigma_1^R$ is equivalent to the *primitive recursive arithmetic* PRA, which will be discussed later.

Among the fragments between EA and $I\Sigma_1^R$ the most interesting for us will be the extension of EA by an axiom stating the totality of the *iterated exponentiation function* $\exp^{(x)}(y)$. (It is easy to see that the graph of this function is naturally Δ_0-definable.) We shall denote this extension by EA^+. EA^+ is strong enough to prove the Cut-elimination theorem for predicate logic and therefore some of its important consequences such as the

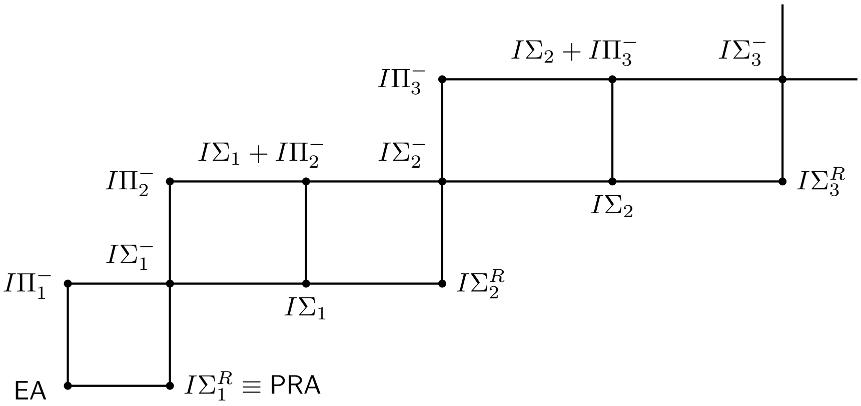

Figure 1. Restricted induction in PA

Herbrand theorem (see [Hájek and Pudlák, 1993; Wilkie and Paris, 1987]). In fact, EA^+ is equivalent to a formalized statement of the Cut-elimination theorem over EA. This is essentially due to the well-known upper and lower bounds on the length of cut-free proofs by R. Statman [Statman, 1978] and V.P. Orevkov [Orevkov, 1979].

From the proof-theoretic point of view the standard fragments of PA are interesting because their properties may differ very much from those of PA itself. The standard questions that one asks about a given fragment are, for example:

- Finite axiomatizability of the fragment;
- The optimal arithmetical complexity of its axiomatization;
- How much reflection is provable in it over a weaker fragment;
- Whether the fragment is conservative over a weaker fragment for sentences of a particular arithmetical complexity.

Later in this section we shall prove some such relationships between the fragments defined by restricted induction. Now we shall introduce one of the central notions in proof theory and formal arithmetic.

10.2 Provably total computable functions

With a system T extending EA we can associate the class $\mathcal{F}(T)$ of all functions $f : \mathbb{N}^k \to \mathbb{N}$ such that for some Σ_1-formula $\varphi(\vec{x}, y)$ there holds:

(i) $f(\vec{x}) = y \iff \mathbb{N} \vDash \varphi(\vec{x}, y)$;

(ii) $T \vdash \forall \vec{x}\, \exists y\, \varphi(\vec{x}, y)$.

Thus, the mapping $T \mapsto \mathcal{F}(T)$ sends sound theories T to classes of number-theoretic functions. The minimal class $\mathcal{F}(\mathsf{EA})$ is known to coincide with *(Kalmar) elementary functions* $\mathcal{E}$. The class $\mathcal{E}$ is defined as the closure of $0, 1, +, \cdot, 2^x$, projection functions and the characteristic function of $\leq$ by composition and bounded recursion, that is, primitive recursion with the restriction that the resulting function is bounded by some previously generated function. Thus, it is easy to see that any elementary function is bounded by some fixed iterate of 2^x.

For T containing EA, the class $\mathcal{F}(T)$ contains $\mathcal{E}$ and is closed under composition, but generally not under the bounded recursion. Also notice that $\mathcal{F}(T)$ only depends on the set of Π_2-consequences of T. Hence, if T is Π_2-conservative over U, then $\mathcal{F}(T) \subseteq \mathcal{F}(U)$.

For many natural theories T the classes $\mathcal{F}(T)$ have been characterized recursion-theoretically. For example, by a well-known result by C. Parsons [Parsons, 1970] and independently by G. Mints [Mints, 1971], $\mathcal{F}(I\Sigma_1)$ coincides with the class of primitive recursive functions. On the other hand, already W. Ackermann [Ackermann, 1940] and G. Kreisel [Kreisel, 1952] established that $\mathcal{F}(\mathsf{PA})$ coincides with the class of $<\epsilon_0$-recursive functions. Characterization of the classes $\mathcal{F}(T)$ for strong systems T is one of the main tasks of *proof-theoretic ordinal analysis*. See [Pohlers, 1998] for a survey of modern developments in this important area of proof theory.

10.3 Reflection principles and restricted induction

Recall (Section 8.3) that $[n]_T\varphi$ denotes *n-provability* of the formula φ, that is, the provability of φ from T and some true Π_n-formulas. The dual statement of *n-consistency* of φ over T is denoted $\langle n \rangle_T \varphi$.

Our applications are based on some fundamental relationships of this notion with the notion of restricted induction, on the one hand, and with that of provably total computable function, on the other. First, we discuss induction.

The following basic result was obtained in [Kreisel and Lévy, 1968]. This was preceded by a somewhat weaker result of [Mostowski, 1953] showing that PA proves consistency of any of its finite subtheories.

THEOREM 152. *Over* EA*,*

$$\mathsf{PA} \equiv \mathsf{RFN}(\mathsf{EA}) \equiv \{\langle n \rangle_{\mathsf{EA}} \top : n < \omega\}.$$

Proof. The second equivalence follows from Proposition 122. We prove the first one.

($\subseteq$) Let $\psi := \varphi(0) \wedge \forall x\, (\varphi(x) \rightarrow \varphi(x+1))$. Obviously, by induction on n, we have $\forall n\, \mathsf{EA} \vdash \psi \rightarrow \varphi(\bar{n})$. This argument is formalizable in EA, so

$$\mathsf{EA} \vdash \forall x\, \Box_{\mathsf{EA}}(\psi \rightarrow \varphi(\dot{x})).$$

Hence, $\mathsf{RFN}(\mathsf{EA})$ implies $\forall x\, (\psi \rightarrow \varphi(x))$.

($\supseteq$) Reason in PA. Assume $\mathsf{EA} \vdash \varphi(\bar{k})$. There is a cut-free proof of $\varphi(\bar{k})$ from the axioms of EA, which are all Π_1. By the subformula property all formulas occurring in the proof belong to some class Π_n, for a standard n. Using a truth-definition for Π_n-formulas prove within PA by induction on depth of a cut-free proof that all sequents in the proof are true. Conclude that $\varphi(\bar{k})$ must be true. ■

From the Unboundedness theorem (Corollary 25) following [Kreisel and Lévy, 1968] we obtain a theorem by M. Rabin [Rabin, 1961].

COROLLARY 153. PA *is not contained in any consistent theory axiomatized by a set of formulas of bounded arithmetical complexity.*

This also implies the important earlier results by C. Ryll-Nardzewski [Ryll-Nardzewski, 1953] and A. Mostowski [Mostowski, 1953] on finite non-axiomatizability of PA.

A more careful account of the complexity of formulas in the proof of Theorem 152 yields the following sharp characterization due to [Leivant, 1983] in the main direction ($\supseteq$). The ($\subseteq$) inclusion seems to have been first stated in [Ono, 1987].

THEOREM 154. *For $n \geq 1$, over EA, $I\Sigma_n \equiv \mathsf{RFN}_{\Pi_{n+2}}(\mathsf{EA}) \equiv \langle n+1 \rangle_{\mathsf{EA}} \top$.*

COROLLARY 155. *$I\Sigma_n$ is not contained in any consistent theory axiomatized by Σ_{n+2}-formulas.*

10.4 1-Consistency and provably total programs

In order to explain the relationships between the notion of n-consistency and that of provably total computable function we consider the corresponding *programs*, or *indices* of such functions. Fix some natural coding of Turing machines and a Σ_1-formula $\varphi_e(x) = y$ expressing the statement that the Turing machine coded by e on input x halts and outputs y.

DEFINITION 156. *Let T be an elementary presented theory. A number e is a T-index, if $e = \langle e_1, e_2 \rangle$ where*

- *e_1 codes a Turing machine;*
- *e_2 codes a T-proof of $\forall x \exists y\, \varphi_{\bar{e}_1}(x) = y$.*

With this indexing of provably total computable functions a universal function ψ^T is associated:

$$\psi^T_e(x) := \begin{cases} \varphi_{e_1}(x), & \text{if } e = \langle e_1, e_2 \rangle \text{ is a } T\text{-index;} \\ 0, & \text{otherwise.} \end{cases}$$

The usual diagonalization argument shows that ψ^T, as a function of arguments e and x, does not belong to $\mathcal{F}(T)$. Therefore, the statement of its totality delivers an independent principle for T.

LEMMA 157. $\mathsf{EA} \vdash \forall e, x \exists y\, \psi_e(x) = y \leftrightarrow \mathsf{RFN}_{\Pi_2}(T)$.

Proof. The totality of ψ is expressed by the formula:

$$\forall e_1, e_2, x\, (\mathsf{Prf}_T(e_2, \ulcorner \forall x \exists y \varphi_{\dot{e}_1}(x) = y \urcorner) \to \exists y \varphi_{e_1}(x) = y). \tag{4}$$

Any Π_2-sentence is equivalent to the one of the form $\forall x \exists y \varphi_{\bar{e}_1}(x) = y$ for a suitable index e_1, so it is not difficult to conclude that (4) is equivalent to $\mathsf{RFN}_{\Pi_2}(T)$. ■

If $f(\vec{x})$ is a function whose graph is definable, let $f{\downarrow}$ denote the formula $\forall \vec{x}\, \exists y f(\vec{x}) = y$. The following basic result almost immediately follows from the Herbrand theorem (cf. [Beklemishev, 1997a] for details).

PROPOSITION 158. *Suppose the graph of f is elementary. Then $g \in \mathcal{F}(\mathsf{EA} + f{\downarrow})$ iff g can be obtained from elementary functions and f by composition.*

We denote by $\mathbf{C}(f)$ the closure of $\mathcal{E} \cup \{f\}$ under composition. We can define the *jump* $\mathcal{F}(T)'$ of the class of provably total computable functions of T as $\mathbf{C}(\psi^T)$. Applying Proposition 158 to $f = \psi^T$ and using Lemma 157 we obtain[17]

COROLLARY 159. *If T is a Σ_1-sound theory, then $\mathcal{F}(\mathsf{EA} + \langle 1 \rangle_T \top) = \mathcal{F}(T)'$.*

Proposition 158 leads to an alternative 'proofs-free' indexing of functions in $\mathcal{F}(T)$ for $T = \mathsf{EA} + f{\downarrow}$. Terms in $\mathbf{C}(f)$ have a natural Gödel numbering. These numbers can be considered as codes of provably total programs. So, with this Gödel numbering we can associate another universal function $\theta^T_e(x)$ that computes the value of the term with index e on x . It has to be noted, however, that the two kinds of indexing are equivalent provably in EA^+ because the Herbrand theorem is verifiable in EA^+ and allows to extract an explicit term from a proof of totality of a computable function.

[17] Strictly speaking, the graph of ψ^T is not elementary. Instead, one considers the function $\tilde{\psi}^T$ such that $\tilde{\psi}^T_e(x)$ encodes the full protocol of the computation of $\psi_e(x)$. It is then routine to check that $\tilde{\psi}^T$ has an elementary graph and $\mathbf{C}(\psi^T) = \mathbf{C}(\tilde{\psi}^T)$.

LEMMA 160. *Suppose f has an elementary graph, is* EA*-provably nondecreasing and satisfies $f(x) > 2^x$. Then*

$$\mathsf{EA} \vdash \lambda x. f^{(x)}(x){\downarrow} \leftrightarrow \langle 1 \rangle_{\mathsf{EA}} f{\downarrow}.$$

Proof (sketch). Let $T := \mathsf{EA}{+}f{\downarrow}$. By Lemma 157, $\langle 1 \rangle_{\mathsf{EA}} f{\downarrow}$ is EA-equivalent to $\psi^T{\downarrow}$. The formula $\psi^T{\downarrow}$ implies EA^+ and hence $\theta^T{\downarrow}$. The argument is reversible, so it is sufficient to show that $\lambda x. f^{(x)}(x){\downarrow}$ is equivalent to the totality of θ^T.

Clearly, if θ is total, then for every k the function $f^{(k)}$ is also total. Indeed, $f^{(k)} \in \mathbf{C}(f)$ and the Gödel number of $f^{(k)}$ is obtained elementarily from k. Hence, $\lambda x. f^{(x)}(x)$ is total.

For the converse implication, we use the monotonicity of f. Under the given assumptions, every term $g \in \mathbf{C}(f)$ can be majorized by a fixed iterate of the function f. A similar bound also holds for the function $\tilde{g}(x)$ computing the full protocol of the computation of $g(x)$:

$$\mathsf{EA} \vdash \forall x\, (g(x) \leq f^{(k)}(x)).$$

The number k can be computed elementarily from the Gödel number of g, say, by a function $j(e)$.

Assume $\lambda x. f^{(x)}(x)$ is total. To show that, for any e and x, the value $\theta^T_e(x)$ is defined, consider the value $f^{(j(e))}(x)$. This value is smaller than $f^{(z)}(z)$, where $z := \max(j(e), x)$, hence it is defined. Therefore, the computation of $\theta^T_e(x)$ converges below this value. ∎

The classes $\mathcal{E} \subseteq \mathcal{E}' \subseteq \mathcal{E}'' \subseteq \ldots$ form the so-called *Grzegorczyk hierarchy* [Grzegorczyk, 1953]. It is well-known that the union of this hierarchy coincides with the class of primitive recursive functions (see also [Rose, 1984]). Theorem 161 below gives a stronger version of this fact.

From the previous proposition we conclude that the class $\mathcal{E}^{(n)}$ coincides with $\mathbf{C}(F_n)$, where

$$F_0(x) := 2^x + 1; \quad F_{n+1}(x) := F_n^{(x)}(x).$$

The functions F_n are all primitive recursive and their graphs are elementary definable. The extension of EA by axioms $F_n{\downarrow}$ for all $n \geq 1$ is an alternative axiomatization of the *primitive recursive arithmetic* PRA.

We define $T^n_\omega := T + \{\langle n \rangle^k_T \top : k < \omega\}$. Lemma 160 yields the following

THEOREM 161.

(i) $\mathsf{EA}^1_\omega \equiv I\Sigma^R_1 \equiv \mathsf{PRA}$*;*

(ii) $\mathcal{F}(\mathsf{EA}^1_\omega)$ *is the class of primitive recursive functions.*

Proof. (i) The totality of all F_n is immediately proved by the Σ_1-induction rule. It is also easy to see by Kreisel's trick (as in the proof of Theorem 152). Once the premise of an application of Σ_1-induction rule is proved from $\langle 1\rangle^k_{\mathsf{EA}}\top$, then the conclusion follows from $\langle 1\rangle^{k+1}_{\mathsf{EA}}\top$. Therefore, $I\Sigma_1^R \subseteq \mathsf{EA}^1_\omega$. Finally, the inclusion $\mathsf{PRA} \subseteq \mathsf{EA}^1_\omega$ follows from Lemma 160.

(ii) Every function in $\mathcal{F}(\mathsf{PRA})$ is primitive recursive, since so are all F_n. In the converse direction, it is immediately seen that the totality of any primitive recursive function is provable in $I\Sigma_1^R$. ■

10.5 Parameter-free induction

The parameter-free induction has been studied in [Kaye *et al.*, 1988; Ratajczyk, 1989; Adamovicz and Bigorajska, 1989; Beklemishev, 1997c; Beklemishev, 1999b] and other papers. $I\Sigma_n^-$ is the theory axiomatized over EA by the schema of induction

$$\varphi(0) \wedge \forall x\, (\varphi(x) \to \varphi(x+1)) \to \forall x \varphi(x),$$

where $\varphi(x)$ is a Σ_n-formula with the only free variable x. $I\Pi_n^-$ is defined similarly.

It is known that the schemata $I\Sigma_n^-$ and $I\Pi_n^-$ show a very different behavior from their counterparts $I\Sigma_n$ and $I\Pi_n$. In particular, for $n \geq 1$, $I\Sigma_n^-$ and $I\Pi_n^-$ are not finitely axiomatizable and $I\Sigma_n^-$ is strictly stronger than $I\Pi_n^-$. Here we shall obtain these results, as well as some conservation results, using graded provability algebras.

The following characterization of parameter-free induction schemata via reflection principles is found in [Beklemishev, 1997c; Beklemishev, 1999b].

THEOREM 162. *For* $n \geq 1$, *over* EA,

(i) $I\Sigma_n^- \equiv \{\pi \to \langle n\rangle_{\mathsf{EA}}\pi : \pi \in \Pi_{n+1}\}$;

(ii) $I\Pi_{n+1}^- \equiv \{\pi \to \langle n\rangle_{\mathsf{EA}}\pi : \pi \in \Pi_{n+2}\}$.

Proof. The inclusion ($\subseteq$) in both cases is proved by a trick similar to the one in the proof of Theorem 152. To prove (ii) we have to derive

$$\varphi(0) \wedge \forall x\, (\varphi(x) \to \varphi(x+1)) \to \forall x \varphi(x),$$

for each Π_{n+1} formula $\varphi(x)$ with the only free variable x. Let ψ denote the Π_{n+2}-sentence (logically equivalent to) $\varphi(0) \wedge \forall x\, (\varphi(x) \to \varphi(x+1))$. By induction on k, we obtain that for each k, $\mathsf{EA} + \psi \vdash \varphi(\bar{k})$. This fact is formalizable in EA, therefore

$$\mathsf{EA} \vdash \forall x\, \mathsf{Prov}_{\mathsf{EA}+\psi}(\ulcorner \varphi(\dot{x}) \urcorner). \qquad (5)$$

Let T denote the theory $\mathsf{EA} + \{\pi \to \langle n\rangle_{\mathsf{EA}}\pi : \pi \in \Pi_{n+2}\}$. Then we have

$$\begin{array}{rll} T + \psi & \vdash & \mathsf{RFN}_{\Pi_{n+1}}(\mathsf{EA} + \psi) \\ & \vdash & \forall x\,(\mathsf{Prov}_{\mathsf{EA}+\psi}(\ulcorner \varphi(\dot{x})\urcorner) \to \varphi(x)) \\ & \vdash & \forall x\varphi(x), \qquad \text{by (5).} \end{array}$$

It follows that $T \vdash \psi \to \forall x\varphi(x)$, as required.

The proof of the converse inclusion is more complicated and we shall omit it (see [Beklemishev, 1997c]). Notably, we will not need this part for the proof of our main conservation result (Theorem 164 below and its corollary). ■

REMARK 163. Statement (ii) of the above theorem also holds for $n = 0$, but only if $[0]$ is understood as a cut-free provability predicate. Over EA^+ there is no difference between the ordinary and the cut-free provability predicates. Thus, we may conclude that over EA^+ the schema $I\Pi_1^-$ is equivalent to $\{\pi \to \langle 0\rangle_{\mathsf{EA}}\pi : \pi \in \Pi_2\}$ which is the same as $\mathsf{Rfn}_{\Sigma_2}(\mathsf{EA})$.

Now we derive some corollaries using methods of Section 4.2 (see [Beklemishev, 1997c; Beklemishev, 1999b]).

THEOREM 164. *$I\Pi_2^-$ is a Π_2-conservative extension of* PRA.

Proof. This is, essentially, a relativized version of Theorem 30.

Assume $I\Pi_2^- \vdash \pi$ with π a Π_2-sentence. By Theorem 162 (ii) we have

$$\mathsf{EA} \vdash \textstyle\bigwedge_{i=1}^{n}(\varphi_i \to \langle 1\rangle_{\mathsf{EA}}\varphi_i) \to \pi,$$

where φ_i are Π_3-sentences. Reading in the proof of Theorem 29 everywhere $[1]_{\mathsf{EA}}$ instead of $\Box_T$ we conclude that, for some k,

$$\mathsf{EA} + \langle 1\rangle_{\mathsf{EA}}^k\top \vdash \pi.$$

However, $\mathsf{PRA} \vdash \langle 1\rangle_{\mathsf{EA}}^k\top$, for any k. ■

Since the provably total computable functions of PRA coincide with the primitive recursive functions, we obtain the following corollary.

COROLLARY 165. *$\mathcal{F}(I\Pi_2^-)$ coincides with the class of all primitive recursive functions.*

Relativization of the proof of Theorem 31 yields the following stronger conservation result.

THEOREM 166. *For $n \geq 1$, $I\Pi_{n+1}^-$ is conservative over $I\Sigma_n^-$ for boolean combinations of Σ_{n+1}-sentences.*

From Theorem 162 and the proof of Theorem 28 we also derive

THEOREM 167. *Neither $I\Sigma_n^-$ nor $I\Pi_n^-$ for $n \geq 1$ are finitely axiomatizable.*

This statement was proved by Kaye, Paris and Dimitracopoulos [Kaye *et al.*, 1988] by model-theoretic methods.

10.6 Graded provability algebras

Our next goal is, essentially, a proof-theoretic (ordinal) analysis of Peano arithmetic. Whereas at the bottom of the proof of Theorem 164 and other statements in the previous section are, essentially, the arguments formalizable in GL, this will not be enough for our further applications. Rather, we have to adopt an algebraic point of view and formulate an additional *reduction property* which is not expressible in GL, nor, for that matter, in expressively stronger Japaridze's logic GLP.

The necessary algebraic structures essentially constitute the algebraic counterpart of a sorted variant of GLP.

Let us first generalize the construction of provability algebras (Section 7) to n-provability algebras. Let T be an elementary presented theory containing EA. Since the formulas $[0]_T, [1]_T, \ldots$ satisfy Bernays–Löb derivability conditions, all of them correctly define operators acting on the Lindenbaum boolean algebra of T. Consider the enriched structure $\mathcal{M}_T^\infty = (\mathcal{B}_T, [0]_T, [1]_T, \ldots)$.

Terms of this algebra correspond to propositional *polymodal formulas* of the Japaridze logic. By Japaridze's theorem, the identities of $\mathcal{M}_T^\infty$ are exactly characterized by the system GLP.

PROPOSITION 168. *For any sound theory T containing* EA*,*

$$\mathsf{GLP} \vdash \varphi(\vec{x}) \iff \mathcal{M}_T^\infty \vDash \forall \vec{x}\, (\varphi(\vec{x}) = \top).$$

Now we enrich $\mathcal{M}_T^\infty$ by an additional *stratification* structure. Stratification is a family of distinguished subsets $P_0 \subset P_1 \subset \ldots \subseteq \mathcal{M}_T^\infty$, which correspond to the degrees of Π_1, Π_2, ... sentences. Obviously, $\bigcup_{i \geq 0} P_i = \mathcal{M}_T^\infty$. Also notice that the operator $\langle n \rangle$ maps $\mathcal{M}_T^\infty$ to P_n and P_n is closed under $\wedge$ and $\vee$. We refer to the elements of P_n as those of *sort* n. Thus, $\mathcal{M}_T^\infty$ together with the natural stratification is a many-sorted algebra. We call this algebra the *graded provability algebra of* T and abusing notations also denote it by $\mathcal{M}_T^\infty$.

The logic of the many-sorted algebra $\mathcal{M}_T^\infty$ is naturally formulated in the language with sorted propositional variables p_i^n, where the upper index n indicates that the variable ranges over sort n, that is, over Π_{n+1}-sentences. The assignment of sorts can be extended to arbitrary polymodal formulas in a natural way (all formulas of the form $\langle n \rangle \varphi$ have sort n). In addition to the identities of GLP, we have an identity expressing the principle of Σ_{n+1}-completeness:

$$\neg p_i^n \to [n] \neg p_i^n.$$

We should also keep in mind that the rule of substitution of the logic in question is restricted to respect the sorts. Then the above principle in particular allows to derive the axiom $\langle n \rangle \varphi \to [n+1] \langle n \rangle \varphi$ of GLP.

We shall call a *graded provability algebra* any many-sorted algebra $\mathcal{M}$ whose identities satisfy the logic described above. Alternatively, it can be defined as an algebra satisfying all the identities of $\mathcal{M}^\infty_{\mathsf{EA}}$ in the many-sorted language.

10.7 Reduction property

The graded provability algebra of T provides a kind of big, universal structure where all the extensions of T formulated in the arithmetical language 'live in'. Any arithmetical theory extending T is embeddable as a filter into the Lindenbaum algebra $\mathcal{B}_T$. In particular, fragments of PA above EA can be viewed as particular filters in $\mathcal{M}^\infty_{\mathsf{EA}}$. However, in order that the machinery of provability algebras could be applicable to these theories, the structure $\mathcal{M}^\infty_{\mathsf{EA}}$ has to 'see' these filters, in other words, they have to be, in some sense, nicely definable in the structure $\mathcal{M}^\infty_{\mathsf{EA}}$.

For the standard fragments of PA obtained by restricting the induction schema this was essentially observed in Section 10.3. By Theorem 154, in $\mathcal{M}^\infty_{\mathsf{EA}}$ the fragments $I\Sigma_n$ correspond to the principal filters generated by the elements $\langle n+1\rangle_{\mathsf{EA}}\top$. By Theorem 152, PA corresponds to the filter generated by $\{\langle n\rangle\top : n<\omega\}$. We also know from Theorem 162 that $I\Pi^-_{n+1}$ is the filter generated by $\{\pi\to\langle n\rangle_{\mathsf{EA}}\pi : \pi\in P_{n+1}\}$ and similarly for $I\Sigma^-_n$. By Theorem 161, PRA corresponds to $\{\langle 1\rangle^n\top : n<\omega\}$.

Stratification also allows us to express the notion of Π_{n+1}-conservative extension of theories. Let U and V be filters in $\mathcal{M}$. We write $U\subseteq_n V$ iff every $\pi\in P_n$ such that $\pi\in U$ also satisfies $\pi\in V$. $U\equiv_n V$ means $U\subseteq_n V$ and $V\subseteq_n U$. The same notation is also applied to arbitrary *sets* of elements of $\mathcal{M}$ and means the corresponding relation between *filters* generated by those sets.

The following proposition proved in [Beklemishev, 2003a; Beklemishev, 2001] is related to the so-called 'Fine Structure Theorem' of [Schmerl, 1979] and generalizes a result of [Parsons, 1972] on the conservativity of $I\Sigma_n$ over $I\Sigma^R_n$.

PROPOSITION 169 (Reduction). *Assume T is a Π_{n+2}-axiomatized theory containing* EA. *Then for all $\varphi\in\mathcal{M}^\infty_T$, the following holds in $\mathcal{M}^\infty_T$:*

$$\langle n+1\rangle_T\varphi \equiv_n \{Q^n_k(\varphi) : k<\omega\},$$

where

$$\begin{array}{rcl} Q^n_0(\varphi) & = & \langle n\rangle_T\varphi, \\ Q^n_{k+1}(\varphi) & = & \langle n\rangle_T(Q^n_k(\varphi)\wedge\varphi). \end{array}$$

Thus, the filter generated by all Π_{n+1}-consequences of an element of the form $\langle n+1\rangle_T\varphi\in\mathcal{M}^\infty_T$ of complexity Π_{n+2} can be generated by specific

Π_{n+1}-elements $Q^n_k(\varphi)$. It is important that these elements are definable by terms in the language of $\mathcal{M}^\infty_T$. Thus, Proposition 169 expresses a specific kind of definitional completeness of $\mathcal{M}^\infty_T$.

The strength of this proposition can be illustrated by the following example. Consider $n = 1$ and $\varphi = \top$ in $\mathcal{M}^\infty_{\mathsf{EA}}$. Then

$$\langle 2\rangle_{\mathsf{EA}}\top \equiv_1 \{\langle 1\rangle^k_{\mathsf{EA}}\top : k < \omega\}.$$

But $\langle 2\rangle_{\mathsf{EA}}\top$ is equivalent to $I\Sigma_1$, by Theorem 154, and $\mathsf{EA}+\{\langle 1\rangle^k_{\mathsf{EA}}\top : k < \omega\}$ is equivalent to PRA, so we obtain the following theorem due to C. Parsons [Parsons, 1970; Parsons, 1972] and G. Mints [Mints, 1971].

COROLLARY 170.

(i) $I\Sigma_1$ *is* Π_2*-conservative over* PRA*.*

(ii) $\mathcal{F}(I\Sigma_1)$ *coincides with primitive recursive functions.*

A proof of Proposition 169 can be obtained rather directly by cut-elimination in predicate logic (see [Beklemishev, 2003a]). Hence, it is formalizable in EA^+. In fact, the proposition can be viewed as an algebraic analog of the cut-elimination theorem in the sense that it reduces the formula $\langle n+1\rangle_T\varphi$ to formulas of lower arithmetical complexity. We call this property of $\mathcal{M}^\infty_T$ *the reduction property.*[18]

We conclude with a corollary of the reduction property concerning n-consistency orderings on $\mathcal{M}^\infty_T$. The *n-consistency ordering* $<_n$ on any graded provability algebra $\mathcal{M}$ is defined by

$$\psi <_n \varphi \Leftrightarrow \mathcal{M} \vDash \varphi \leq \langle n\rangle\psi.$$

Clearly, $<_n$ is transitive and irreflexive on $\mathcal{M} \setminus \{\bot\}$. Unlike the usual ordering $\leq$ of the Lindenbaum algebra of T, these orderings are certainly not dense: e.g., there are no other elements between $\top$ and $\langle 0\rangle\top$ w.r.t. the ordering $<_0$. However, recall that using Shavrukov's theorem it is possible to construct a dense linear chain in $\mathcal{M}^\infty_T$ w.r.t. $<_0$.

Define: if $\alpha = \langle n+1\rangle\varphi$, then $\alpha[\![k]\!] := Q^n_k(\varphi)$. Reduction property yields the following corollary, which tells us that the limit of the sequence $\alpha[\![k]\!]$ (in the sense of the ordering $<_n$ on $\mathcal{M}^\infty_T$) is α.

COROLLARY 171. *Assume* $\mathcal{M}^\infty_T$ *satisfies the reduction property. If* $\psi <_n \alpha$*, then* $\exists k : \psi <_n \alpha[\![k]\!]$*. Hence* $\alpha[\![0]\!] <_n \alpha[\![1]\!] <_n \ldots \longrightarrow \alpha$*.*

Proof. It is obvious that $\alpha[\![k]\!] <_n \alpha[\![k+1]\!]$ for any k in any graded provability algebra. If $T \vdash \alpha \to \langle n\rangle_T\psi$, then $\langle n\rangle_T\psi$ belongs to the filter generated by $\{\alpha[\![k]\!] : k < \omega\}$. Hence, $T \vdash \alpha[\![k]\!] \to \langle n\rangle_T\psi$. ■

[18] Not all graded provability algebras have this property: e.g., if we throw away the operation $\langle 1\rangle$ from the structure, the logic of the algebra remains the same, but the Π_1-consequences of $\langle 2\rangle\top$ cannot be expressed in terms of $\langle 0\rangle$ alone.

In the remaining part of the section we show how the notions involved can be used to give a proof-theoretic analysis of Peano arithmetic.

10.8 An algebraic view of ϵ_0

Work in the letterless fragment of GLP. Let S be the set of formulas generated from $\top$ by $\langle 0\rangle$, $\langle 1\rangle$, ... An element of S typically has the form

$$\alpha = \langle n_1\rangle\langle n_2\rangle \dots \langle n_k\rangle\top.$$

We identify such elements with words in the alphabet of natural numbers

$$\alpha = n_1 n_2 \dots n_k.$$

The empty word $\varnothing$ is identified with $\top$. Let S_n be the restriction of S to the alphabet $\{n, n+1, \dots\}$.

THEOREM 172. *$(S_n, <_n)$ is a well-founded ordering of height ϵ_0. Modulo provable equivalence in GLP this ordering is linear.*

We shall use elements of S as our codes for the ordinals below ϵ_0. Recall that GLP is elementary decidable. The reader not familiar with ϵ_0 may consider the order type of $(S, <_0)$ modulo GLP as a *definition* of ϵ_0.

The proof of this theorem is given in [Beklemishev, 2001]. Here we only formulate an easy correspondence between S and the ordinals below ϵ_0.

Define $o(0^k) = k$. If $\alpha = \alpha_1 0 \alpha_2 0 \cdots 0 \alpha_n$, where all $\alpha_i \in S_1$ and not all of them empty, then recursively define

$$o(\alpha) = \omega^{o(\alpha_n^-)} + \cdots + \omega^{o(\alpha_1^-)},$$

where β^- is obtained from $\beta \in S_1$ by replacing every letter $m+1$ by m.

We have: for all $\alpha, \beta \in S$,

$$\begin{aligned} \mathsf{GLP} \vdash \alpha \leftrightarrow \beta \quad &\text{iff} \quad o(\alpha) = o(\beta); \\ \mathsf{GLP} \vdash \beta \to \Diamond\alpha \quad &\text{iff} \quad o(\alpha) < o(\beta). \end{aligned}$$

EXAMPLE 173. $o(2101) = \omega^{o(0)} + \omega^{o(10)} = \omega + \omega^{\omega^0+\omega^1} = \omega^\omega$. Accordingly, we have

$$\mathsf{GLP} \vdash 2101 \leftrightarrow (21 \wedge 01) \leftrightarrow 21 \leftrightarrow 2.$$

Theorem 172 derives from the paper [Ignatiev, 1993a] by K. Ignatiev, who obtained normal forms for arbitrary letterless formulas of GLP. Letterless formulas constitute the *prime subalgebra* $\mathcal{P} \subset \mathcal{M}_T^\infty$.

THEOREM 174 (Ignatiev). *Suppose T is sound. On $\mathcal{P} \setminus \{\bot\}$ the ordering $<_0$ is well-founded of height ϵ_0.*

Technically, we do not need this stronger result, but it shows that ϵ_0 is an intrinsic characteristic of the algebra $\mathcal{M}_T^\infty$.

Before giving a consistency proof of PA we have to establish that modulo GLP the set S is closed under the operation $\alpha \mapsto \alpha[\![n]\!]$.

LEMMA 175. *Some derivations in* GLP*:*

(i) If $m < n$, *then* $\vdash \langle n\rangle\varphi \wedge \langle m\rangle\psi \leftrightarrow \langle n\rangle(\varphi \wedge \langle m\rangle\psi)$*;*

(ii) If $\alpha \in S_{n+1}$, *then* $\vdash \alpha \wedge n\beta \leftrightarrow \alpha n\beta$.

(iii) If $m \leq n$, *then* $\vdash nm\alpha \to m\alpha$.

Proof. Statement (i):

$$\begin{aligned}\mathsf{GLP} \vdash \langle n\rangle\varphi \wedge \langle m\rangle\psi \quad &\to \quad [n]\langle m\rangle\psi \quad \text{by Axiom (iii)}\\ &\to \quad \langle n\rangle(\varphi \wedge \langle m\rangle\psi).\end{aligned}$$

Statement (ii) follows by repeated application of (i). Statement (iii) is axiom $[m]\varphi \to [m][m]\varphi$ of GLP. ■

LEMMA 176. *If* $\alpha = \langle n+1\rangle\varphi \in S$, *then* $\exists\beta \in S$ $\mathsf{GLP} \vdash \beta \leftrightarrow \alpha[\![k]\!]$.

Proof. We argue by induction on k. For $k = 0$ we have $\alpha[\![0]\!] = \langle n\rangle\varphi \in S$.

Write $\alpha[\![k]\!] \in S$ in the form $n\gamma m\beta$, where $\gamma \in S_{n+1}$ and $m \leq n$.

$$\begin{aligned}\mathsf{GLP} \vdash \alpha[\![k+1]\!] \quad &\leftrightarrow \quad \langle n\rangle(\gamma m\beta \wedge n\gamma m\beta)\\ &\leftrightarrow \quad \langle n\rangle(\gamma(m\beta \wedge n\gamma m\beta)) \quad \text{by Lemma 175(i)}\\ &\leftrightarrow \quad \langle n\rangle(\gamma n\gamma m\beta) \quad \text{by Lemma 175(iii).}\end{aligned}$$

■

COROLLARY 177. *For any* k, $\mathsf{GLP} \vdash \alpha[\![k]\!] \leftrightarrow (n\gamma)^{k+1}m\beta$.

10.9 A consistency proof for PA

Work in $\mathcal{M}^{\infty}_{\mathsf{EA}}$. We shall denote by α^* the arithmetical interpretation of a formula (or a word) $\alpha \in S$. All the modalities will refer to the operators of $\mathcal{M}^{\infty}_{\mathsf{EA}}$. The function $(\cdot)^*$ as a mapping between Gödel numbers is elementary and thus is also representable in EA.

First of all, recall that PA is embeddable into $\mathcal{M}^{\infty}_{\mathsf{EA}}$ as a filter generated by $\{\langle n\rangle\top : n < \omega\}$, by Theorem 152. Moreover, this fact is formalizable in EA, so we obtain

$$\begin{aligned}\mathsf{EA} \vdash \forall n\Diamond(\langle n\rangle\top)^* \quad &\leftrightarrow \quad \mathsf{Con}(\mathsf{EA} + \{\langle n\rangle\top : n < \omega\})\\ &\leftrightarrow \quad \mathsf{Con}(\mathsf{PA}).\end{aligned}$$

We are going to prove $\forall\alpha \in S\, \Diamond\alpha^*$ by transfinite induction over EA^+. We claim:

$$\mathsf{EA}^+ \vdash \forall\alpha \in S\, (\forall\beta <_0 \alpha\, \Diamond\beta^* \to \Diamond\alpha^*).$$

Assume $\forall\beta <_0 \alpha\, \Diamond\beta^*$.

If $\alpha = 0\beta$, then $\Diamond\beta^*$, hence $\Diamond\Diamond\beta^*$ using $\langle 1\rangle\top$ in EA^+.

If $\alpha = \langle n+1\rangle\beta$, then $\forall k\, \Diamond\alpha[\![k]\!]^*$ because $\alpha[\![k]\!] <_0 \alpha$. By Proposition 169, (provably in EA^+)

$$\alpha^* \equiv_n \{\alpha[\![k]\!]^* : k < \omega\}.$$

Therefore, $\forall k\, \Diamond\alpha[\![k]\!]^*$ yields $\Diamond\alpha^*$. So,

$$\begin{array}{rcl} \mathsf{EA}^+ + (S,<_0)\text{-induction} & \vdash & \forall\alpha \in S\, \Diamond\alpha^* \\ & \vdash & \mathsf{Con}(\mathsf{PA}),\ \text{by (6)}. \end{array}$$

A simple inspection of the above argument shows that transfinite induction is applied once in the form of a rule for the Π_1-formula $\varphi(\alpha) := \Diamond\alpha^*$:

$$\frac{\forall\beta <_0 \alpha\ \varphi(\beta) \to \varphi(\alpha)}{\forall\alpha\varphi(\alpha)}.$$

We therefore can state a more formal version of Gentzen's famous consistency proof for PA [Gentzen, 1936; Gentzen, 1938].

THEOREM 178. $\mathsf{EA}^+ +$ $(S,<_0)$*-induction rule for* Π_1*-formulas is equivalent to*

$$\mathsf{EA}^+ + \mathsf{Con}(\mathsf{PA}) + \mathsf{Con}(\mathsf{PA} + \mathsf{Con}(\mathsf{PA})) + \ldots$$

We have shown that one application of the rule derives the consistency of PA. It is proved in a similar manner that nested applications of transfinite induction for $(S,<_0)$ derive iterated consistency assertions. A full proof of the above theorem, including the one of the converse direction that we omit here, is given in [Beklemishev, 2001].

10.10 The Worm Principle

Here we present a simple statement of combinatorial nature that is independent of Peano Arithmetic and is motivated by graded provability algebras. It asserts the termination of a certain combinatorial game reminiscent of the well-known *Hydra battle* of L. Kirby and J. Paris [Kirby and Paris, 1982].

The game deals with objects called *worms*. A worm is a finite function $f : [0,n] \to \mathbb{N}$. Worms can be specified as lists of natural numbers $w = (f(0), f(1), \ldots, f(n))$. For example, $w = 2102031$ is a worm (where we omit commas assuming all elements are <10). $f(n)$ is called the *head* of the worm. The empty worm is denoted by $\varnothing$.

Now we describe the rules of the game. Informally, the game starts with an arbitrary worm and at each step we hit the head of the worm so that

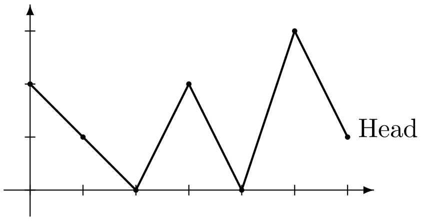

Figure 2. A Worm

it decreases by 1. In response the worm grows according to the two simple rules below. Unlike the original Hydra battle, the Worm game is fully deterministic.

We specify a function next(w, m), where $w = (f(0), f(1), \ldots, f(n))$ is a worm and m is a step of the game:

1. If $f(n) = 0$ then next$(w, m) := (f(0), \ldots, f(n-1))$. In this case the head of the worm is cut away.

2. If $f(n) > 0$ let $k := \max_{i<n} f(i) < f(n)$.

   The worm w (with the head decreased by 1) is then the concatenation of two parts, the *good*[19] part $r := (f(0), \ldots, f(k))$, and the *bad* part $s := (f(k+1), \ldots, f(n-1), f(n) - 1)$. We define

$$\text{next}(w, m) := r * \underbrace{s * s * \cdots * s}_{m+1 \text{ times}}.$$

Now let $w_0 := w$ and $w_{n+1} := \text{next}(w_n, n+1)$.

As an example consider the worm $w = 2102031$ depicted in Figure 2. At the first step we obtain $k = 4$; $r = 21020$; $s = 30$; next$(w, 1) = 210203030$. Then the game proceeds as follows:

[19] This part can also be empty.

$$\begin{aligned}
&w_0 = 2102031\\
&w_1 = 210203030\\
&w_2 = 21020303\\
&w_3 = 21020302222\\
&w_4 = 2102030222122221222122212221\\
&w_5 = 2102030(22212221222122212220)^6\\
&\dots
\end{aligned}$$

Notice that w_n is defined by primitive recursion. In fact, w_n is an elementary function of n and (the code of) w. This can be seen from the estimate

$$|w_n| \leq (n+2)! \cdot |w_0|$$

showing that the length of a worm grows only elementarily in the course of the game. Also notice that the maximal size of the elements of the worm can only decrease. This allows to write out a Δ_0-formula in three variables stating $w_n = u$.

The intended true PA-unprovable principle asserts that any initial worm is eventually reduced to nothing:

$$\textbf{E}\text{very } \textbf{W}\text{orm } \textbf{D}\text{ies} \Leftrightarrow \forall w \exists n\, w_n = \varnothing.$$

THEOREM 179. *EWD is true but unprovable in PA. In fact, EWD is equivalent to 1-Con(PA) in EA.*

Proof. First we prove $\mathsf{EA} + 1\text{-}\mathsf{Con}(\mathsf{PA}) \vdash \mathsf{EWD}$. For methodological reasons we would like to give a termination proof of the Worm game that does not refer to any ordinal assignments. Instead, we interpret worms as elements of a graded provability algebra.

We work in $\mathcal{M}^\infty_{\mathsf{EA}}$. Let $\alpha \in S$ be the converse of w, that is, the word w written in the reverse order. Then we define $w^\star := (\alpha^+)^*$, where α^+ means increasing every element of α by 1.

Thus, for example, $(103)^\star = \langle 4\rangle\langle 1\rangle\langle 2\rangle\top$.

LEMMA 180. *For any w, $\mathsf{PA} \vdash w^\star$.*

Proof. We argue by induction on $|w|$. If $w = vn$ and m is greater than any letter in w, then

$$\begin{aligned}
\mathsf{EA} \vdash v^\star \wedge \langle m+1\rangle\top \quad &\to \quad \langle m+1\rangle v^\star, \quad \text{by Lemma 175}\\
&\to \quad \langle n+1\rangle v^\star.
\end{aligned}$$

By Theorem 152 and the induction hypothesis,

$$\mathsf{PA} \vdash v^\star \wedge \langle m+1\rangle\top,$$

so $\mathsf{PA} \vdash \langle n+1\rangle v^\star$, which yields the induction step. ■

LEMMA 181. *For any w,*

$$\mathsf{EA} \vdash \forall n\, (w_n \neq \varnothing \to \Box(w_n^\star \to \langle 1\rangle w_{n+1}^\star)).$$

Proof. It is sufficient to prove

$$\forall w \neq \varnothing\ \forall n\ \mathsf{EA} \vdash w^\star \to \langle 1\rangle \mathrm{next}(w,n)^\star$$

by an argument formalizable in EA.

Let α be the converse of w. If α begins with 0, the claim is obvious. If α begins with $k+1$, the function $\mathrm{next}(w,n)$ is defined in such a way as to exactly agree with the function $\alpha[\![n]\!]$. Thus, by Corollary 177, $\alpha[\![n]\!]$ is the converse of $\mathrm{next}(w,n)$. Corollary 171 yields

$$\mathsf{GLP} \vdash \alpha \to \Diamond\alpha[\![n]\!].$$

GLP is stable under $(\cdot)^+$, so

$$\mathsf{GLP} \vdash \alpha^+ \to \langle 1\rangle\alpha[\![n]\!]^+.$$

This proves the claim, by the arithmetical soundness of GLP. ■

LEMMA 182. *For any w,* $\mathsf{EA} \vdash \langle 1\rangle w_0^\star \to \exists n\, w_n = \varnothing$.

Proof. We prove $\forall n\, w_n \neq \varnothing \to \forall n\,[1]\neg w_n^\star$ essentially using Löb's principle.

$$\begin{aligned} \mathsf{EA} \vdash \forall n\, w_n \neq \varnothing \wedge [1]\forall n[1]\neg w_n^\star &\to [1]\forall n[1]\neg w_{n+1}^\star \\ &\to \forall n[1][1]\neg w_{n+1}^\star \\ &\to \forall n[1]\neg w_n^\star,\ \text{by Lemma 181.} \end{aligned}$$

$$\begin{aligned} \mathsf{EA} \vdash [1]\forall n\, w_n \neq \varnothing &\to [1]([1]\forall n[1]\neg w_n^\star \to \forall n[1]\neg w_n^\star) \\ &\to [1]\forall n[1]\neg w_n^\star,\ \text{by Löb.} \end{aligned}$$

$$\begin{aligned} \mathsf{EA} \vdash \forall n\, w_n \neq \varnothing &\to [1]\forall n\, w_n \neq \varnothing,\ \text{by } \Sigma_2\text{-completeness} \\ &\to [1]\forall n[1]\neg w_n^\star \\ &\to \forall n[1]\neg w_n^\star \\ &\to [1]\neg w_0^\star, \end{aligned}$$

as required. ■

We conclude the first part of the proof of Theorem 179. From Lemmas 180 and 182 we obtain

$$\begin{aligned} \mathsf{PA} &\vdash \langle 1\rangle w^\star, \\ \mathsf{EA} &\vdash \langle 1\rangle w^\star \to \exists n\, w_n = \varnothing. \end{aligned}$$

Hence, provably in EA, $\forall w\ \mathsf{PA} \vdash \exists n\, w_n = \varnothing$. This proof is formalizable in EA, so $1\text{-}\mathsf{Con}(\mathsf{PA})$ implies $\forall w \exists n\, w_n = \varnothing$. ■

10.11 Independence of EWD

Let $w[\![n]\!] := \text{next}(w, n)$ and

$$w[\![n \dots n+k]\!] := w[\![n]\!][\![n+1]\!] \dots [\![n+k]\!].$$

We introduce an analogue of Hardy functions as follows. Let $h_w(n)$ be the smallest k such that

$$w[\![n \dots n+k]\!] = \varnothing.$$

We need some nice properties of h established by elementary reasoning formalizable in EA.

The following notion will be used to establish the monotonicity of h functions. Let $v \trianglelefteq u$ iff $v = u[\![0]\!][\![0]\!] \dots [\![0]\!]$. This essentially means that v is an initial segment of u except possibly for the last letter, which should be not larger than the corresponding letter in u.

LEMMA 183. *If $h_w(m)$ is defined and $u \trianglelefteq w$, then*

$$\exists k \; w[\![m \dots m+k]\!] = u.$$

Proof. The n-th letter in w can only change if all letters to the right of it are deleted. So, if w rewrites to $\varnothing$, it cannot possibly miss the u state. ■

COROLLARY 184. *If $h_w(n)$ is defined, then $\forall m \le n \, \exists k \; w[\![n \dots n+k]\!] = w[\![m]\!]$.*

LEMMA 185. *If $v \trianglelefteq u$ and $x \le y$, then $h_v(x) \le h_u(y)$.*

Proof. Repeating Corollary 184, obtain $s_0, s_1, \dots$ such that

$$\begin{array}{rcl} u[\![y \dots y+s_0]\!] & = & v[\![x]\!] \\ u[\![y \dots y+s_0+s_1]\!] & = & v[\![x]\!][\![x+1]\!] \\ & \dots & \end{array}$$

Therefore, all steps of the rewrite sequence for v occur in the rewrite sequence for u. ■

LEMMA 186. $h_{u0v}(n) = h_u(n + h_v(n) + 2) + h_v(n) + 1 > h_u(h_v(n))$.

Proof. Nothing can happen to a 0 between u and v until the v part is eliminated. So, the worm $u0v$ first rewrites to $u0$ and then to $\varnothing$. ■

COROLLARY 187. *If $w \in S_1$, then $h_{w1}(n) > h_w^{(n)}(n)$.*

Proof. Observe that $w1[\![n]\!] = w0w0 \dots w0$. ■

Now we can formulate the main lemma. As usual, $h_w\!\downarrow$ denotes the formula $\forall x \exists y\, h_w(x) = y$. Let $w^* := \alpha^*$, where α is the converse of w.

LEMMA 188. $\mathsf{EA} \vdash \forall w \in S_1\ (h_{1111w}\!\downarrow \to \langle 1\rangle w^*)$.

Using this lemma we can easily give

Proof of the independence of EWD:

$$\begin{aligned} \mathsf{EA} \vdash \forall w \exists n\, w_n = \varnothing &\to \forall w \in S_1\ h_w\!\downarrow \\ &\to \forall n\ \langle 1\rangle\langle n\rangle\top \\ &\to \text{1-}\mathsf{Con}(\mathsf{PA}). \end{aligned}$$

Here, the first implication holds because for every worm w and a number x we can find another worm $w' := w0^x$ such that $w'[\![0\dots x-1]\!] = w$. So, w' dies iff $h_w(x)$ is defined. ■

The proof relies on the reduction property of $\mathcal{M}^\infty_{\mathsf{EA}}$. More precisely, we shall use the following corollary.

COROLLARY 189. *Suppose $\alpha \in S_1$ begins with $m > 1$. Then*

$$\mathsf{EA}^+ \vdash \langle 1\rangle\alpha^* \leftrightarrow \forall n \langle 1\rangle \alpha[\![n]\!]^*.$$

Proof. In $\mathcal{M}^\infty_{\mathsf{EA}}$, by the reduction property, $\alpha^* \equiv_1 \{\alpha[\![n]\!]^* : n < \omega\}$. Therefore, formalizably in EA^+, α^* proves a false Σ_1-sentence iff $\alpha[\![n]\!]^*$ does, for some n, q.e.d. ■

We shall also essentially use Proposition 160. Notice that by Corollary 187 we have $h_{111}(x) > 2^x$, therefore the same inequality holds for the function $h_{111w}(x)$, where w is any worm.

Proof of Lemma 188. Reason in EA. By Löb, we can use as an additional assumption

$$\forall w \in S_1\ [1](h_{1111w}\!\downarrow \to \langle 1\rangle w^*).$$

If $1111w = v1$, then $h_{v1}\!\downarrow \to \lambda x.h_v^{(x)}(x)\!\downarrow$.
The function h_v is increasing, has an elementary graph and grows at least exponentially. So, if $w = \varnothing$, the claim is obvious: $h_{1111}\!\downarrow$ implies the totality of superexponentiation and hence $\langle 1\rangle\top$. If w is nonempty, we reason as follows:

$$\begin{aligned} \lambda x.h_v^{(x)}(x)\!\downarrow &\to \langle 1\rangle h_v\!\downarrow \\ &\to \langle 1\rangle\langle 1\rangle v^*, \quad \text{by the assumption} \\ &\to \langle 1\rangle w^*. \end{aligned}$$

If $1111w = v$ ends with $m > 1$, then

$$\begin{aligned} h_v\!\downarrow &\to \lambda x.h_{v[\![x]\!]}(x+1)\!\downarrow \\ &\to \forall n\ h_{v[\![n]\!]}\!\downarrow. \end{aligned}$$

Argument: Fix n. If $x \leq n$, then $h_{v[\![n]\!]}(x) \leq h_{v[\![n]\!]}(n+1)$.
If $x \geq n$, then $h_{v[\![n]\!]}(x) \leq h_{v[\![x]\!]}(x+1)$.

$$\begin{array}{rcl} \forall n\, h_{v[\![n+1]\!]}\!\downarrow & \to & \forall n\, h_{v[\![n]\!]1}\!\downarrow, \quad \text{as } v[\![n]\!]1 \trianglelefteq v[\![n+1]\!] \\ & \to & \forall n\, \langle 1\rangle h_{v[\![n]\!]}\!\downarrow, \quad \text{as before} \\ & \to & \forall n\, \langle 1\rangle\langle 1\rangle w[\![n]\!]^* \\ & \to & \langle 1\rangle w^*, \ \text{by Corollary 189.} \end{array}$$

This ends the proofs of Lemma 188 and Theorem 179. ■

Part II, Logic of Proofs

11 THE ORIGIN OF THE LOGIC OF PROOFS

Around BHK semantics. The general *BHK* idea of understanding logical connectives as operations on truth justifications proved very fruitful. [Kleene, 1945] introduced a *computational interpretation* of intuitionistic logic by considering computable functions rather than proofs to be truth justifications. It showed that a constructive (intuitionistic) logical derivation may be regarded as both a computational program and a proof of its correctness. This approach gave rise to the whole class of realizability semantics for constructive mathematical theories, as well as impressive array of applications, cf. [Troelstra, 1998]. However, neither Kleene's realizability nor its variants can be considered a *BHK*-semantics. Computational programs behave quite differently from mathematical proofs. Ordinary proofs allow for a verification, i.e. an algorithmic test of their correctness, whereas computational programs cannot have general verification algorithms. Thus, predicate

$$p \textit{ is a proof of } F$$

is **decidable**, whereas predicate

$$r \textit{ realizes } F$$

is not decidable. [Plisko, 1977] proved that the set of realizable first order formulas was not recursively axiomatizable. It is still an open problem whether the set of realizable propositional formulas is axiomatizable (recursively enumerable). Kleene himself protested against attempts to identify realizability with *BHK*.

Another paradigmatic example of computational semantics for intuitionistic logic is the *Curry-Howard isomorphism* between intuitionistic derivations and typed λ-*terms* (see for example [Girard *et al.*, 1989; Troelstra and Schwichtenberg, 1996]). From the foundational point of view, the significance of the Curry-Howard isomorphism is limited to the framework of computational semantics. It does not present a *BHK*-semantics because λ-terms are elementary prototypes of computational programs rather then proofs. If considered as proofs, they are nothing else but natural deduction derivations in the very Heyting's calculus that *BHK* is called to lay ground for. Thus, provability reading of the Curry-Howard isomorphism reduces to a trivial observation "formula F is provable in IPC iff formula F is provable in the natural deduction version of IPC," and therefore it does not give a semantics independent from the original Heyting's calculus. Surveys [Uspensky and Plisko, 1985; van Dalen, 1986; Troelstra and van Dalen, 1988] serve as good sources on the computational semantics of intuitionistic logic.

There were several *BHK*-like semantics where a role of proofs was played by abstract objects unrelated to generally accepted mathematical models of proofs (cf. monograph [Beeson, 1980] and surveys [Avigad and Feferman, 1998; Troelstra, 1998]). In particular, in [Medvedev, 1962] a "problem" is defined as an abstract nonempty finite set whose elements are called "solutions" and its subset of "actual solutions". The propositional logic of finite problems is different from IPC. It is even unknown whether the former is decidable (cf. [Uspensky and Plisko, 1985; Uspensky, 1992] for a more detailed analysis of this approach). In [Läuchli, 1970] a realizability of intuitionistic logic by abstract, not necessarily computable functionals was considered. [Kreisel, 1962a; Kreisel, 1962b; Kreisel, 1965] made an attempt to formalize the *BHK*-semantics in his *theory of constructions*, the original variant of which was inconsistent. The subsequent patch due to [Goodman, 1970] resulted in a loss of a *BHK* character of this interpretation since a "proof" of implication $A \to B$ was no longer applicable to all "proofs" of A (a comprehensive analysis of Kreisel-Goodman theory may be found in [Weinstein, 1983]).

The Kuznetsov-Muravitsky-Goldblatt-Boolos semantics ([Kuznetsov and Muravitsky, 1976; Goldblatt, 1978; Boolos, 1979b; Boolos, 1993]) for IPC was already related to a real mathematical model of provability, that of Gödel's arithmetical provability predicate. Still that semantics involved neither individual proofs nor operations on them. It is highly non-constructive since realizability in that model has a hyperarithmetical complexity, which is far from the *BHK*-semantics.

A certain summary of attempts to build a *BHK*-semantics was presented in [Weinstein, 1983]:

> "The interpretation of intuitionistic theories in terms of the notions of proof and construction ... has yet, however, failed to receive a definitive formulation."

A survey [van Dalen, 1986] says:

> "The intended interpretation of intuitionistic logic as presented by Heyting [i.e. the *BHK*-semantics, A.& B.]... so far has proved to be rather elusive. "

Kolmogorov and Gödel's approach. Kolmogorov's idea of 1932 was to develop a joint logic of propositions and "problem solutions" in the usual classical mathematics and then to interpret in this framework the intuitionistic logic without references to specific intuitionistic foundations. A short note [Kolmogorov, 1985] said:

> "The paper [Kolmogoroff, 1932] was written with the hope that the logic of solutions of problems would later become a regular part of courses on logic. It was intended to construct a unified logical apparatus dealing with objects of two types—propositions and problems."

[Kolmogoroff, 1932] operated with an informal notion of *problem solution* which left a possibility of considering different mathematical models of it. However, since 1980s the **provability** reading of Kolmogorov's "problem solution" has become widely accepted (cf. [Troelstra and van Dalen, 1988; van Dalen, 1994; Troelstra, 1998; Troelstra and Schwichtenberg, 1996]). There was a good reason for that. In mathematical logic there is one canonical model of the notions of problem and problem solution: a formula in an appropriate formal mathematical theory (for example, Peano arithmetic, or Zermelo-Frenkel set theory, etc.) and its formal proof in the given theory.

An approach to explain intuitionistic logic from the point of view of classical provability may be found in Gödel's works. As it was mentioned earlier, formalizing Brouwer's understanding of logical truth as provability, Gödel defines translation $tr(F)$ of propositional formula F in the intuitionistic language into the language of classical logic with modality $\Box$, namely (in an equivalent formulation) $tr(F)$ is obtained by $\Box$'ing every subformula of formula F. Informally speaking, the usual procedure of determining the classical truth by parsing the syntactic tree of a formula, when applied to $tr(F)$ will, for each new subformula of F, test its provability rather than truth, in agreement with Brouwer's ideas. It was established in [Gödel, 1933; McKinsey and Tarski, 1948] that such a translation provides a proper embedding of intuitionistic logic IPC into S4, i.e. into classical logic extended by the provability operator. Therefore the initial problem of defining IPC in terms of classical provability was reduced to finding an exact provability model for S4. Gödel noticed that the straightforward reading of $\Box F$ as "F is provable in a given formal theory" is inconsistent with S4 because formal provability is not reflexive. In a lecture in Vienna in 1938 Goedel revisited this problem and pointed out

> "...I supplemented the usual propositional calculus with B ("is provable in the absolute sence"), and axioms [S4 axioms are listed]. *Intuitionism is derivable from this.* A curious result, although these axioms are all extraordinary plausible: *nevertheless propositions about B are derivable from them which are surely false for every defined B ...*"

In the same lecture Gödel suggested using the format of explicit proofs *t is a proof of F* for interpreting his provability calculus S4, thus turning to *BHK*-style semantics of IPC. The definitive solution of this problem in [Artemov, 1995] was found along this way.

A compact survey of studies on provability semantics for $\mathsf{S4}$ may be found in [Artemov, 2001]. Here is a partial list of books and papers that discussed this matter: [Lemmon, 1957; Myhill, 1960; Kripke, 1963; Montague, 1963; Mints, 1974; Kuznetsov and Muravitsky, 1977; Goldblatt, 1978; Boolos, 1979b; Myhill, 1985; Shapiro, 1985b; Shapiro, 1985a; Kuznetsov and Muravitsky, 1986; Artemov, 1990; Buss, 1990; Boolos, 1993].

11.1 The logic of proofs: formal system

The results presented is Section 11 all came from [Artemov, 1995; Artemov, 2001], unless stated otherwise.

DEFINITION 190. The language of logic of proofs LP contains

- the language of classical propositional logic which includes propositional variables, truth constants $\top$, $\bot$, and boolean connectives
- proof variables $x_0, \ldots, x_n, \ldots$, proof constants $a_0, \ldots, a_n, \ldots$
- function symbols: monadic !, binary $\cdot$ and $+$
- operator symbol of the type "*term : formula*".

We will use $a, b, c, \ldots$ possibly with indices for proof constants, $x, y, z, \ldots$ for proof variables, i, j, k, l, m, n for natural numbers. Terms are defined by the grammar

$$t ::= x \mid a \mid !t \mid t_1 \cdot t_2 \mid t_1 + t_2$$

We call these terms *proof polynomials* and denote them by p,r,s.... Constants correspond to proofs of a finite fixed set of axiom schemas. We will omit "$\cdot$" whenever it is safe. We also assume that $p\cdot r\cdot s\ldots$ should be read as $(\ldots((p \cdot r) \cdot s)\ldots)$, and $p + r + s \ldots$ as $(\ldots((p + r) + s)\ldots)$.

Using t to stand for any term and S for any propositional letter, $\top$ or $\bot$, formulas are defined by the grammar

$$F ::= S \mid F_1 {\rightarrow} F_2 \mid F_1 {\wedge} F_2 \mid F_1 {\vee} F_2 \mid \neg F \mid t{:}F$$

We will use A, B, C, F, G, H for the formulas in this language, and $\Gamma, \Delta, \ldots$ for the finite sets of formulas unless otherwise explicitly stated. We will also use $\vec{x}, \vec{y}, \vec{z}, \ldots$ and $\vec{p}, \vec{r}, \vec{s}, \ldots$ for vectors of proof variables and proof polynomials respectively. If $\vec{s} = (s_1, \ldots, s_n)$ and $\Gamma = (F_1, \ldots, F_n)$, then $\vec{s}{:}\Gamma$ denotes $(s_1 : F_1, \ldots, s_n : F_n)$, $\bigvee \Gamma = F_1 \vee \ldots \vee F_n$, $\bigwedge \Gamma = F_1 \wedge \ldots \wedge F_n$. We assume the following precedences from highest to lowest: $!, \cdot, +, :, \neg, \wedge, \vee, \rightarrow$. We will use the symbol $=$ in different situations, both formal and informal. Symbol $\equiv$ denotes syntactical identity, $\ulcorner E \urcorner$ is the Gödel number of E, $|s|$ is the length of s, i.e. the total number of symbols in s. We will skip the

Gödel number symbol "$\ulcorner\urcorner$" inside proof formulas and provability formulas (such as Prf, Proof, etc.) when it is safe.

The intended semantics for $p : F$ is "p is a proof of F", which will be formalized in the next section. Note that proof systems which provide a semantics for $p : F$ are *multi-conclusion* ones, i.e. p may be a proof of several different F's.

We define the system LP_0 in the language of LP.

Axiom schemes:

A0. Finite set of axiom schemes of classical propositional logic

A1. $t{:}F \rightarrow F$ (*reflection*)

A2. $t{:}(F \rightarrow G) \rightarrow (s{:}F \rightarrow (t{\cdot}s){:}G)$ (*application*)

A3. $t{:}F \rightarrow !t{:}(t{:}F)$ (*proof checker*)

A4. $s{:}F \rightarrow (s{+}t){:}F, \quad t{:}F \rightarrow (s{+}t){:}F$ (*union*)

Rule of inference:

R1. $\varphi,\ \varphi \rightarrow \psi/\psi$ (*modus ponens*).

The system LP is LP_0 plus the rule

R2. $A \vdash c{:}A$, if A is an axiom A0 – A4, and c a proof constant (*axiom necessitation*)

A *Constant Specification (CS)* is a finite set of formulas $c_1 : A_1, \dots, c_n : A_n$ such that c_i is a constant, and A_i an axiom *A0 – A4*. *CS* is *injective* if for each constant c there is at most one formula $c{:}A \in CS$ (each constant denotes a proof of not more than one axiom). Each derivation in LP naturally generates the *CS* consisting of all formulas introduced in this derivation by the *axiom necessitation* rule. For a constant specification *CS*, by LP*(CS)* we mean LP_0 plus formulas from *CS* as additional axioms.

Atomic constant terms (*combinators*) of typed combinatory logic (cf. [Troelstra and Schwichtenberg, 1996]) may be regarded as proof constants. The combinator $\mathbf{k}^{A,B}$ of the type $A \rightarrow (B \rightarrow A)$ can be identified with a constant a specified as $a : (A \rightarrow (B \rightarrow A))$. The combinator $\mathbf{s}^{A,B,C}$ of the type $(A \rightarrow (B \rightarrow C)) \rightarrow ((A \rightarrow B) \rightarrow (A \rightarrow C))$ corresponds to a constant b such that $b : [(A \rightarrow (B \rightarrow C)) \rightarrow ((A \rightarrow B) \rightarrow (A \rightarrow C))]$. Term variables of combinatory logic may be regarded as proof variables in LP, application as operation "$\cdot$". A combinatory term t of the type F is represented in LP by an formula $t{:}F$. Typed combinatory logic $\mathsf{CL}_{\rightarrow}$ corresponds to a fragment of LP consisting of formulas of the sort $t{:}F$ where t contains no operations other than "$\cdot$" and F is a formula built from the propositional letters by "$\rightarrow$" only.

There is no restriction on the choice of a constant c in *R2* within a given derivation. In particular, *R2* allows us to introduce a formula $c\!:\!A(c)$, or to specify a constant several times as a proof of different axioms from *A0* – *A4*. One might restrict LP to injective constant specifications only without changing the ability of LP to emulate modal logic, or the functional and arithmetical completeness theorems for LP (below). On the other hand, LP allows us to choose the same constant in *R2* all the time.

Both LP_0 and LP enjoy the deduction theorem

$$\Gamma, A \vdash B \qquad \Rightarrow \qquad \Gamma \vdash A \rightarrow B,$$

and the substitution lemma: *If* $\Gamma(x,P) \vdash B(x,P)$*, then for any* t*,* F

$$\Gamma(x/t, P/F) \vdash B(x/t, P/F).$$

Obviously,

F is derivable in LP *with*
a constant specification CS iff $\mathsf{LP}(\mathit{CS}) \vdash F$ iff $\mathsf{LP}_0 \vdash \bigwedge \mathit{CS} \rightarrow F$.

LEMMA 191 (Lifting lemma). *If* $\vec{s}:\Gamma,\ \Delta \vdash_{\mathsf{LP}} F$*, then there is a proof polynomial* $t(\vec{x},\vec{y})$ *such that*

$$\vec{s}\!:\!\Gamma,\ \vec{y}\!:\!\Delta \vdash_{\mathsf{LP}} t(\vec{s},\vec{y})\!:\!F.$$

Moreover, if the constant specification CS in the original derivation is injective then the resulting constant specification is also injective and extends CS.

Proof. By induction on the derivation $\vec{s}\!:\!\Gamma, \Delta \vdash F$. If $F = s\!:\!G \in \vec{s}\!:\!\Gamma$, then put $t :=!s$ and use *A3*. If $F = D_j \in \Delta$, then put $t := y_j$. If F is an axiom *A0* – *A4*, then pick a fresh proof constant c and put $t := c$; by *R2*, $\vdash c\!:\!F$. Let F be derived by *modus ponens* from $G \rightarrow F$ and G. Then, by the induction hypothesis, there are proof polynomials $u(\vec{s},\vec{y})$ and $v(\vec{s},\vec{y})$ such that $u\!:\!(G \rightarrow F)$ and $v\!:\!G$ are both derivable from $\vec{s}\!:\!\Gamma, \vec{y}\!:\!\Delta$. By *A2*, $\vec{s}\!:\!\Gamma, \vec{y}\!:\!\Delta \vdash (u\cdot v)\!:\!F$, and we put $t := u\cdot v$. If F is derived by *R2*, then $F = c\!:\!A$ for some axiom A. Use the same *R2* followed by *A3*: $c\!:\!A \rightarrow !c\!:\!c\!:\!A$ and *modus ponens* to get $!c\!:\!F$, and put $t :=!c$. ■

It is easy to see from the proof that the lifting polynomial t is nothing but a blueprint of a given derivation of F. Thus, LP internalizes its own proofs as proof terms.

COROLLARY 192 (Internalization property for LP). *If*

$$A_1, \ldots, A_n \vdash B\ ,$$

then it is possible to construct a proof polynomial $t(x_1,\dots,x_n)$ *depending on fresh variables* $x_1,\dots,x_n$*, such that*

$$x_1{:}A_1,\dots,x_n{:}A_n \vdash t(x_1,\dots,x_n){:}B\ .$$

One might notice that the Curry-Howard isomorphism covers only a simple instance of the internalization property when all of $A_1,\dots,A_n,B$ are purely propositional formulas without proof terms.

COROLLARY 193 (Necessitation rule for LP).

$$\vdash F \quad\Rightarrow\ \vdash p{:}F \textit{ for some ground proof polynomial } p\ .$$

Example below shows how to derive in LP vs. S4. Note, that LP suffices to emulate all S4-derivations, as it will be shown in Theorem 198.

EXAMPLE 194. We first derive $\Box A\vee\Box B\to\Box(\Box A\vee\Box B)$ in S4.

1. $\Box A\to\Box A\vee\Box B$, $\Box B\to\Box A\vee\Box B$, axioms;
2. $\Box(\Box A\to\Box A\vee\Box B)$, $\Box(\Box B\to\Box A\vee\Box B)$, by necessitation, from 1;
3. $\Box A\to\Box\Box A$, $\Box B\to\Box\Box B$, axioms;
4. $\Box\Box A\to\Box(\Box A\vee\Box B)$, $\Box\Box B\to\Box(\Box A\vee\Box B)$, from 2;
5. $\Box A\to\Box(\Box A\vee\Box B)$, $\Box B\to\Box(\Box A\vee\Box B)$, from 3, 4;
6. $\Box A\vee\Box B\to\Box(\Box A\vee\Box B)$, from 5.

And here is the corresponding derivation in LP:

1. $x{:}A\to x{:}A\vee y{:}B$, $\quad y{:}B\to x{:}A\vee y{:}B$, axioms;
2. $a{:}(x{:}A\to x{:}A\vee y{:}B)$, $\quad b{:}(y{:}B\to x{:}A\vee y{:}B)$, constant specification;
3. $x{:}A\to !x{:}x{:}A$, $\quad y{:}B\to !y{:}y{:}B$, axioms;
4. $!x{:}x{:}A\to(a{\cdot}!x){:}(x{:}A\vee y{:}B)$, $\quad !y{:}y{:}B\to(b{\cdot}!y){:}(x{:}A\vee y{:}B)$, from 2;
5. $x{:}A\to(a{\cdot}!x){:}(x{:}A\vee y{:}B)$, $\quad y{:}B\to(b{\cdot}!y){:}(x{:}A\vee y{:}B)$, from 3, 4;

$5'$. $(a{\cdot}!x){:}(x{:}A\vee y{:}B)\to(a{\cdot}!x+b{\cdot}!y){:}(x{:}A\vee y{:}B)$, an axiom;
$5''$. $(b{\cdot}!y){:}(x{:}A\vee y{:}B)\to(a{\cdot}!x+b{\cdot}!y){:}(x{:}A\vee y{:}B)$, an axiom;
6. $x{:}A\to(a{\cdot}!x+b{\cdot}!y){:}(x{:}A\vee y{:}B)$, from 5, $5'$
$6'$. $y{:}B\to(a{\cdot}!x+b{\cdot}!y){:}(x{:}A\vee y{:}B)$, from 5, $5''$;
$6''$. $x{:}A\vee y{:}B\to(a{\cdot}!x+b{\cdot}!y){:}(x{:}A\vee y{:}B)$, from 6, $6'$.

The operations "·" and "!" are present in single-conclusion as well as in multi-conclusion proof systems. On the other hand, "+" is an operation for multi-conclusion proof systems only. Indeed, by *A4* we have $s{:}F\wedge t{:}G\to(s+t){:}F\wedge(s+t){:}G$, thus $s+t$ proves both F and G. Proof theoretical differences between single-conclusion and multi-conclusion proof systems are mostly cosmetic. Usual proof systems (Hilbert or Gentzen -style) may be considered as single-conclusion if one assumes that a proof derives only the end formula (sequent) of a proof tree. On the other hand, the same systems may be regarded as multi-conclusion by assuming that a proof derives all formulas assigned to the nodes of the proof tree. As we have seen from the introduction, logical identities of single-conclusion proofs alone are not compatible with normal modal logics. Hence, provability considered from the modal point of view corresponds to possibly multi-conclusion proofs.

Note, that individual operators "$t{:}(\ \ )$" in LP are not normal modalities since they do not satisfy the property $t:(P \to Q) \to (t:P \to t:Q)$. This makes LP essentially different from polymodal logics, e.g. the dynamic logic of programs ([Kozen and Tiuryn, 1990]), where the modality is upgraded by some additional features. Rather the modality in the logic of proofs has been decomposed into a family of proof polynomials.

11.2 Realization of modal and intuitionistic logics by proof polynomials

The formulation of LP and above Example 194 left an impression that LP was something like an explicit version of $\mathsf{S4}$. The main idea of the logic of proofs project was based on the observation that proof polynomials apparently denoted classical proof objects (cf. Subsection 11.3) and a hope that LP indeed was capable of realizing derivations in $\mathsf{S4}$ by recovering proof polynomials for every occurrence of modality. It would immediately deliver a realizability-style provability semantics for Gödel's provability calculus $\mathsf{S4}$ and hence a *BHK*-style semantics of proofs for intuitionistic propositional calculus IPC. Both facts have been first established in [Artemov, 1995].

The inverse operation to realization of modalities of proof polynomials is the forgetful projection of LP-formulas to the usual modal formulas obtained by replacing all $t{:}X$'s by $\Box X$'s. It is easy to see that the forgetful projection of LP is $\mathsf{S4}$-compliant. Let F^o be the forgetful projection of F. By a straightforward induction on a derivation in LP one could show that

LEMMA 195. *If* $\mathsf{LP} \vdash F$, *then* $\mathsf{S4} \vdash F^o$.

The goal of the current subsection is to establish the converse, namely that LP suffices to realize any theorem of $\mathsf{S4}$.

DEFINITION 196. By an LP-*realization* of a modal formula F we mean an assignment of proof polynomials to all occurrences of the modality in F along with a constant specification of all constants occurring in those proof polynomials. By F^r we understand the image of F under a realization r.

Positive and negative occurrences of modality in a formula and a sequent are defined in the conventional way. Since we will be using sequent calculus language in the realization algorithm below, we recall the polarity definition for a modal formula F within a given sequent. Namely, (1) the indicated occurrence of $\Box$ in $\Box F$ is positive; (2) any occurrence of $\Box$ from F in $G{\to}F$, $G{\wedge}F$, $F{\wedge}G$, $G{\vee}F$, $F{\vee}G$, $\Box F$ and $\Gamma \Rightarrow \Delta, F$ has the same polarity as the corresponding occurrence of $\Box$ in F; (3) any occurrence of $\Box$ from F in $\neg F$,

$F \rightarrow G$ and $F, \Gamma \Rightarrow \Delta$ has a polarity opposite to that of the corresponding occurrence of $\Box$ in F.

In a provability context $\Box F$ is intuitively understood as "*there exists a proof x of F*". After an informal skolemization, i.e. replacing quantifiers by functions, all negative occurrences of $\Box$ produce arguments of Skolem functions, whereas positive ones give functions of those arguments. For example, $\Box A \rightarrow \Box B$ should be read informally as

$$\exists x \text{ ``}x \text{ is a proof of } A\text{''} \rightarrow \exists y \text{ ``}y \text{ is a proof of } B\text{''},$$

with the Skolem form

$$\text{``}x \text{ is a proof of } A\text{''} \rightarrow \text{``}f(x) \text{ is a proof of } B\text{''}.$$

The following definition captures this feature.

DEFINITION 197. A realization r is called *normal* if all negative occurrences of $\Box$ are realized by proof variables and the corresponding constant specification is injective.

THEOREM 198. *Given a derivation* $\mathsf{S4} \vdash F$ *one could recover a normal realization* r *such that* $\mathsf{LP} \vdash F^r$

Proof. Consider a cut-free sequent formulation of $\mathsf{S4}$ (cf. [Avron, 1984], [Mints, 1974]), with sequents $\Gamma \Rightarrow \Delta$, where Γ and Δ are finite multisets of modal formulas. Without loss of generality we may assume that axioms are sequents of the form $S \Rightarrow S$, where S is a propositional letter, and the sequent $\bot \Rightarrow$. Along with the usual structural rules (weakening, contraction, cut) and rules introducing boolean connectives there are also two proper modal rules:

$$\frac{A, \Gamma \Rightarrow \Delta}{\Box A, \Gamma \Rightarrow \Delta}\,(\Box\Rightarrow) \qquad \text{and} \qquad \frac{\Box\Gamma \Rightarrow A}{\Box\Gamma \Rightarrow \Box A}\,(\Rightarrow\Box)$$

$(\Box\{A_1, \ldots, A_n\} = \{\Box A_1, \ldots, \Box A_n\})$.

Given $\mathsf{S4} \vdash F$ one could find a cut-free derivation $\mathcal{T}$ of a sequent $\Rightarrow F$. It suffices now to construct a normal realization r with an injective constant specification CS such that $\mathsf{LP}(CS) \vdash \bigwedge \Gamma^r \rightarrow \bigvee \Delta^r$ for any sequent $\Gamma \Rightarrow \Delta$ occurring in $\mathcal{T}$. We will also speak about a sequent $\Gamma \Rightarrow \Delta$ being derivable in LP meaning $\mathsf{LP} \vdash \bigwedge \Gamma \rightarrow \bigvee \Delta$, or, equivalently, $\mathsf{LP}^G \vdash \Gamma \Rightarrow \Delta$. Note that all $\Box$'s introduced by $(\Rightarrow\Box)$ are positive, and all negative $\Box$'s are introduced by $(\Box\Rightarrow)$ or by weakening.

In each rule in $\mathcal{T}$ every occurrence of $\Box$ in the conclusion sequent of the rule has one, two or none predecessors in premise sequents: two, if this occurrence is in a contraction formula; none, if this $\Box$ was just introduced by a modal rule or a weakening; one in all other cases. We call these occurrences of $\Box$ *related* and extend this relationship by transitivity. Hence, all occurrences of $\Box$ in $\mathcal{T}$ are naturally split into disjoint *families* of related ones. Since cut-free derivations in $\mathsf{S4}$ respect polarities, in any given family either all $\Box$'s are negative or all are positive. We call a family *essential* if it contains at least one instance of the $(\Rightarrow\Box)$ rule. Clearly, essential families are all positive.

Now the desired r will be constructed by stages 1 – 3 described below. We reserve a large enough set of proof variables as *provisional variables.*

Stage 1. For each negative family $\Box B$ or nonessential positive family we pick a fresh proof variable x and replace all occurrences of $\Box B$ by "$x{:}B$".

Stage 2. Pick an essential family f, enumerate all the occurrences of rules $(\Rightarrow\Box)$ which introduce boxes of this family. Let n_f be the total number of such rules for f. Replace all boxes of f by the polynomial

$$v_1 + \ldots + v_{n_f},$$

where v_i's are fresh provisional variables. The resulting tree $\mathcal{T}'$ is labelled by LP-formulas, since all $\Box$'s have been replaced by proof polynomials.

Stage 3. Run a process going from the leaves of the tree to its root which will update a constant specification CS and replace the provisional variables by proof polynomials of the usual variables from stage (1) and constants from CS as follows. By induction on the depth of a node in $\mathcal{T}'$ we establish that after the process passes a node the sequent assigned to this node becomes derivable in $\mathsf{LP}(CS)$ for a current $\mathcal{CS}$.

At the initial moment CS is empty. The axioms $S \Rightarrow S$ and $\bot \Rightarrow$ are derivable in LP_0. For every rule other than $(\Rightarrow\Box)$ we change neither the realization of formulas nor CS, and just notice that the concluding sequent is provable in $\mathsf{LP}(CS)$ given that the premises are. It is easy to see that every move down in the tree other than $(\Rightarrow\Box)$ is provable in $\mathsf{LP}(CS)$.

Consider a rule $(\Rightarrow\Box)$ of a family f, and let this rule have number i in the numbering of all rules $(\Rightarrow\Box)$ from a given family f. The corresponding node in $\mathcal{T}'$ is labelled by

$$\frac{y_1{:}B_1,\ldots,y_k{:}B_k \Rightarrow B}{y_1{:}B_1,\ldots,y_k{:}B_k \Rightarrow (u_1+\ldots+u_{n_f}){:}B}\ ,$$

where $y_1,\ldots,y_k$ are proof variables introduced in (1), $u_1,\ldots,u_{n_f}$ are proof polynomials, and u_i is a provisional variable. By the induction hypothesis, the premise sequent $y_1 : B_1,\ldots,y_k : B_k \Rightarrow B$ is derivable in $\mathsf{LP}(CS)$. By Lemma 191, construct a proof polynomial $t(y_1,\ldots,y_n)$ and extend the constant specification to get a new injective CS such that

$$\mathsf{LP}(CS) \vdash y_1{:}B_1,\ldots,y_k{:}B_k \Rightarrow t(y_1,\ldots,y_n){:}B.$$

Since

$$\mathsf{LP}_0 \vdash t{:}B \rightarrow (u_1{+}\ldots{+}u_{i-1}{+}t{+}u_{i+1}{+}\ldots{+}u_{n_f}){:}B\ ,$$

$$\mathsf{LP}\textit{(CS)} \vdash y_1{:}B_1,\ldots,y_k{:}B_k \Rightarrow (u_1{+}\ldots{+}u_{i-1}{+}t{+}u_{i+1}{+}\ldots{+}u_{n_f}){:}B.$$

Now substitute $t(y_1,\ldots,y_n)$ for u_i everywhere in the derivation tree and in the current *CS*. The latter remains injective after such a substitution, though this operation may lead to constant specifications of the sort $c{:}A(c)$ where $A(c)$ contains c.

Note that $t(y_1,\ldots,y_n)$ has no provisional variables, hence such a substitution is always possible. Moreover, after the substitution there is one less provisional variable (namely u_i) left, which guarantees termination. The conclusion of the rule ($\Rightarrow\Box$) under consideration becomes derivable in LP*(CS)*, and the induction step is complete.

Eventually, we substitute polynomials of non-provisional variables for all provisional variables and build a realization of the root sequent of the proof tree derivable in LP*(CS)*. Obviously, the realization r built by this procedure is normal. ■

COROLLARY 199 (Realization of $\mathsf{S4}$).

$$\mathsf{S4} \vdash F \quad \Leftrightarrow \quad \mathsf{LP} \vdash F^r \ \textit{for some realization } r.$$

The realization algorithm above is in fact exponential in the length of a given cut-free derivation of $\mathsf{S4}$ mostly because of a repeating use of the Lifting Lemma. A polynomial time realization algorithm was recently offered by V. Brezhnev and R. Kuznets (not yet published).

Some $\mathsf{S4}$-theorems admit essentially different realizations in LP. For example, among possible realizations of $\Box F \vee \Box F \rightarrow \Box F$ there are

$$x{:}F{\vee}y{:}F{\rightarrow}(x+y){:}F \quad\text{and}\quad x{:}F{\vee}x{:}F{\rightarrow}x{:}F.$$

The former of these formulas is a meaningful specification of the operation "+", the latter one is a trivial tautology.

Modal formulas can be realized by some restricted classes of proof polynomials. For example, the standard realization of the $\mathsf{S4}$-theorem $(\Box A{\vee}\Box B){\rightarrow}\Box(A{\vee}B)$ gives $(x{:}A{\vee}y{:}B){\rightarrow}(a{\cdot}x+b{\cdot}y){:}(A{\vee}B)$ with the injective constant specification $a{:}(A{\rightarrow}A{\vee}B)$, $b{:}(B{\rightarrow}A{\vee}B)$. The same modal formula can be realized in LP as $(c{:}A\vee c{:}B){\rightarrow}(c{\cdot}c){:}(A{\vee}B)$ with the constant specification $c:(A\rightarrow A\vee B)$, $c:(B\rightarrow A\vee B)$. However, the idea behind LP design has been to keep its language reasonably general, since realization of $\mathsf{S4}$ is not the only job the logic of proofs has been created for.

11.3 Standard provability semantics of LP

We shall work in PA. By Δ_1 and Σ_1 we mean the corresponding classes of arithmetical predicates. We will use x, y, z to denote individual variables in arithmetic and hope that the reader is able to distinguish them from the proof variables.

DEFINITION 200. We assume here that PA contains terms for all primitive recursive functions (cf. [Smoryński, 1985], [Takeuti, 1975]), called *primitive recursive terms*. Formulas $f(\vec{x}) = 0$ where $f(\vec{x})$ is a primitive recursive term are *standard primitive recursive formulas. A standard* Σ_1*-formula* is a formula $\exists x \varphi(x, \vec{y})$ where $\varphi(x, \vec{y})$ is a standard primitive recursive formula. An arithmetical formula φ is *provably* Σ_1 if it is provably equivalent in PA to a standard Σ_1-formula; φ is *provably* Δ_1 iff both φ and $\neg\varphi$ are provably Σ_1. Sometimes, we will omit a Gödel number symbol and write φ instead of $\ulcorner\varphi\urcorner$ when safe.

DEFINITION 201. A *proof predicate* is a provably Δ_1-formula $\mathsf{Prf}(x, y)$ such that for every arithmetical sentence φ

$$\mathsf{PA} \vdash \varphi \ \text{ iff } \ \text{for some } n \in \omega, \quad \mathsf{Prf}(n, \ulcorner\varphi\urcorner) \text{ holds.}$$

$\mathsf{Prf}(x, y)$ is *normal* if it satisfies the following two conditions:

1) (*finiteness of proofs*) For any k, the set $T(k) = \{l \mid \mathsf{Prf}(k, l)\}$ is finite. The function from k to the code of $T(k)$ is computable.

2) (*conjoinability of proofs*) For any k and l, there is n such that

$$T(k) \cup T(l) \subseteq T(n).$$

The conjoinability property yields that normal proof predicates are multi-conclusion ones.

EXAMPLE 202. The natural arithmetical proof predicate $\mathsf{Proof}(x, y)$

"*x is the code of a derivation containing a formula with the code y*".

is the standard example of a normal proof predicate.

Note, that every normal proof predicate can be transformed into a single-conclusion one by straightforwardly changing from

$$\text{“}p \text{ proves } F_1, \ldots, F_n\text{”} \qquad \text{to} \qquad \text{“}(p, i) \text{ proves } F_i,\ i = 1, \ldots, n\text{”}.$$

Moreover, every single-conclusion proof predicate may be regarded as normal multi-conclusion, e.g. by reading

$$\text{“}p \text{ proves } F_1 \wedge \ldots \wedge F_n\text{”} \qquad \text{as} \qquad \text{“}p \text{ proves each of } F_i,\ i = 1, \ldots, n\text{”}.$$

PROPOSITION 203. *For every normal proof predicate* Prf *there are computable functions* $\mathbf{m}(x,y)$, $\mathbf{a}(x,y)$ *and* $\mathbf{c}(x)$ *such that for all arithmetical formulas* φ, ψ *and all natural numbers* k, n *the following formulas are valid:*

$$\mathsf{Prf}(k, \varphi \to \psi) \wedge \mathsf{Prf}(n, \varphi) \to \mathsf{Prf}(\mathbf{m}(k,n), \psi)$$

$$\mathsf{Prf}(k, \varphi) \to \mathsf{Prf}(\mathbf{a}(k,n), \varphi), \qquad \mathsf{Prf}(n, \varphi) \to \mathsf{Prf}(\mathbf{a}(k,n), \varphi)$$

$$\mathsf{Prf}(k, \varphi) \to \mathsf{Prf}(\mathbf{c}(k), \mathsf{Prf}(k, \varphi)).$$

Proof. For example, the following function can be taken as $\mathbf{m}$:

$$\mathbf{m}(k,n) = \mu z.\text{“}\mathsf{Prf}(z,\psi) \textit{ holds for all } \psi \textit{ such that there are } \ulcorner \varphi \to \psi \urcorner \in T(k) \textit{ and } \ulcorner \varphi \urcorner \in T(n)\text{”} .$$

Likewise, for $\mathbf{a}$ one could take

$$\mathbf{a}(k,n) = \mu z. T(k) \cup T(n) \subseteq T(z).$$

Finally, $\mathbf{c}$ may be given by

$$\mathbf{c}(k) = \mu z.\text{“}\mathsf{Prf}(z, \mathsf{Prf}(k,\varphi)) \textit{ for all } \ulcorner \varphi \urcorner \in T(k)\text{”}.$$

Such a z always exists. Indeed, $\mathsf{Prf}(k,\varphi)$ is a true Δ_1-sentence for every $\ulcorner \varphi \urcorner \in T(k)$, therefore they are all provable in PA. Use conjoinability to find a uniform proof of all of them. ■

DEFINITION 204. An arithmetical *interpretation* $*$ of the LP-language has the following parameters:

- a normal proof predicate Prf with the functions $\mathbf{m}(x,y)$, $\mathbf{a}(x,y)$ and $\mathbf{c}(x)$ as in Proposition 203,
- an evaluation of propositional letters by sentences of arithmetic,
- an evaluation of proof variables and constants by natural numbers.

Let $*$ commute with boolean connectives,

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

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

Under an interpretation $*$ a proof polynomial t becomes the natural number t^*, an LP-formula F becomes the arithmetical sentence F^*. A formula $(t{:}F)^*$ is always provably Δ_1. For a set X of LP-formulas by X^* we mean the set of all F^*'s such that $F \in X$. Given a constant specification CS, an arithmetical interpretation $*$ is a *CS-interpretation*, if all formulas from CS^*

are true (equivalently, are provable in PA). An LP-formula F is *valid* (with respect to the arithmetical semantics), if F^* is true under all interpretations $*$. F is *provably valid* if $\mathsf{PA} \vdash F^*$ for any interpretation $*$. F is *valid under constant specification CS*, if F^* is true under all CS -interpretations $*$. F is *provably valid under constant specification CS*, if $\mathsf{PA} \vdash F^*$ for any *CS -interpretation* $*$. It is obvious that *provably valid* yields *valid*.

PROPOSITION 205 (Arithmetical soundness of LP_0).

If $\mathsf{LP}_0 \vdash F$ *then* F *is provably valid (hence, valid).*

Proof. A straightforward induction on the derivation in LP_0. Let us check the axiom $t{:}F \rightarrow F$. Under an interpretation $*$

$$(t{:}F \rightarrow F)^* \equiv \mathsf{Prf}(t^*, F^*) \rightarrow F^*.$$

Consider two possibilities. Either $\mathsf{Prf}(t^*, F^*)$ is true, in which case t^* is indeed a proof of F^*, thus, $\mathsf{PA} \vdash F^*$ and $\mathsf{PA} \vdash (t{:}F \rightarrow F)^*$. Otherwise $\mathsf{Prf}(t^*, F^*)$ is false, in which case being a false Δ_1-formula it is refutable in PA. Hence, $\mathsf{PA} \vdash \neg\mathsf{Prf}(t^*, F^*)$ and again $\mathsf{PA} \vdash (t{:}F \rightarrow F)^*$. ■

COROLLARY 206 (Arithmetical soundness of LP).

LP*(CS)* $\vdash F \;\Rightarrow\;$ F *is provably valid under the constant specification CS.*

The above provability semantics for LP may be characterized as a *call-by-value* semantics, since the evaluation F^* of a given LP-formula F depends upon the value of participating functions. A different *call-by-name* provability semantics for LP was introduced in [Artemov, 1995] and then used in [Krupski, 1997], [Sidon, 1997]. In the latter semantics, F^* depends upon the particular programs for the functions participating in $*$.

Following [Artemov, 2001], we proceed with establishing an arithmetical completeness of the logic of proofs by building a decidable version of the canonical model for LP and then embedding this model into PA. As a side product we will get normalization theorem for a Gentzen-style formulation of LP.

11.4 A sequent formulation of logic of proofs

As before, by a *sequent* we mean a pair $\Gamma \Rightarrow \Delta$, where Γ and Δ are finite multisets of LP-formulas. For Γ, F we mean $\Gamma \cup \{F\}$. Without loss of generality we assume a boolean basis $\rightarrow, \bot$ and treat the remaining boolean connectives as definable ones.

Axioms of LP_0^G are sequents of the form $\Gamma, F \Rightarrow F, \Delta$ and $\Gamma, \bot \Rightarrow \Delta$. Along with the usual Gentzen sequent rules of classical propositional logic, including the cut and contraction rules (e.g. like **G2c** from [Troelstra and Schwichtenberg, 1996]), the system LP_0^G contains the rules

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

$$\frac{\Gamma \Rightarrow \Delta, t{:}A}{\Gamma \Rightarrow \Delta, (t+s){:}A}\ (\Rightarrow +) \qquad\qquad \frac{\Gamma \Rightarrow \Delta, t{:}A}{\Gamma \Rightarrow \Delta, (s+t){:}A}\ (\Rightarrow +)$$

$$\frac{\Gamma \Rightarrow \Delta, s{:}(A \rightarrow B) \qquad \Gamma \Rightarrow \Delta, t{:}A}{\Gamma \Rightarrow \Delta, (s \cdot t){:}B}\quad (\Rightarrow \cdot)$$

The system LP^G is LP_0^G plus the rule

$$\frac{\Gamma \Rightarrow A, \Delta}{\Gamma \Rightarrow c{:}A, \Delta}\quad (\Rightarrow c),$$

where A is an axiom *A0 – A4* of LP, and c is a proof constant.

LP^{G-} and LP_0^{G-} are the corresponding systems without the rule Cut.

By a straightforward induction both ways it is easy to establish

PROPOSITION 207. $\mathsf{LP}_0^G \vdash \Gamma \Rightarrow \Delta$ *iff* $\mathsf{LP}_0 \vdash \bigwedge \Gamma \rightarrow \bigvee \Delta$,
$\mathsf{LP}^G \vdash \Gamma \Rightarrow \Delta$ *iff* $\mathsf{LP} \vdash \bigwedge \Gamma \rightarrow \bigvee \Delta$.

COROLLARY 208. $\mathsf{LP}(CS) \vdash F$ *iff* $\mathsf{LP}_0^G \vdash CS \Rightarrow F$.

DEFINITION 209. The sequent $\Gamma \Rightarrow \Delta$ is *saturated* if

1. $A \rightarrow B \in \Gamma$ implies $B \in \Gamma$ or $A \in \Delta$,
2. $A \rightarrow B \in \Delta$ implies $A \in \Gamma$ and $B \in \Delta$,
3. $t{:}A \in \Gamma$ implies $A \in \Gamma$,
4. $!t{:}t{:}A \in \Delta$ implies $t{:}A \in \Delta$,
5. $(s+t){:}A \in \Delta$ implies $s{:}A \in \Delta$ and $t{:}A \in \Delta$
6. $(s \cdot t){:}B \in \Delta$ implies *for each* $X \rightarrow B$ *occurring as a subformula in* Γ, Δ *either* $s{:}(X \rightarrow B) \in \Delta$ *or* $t{:}X \in \Delta$.

LEMMA 210 (Saturation lemma). *Suppose* $\mathsf{LP}_0^{G-} \nvdash \Gamma \Rightarrow \Delta$. *Then there exists a saturated sequent* $\Gamma' \Rightarrow \Delta'$ *such that*

1. $\Gamma \subseteq \Gamma'$, $\Delta \subseteq \Delta'$,

2. $\Gamma' \Rightarrow \Delta'$ *is not derivable in* LP_0^{G-}.

Proof. Run a straightforward saturation algorithm, prove its termination ([Artemov, 2001]). ■

Note that in a saturated sequent $\Gamma \Rightarrow \Delta$ which is not LP_0^{G-}-derivable the set Γ is closed under the rules $t{:}X/X$ and $X \to Y, X/Y$.

LEMMA 211 (Completion Lemma). *For each saturated sequent $\Gamma \Rightarrow \Delta$ not derivable in LP_0^{G-} there is a set of LP-formulas $\widetilde{\Gamma}$ (a completion of $\Gamma \Rightarrow \Delta$) such that*

1. $\widetilde{\Gamma}$ is decidable, for each t the set $I(t) = \{X \mid t{:}X \in \widetilde{\Gamma}\}$ is finite and a function from a code[20] of t to a code[21] of $I(t)$ is computable,

2. $\Gamma \subseteq \widetilde{\Gamma}$, $\Delta \cap \widetilde{\Gamma} = \varnothing$,

3. if $t{:}X \in \widetilde{\Gamma}$, then $X \in \widetilde{\Gamma}$,

4. if $s{:}(X \to Y) \in \widetilde{\Gamma}$ and $t{:}X \in \widetilde{\Gamma}$, then $(s \cdot t){:}Y \in \widetilde{\Gamma}$,

5. if $t{:}X \in \widetilde{\Gamma}$, then $!t{:}t{:}X \in \widetilde{\Gamma}$,

6. if $t{:}X \in \widetilde{\Gamma}$, then $(t+s){:}X \in \widetilde{\Gamma}$ and $(s+t){:}X \in \widetilde{\Gamma}$.

Proof. By a straightforward completion algorithm ([Artemov, 2001]). ■

11.5 Completeness theorems

In this section we establish completeness and cut-elimination theorems for the logic of proofs.

THEOREM 212. *The following are equivalent*

1. $\mathsf{LP}_0^{G-} \vdash \Gamma \Rightarrow \Delta$,

2. $\mathsf{LP}_0^{G} \vdash \Gamma \Rightarrow \Delta$,

3. $\mathsf{LP}_0 \vdash \bigwedge \Gamma \to \bigvee \Delta$,

4. $\bigwedge \Gamma \to \bigvee \Delta$ is provably valid in arithmetic,

5. $\bigwedge \Gamma \to \bigvee \Delta$ is valid in arithmetic.

Proof. The steps from 1 to 2 and from 4 to 5 are trivial. The step from 2 to 3 follows from Proposition 207 and the step from 3 to 4 follows from Proposition 205. It suffices now to prove that 1 follows from 5. We assume "not 1" and establish "not 5". Suppose $\mathsf{LP}_0^{G-} \nvdash \Gamma \Rightarrow \Delta$. Our aim now will be to construct an arithmetical interpretation $*$ such that $(\bigwedge \Gamma \to \bigvee \Delta)^*$ is false in the standard arithmetical sense.

From Lemma 210 get a saturated sequent $\Gamma' \Rightarrow \Delta'$, and then perform a completion (Lemma 211) to get a set of formulas $\widetilde{\Gamma}'$.

We define the desired interpretation $*$ on propositional letters S_i, proof variables x_j and proof constants a_j first. We assume that Gödel numbering of the joint language of LP and PA is injective, i.e.

$$\ulcorner E_1 \urcorner = \ulcorner E_2 \urcorner \quad \leftrightarrow \quad E_1 \equiv E_2$$

[20] For example, the Gödel number of t.

[21] For example, the code of the finite set of Gödel numbers of formulas from $I(t)$.

for any expressions E_1, E_2, and that 0 is not a Gödel number of any expression. For a propositional letter S, proof variable x and proof constant a let

$$S^* = \begin{cases} \ulcorner S \urcorner = \ulcorner S \urcorner, & \text{if } S \in \widetilde{\Gamma}' \\ \ulcorner S \urcorner = 0, & \text{if } S \notin \widetilde{\Gamma}', \end{cases} \qquad x^* = \ulcorner x \urcorner, \qquad a^* = \ulcorner a \urcorner.$$

The remaining parts of $*$ are constructed by an arithmetical fixed point equation below.

For any arithmetical formula $\mathsf{Prf}(x,y)$ define an auxiliary translation † of proof polynomials to numerals and LP-formulas to PA-formulas such that $S^\dagger = S^*$ for any propositional letter S, $t^\dagger = \ulcorner t \urcorner$ for any proof polynomial t, $(t{:}F)^\dagger = \mathsf{Prf}(t^\dagger, \ulcorner F^\dagger \urcorner)$, and † commutes with the propositional connectives.

It is clear that if $\mathsf{Prf}(x,y)$ contains quantifiers, then † is injective, i.e. $F^\dagger \equiv G^\dagger$ yields $F \equiv G$. Indeed, from $F^\dagger \equiv G^\dagger$ it follows that the principal connectives in F and G coincide. We consider one case: $(F_1 \to F_2)^\dagger \equiv (s{:}G)^\dagger$ is impossible. Since $(s{:}G)^\dagger \equiv \mathsf{Prf}(k,n)$ for the corresponding k and n, this formula contains quantifiers. Therefore, the formula $(F_1 \to F_2)^\dagger \equiv {F_1}^\dagger \to {F_2}^\dagger$ also contains quantifiers and thus contains a subformula of the form $\mathsf{Prf}(k_1, n_1)$. However, $(s{:}G)^\dagger \equiv {F_1}^\dagger \to {F_2}^\dagger$ is impossible, since the numbers of logical connectives and quantifiers in both parts of $\equiv$ are different. Now the injectivity of † can be shown by an easy induction on the construction of an LP-formula. Moreover, one can construct primitive recursive functions f and g such that

$$f(\ulcorner B \urcorner, \ulcorner \mathsf{Prf} \urcorner) = \ulcorner B^\dagger \urcorner, \quad g(\ulcorner B^\dagger \urcorner, \ulcorner \mathsf{Prf} \urcorner) = \ulcorner B \urcorner.$$

Let $(\mathsf{Proof}, \otimes, \oplus, \Uparrow)$ be the standard multi-conclusion proof predicate from Example 202, with $\otimes$ standing for "application", $\oplus$ for "union" and $\Uparrow$ for "proof checker" operations associated with Proof. In particular, for any arithmetical formulas φ, ψ and any natural numbers k, n the following arithmetical formulas are true:

$$\begin{array}{l} \mathsf{Proof}(k, \varphi \to \psi) \wedge \mathsf{Proof}(n, \varphi) \to \mathsf{Proof}(k \otimes n, \psi) \\ \mathsf{Proof}(k, \varphi) \to \mathsf{Proof}(k \oplus n, \varphi) \\ \mathsf{Proof}(n, \varphi) \to \mathsf{Proof}(k \oplus n, \varphi) \\ \mathsf{Proof}(k, \varphi) \to \mathsf{Proof}(\Uparrow k, \mathsf{Proof}(k, \varphi)). \end{array}$$

Without loss of generality we assume that $\mathsf{Proof}(\ulcorner t \urcorner, k)$ is false for any proof polynomial t and any $k \in \omega$.

Let $\varphi(\vec{y}, z)$ be a provably Σ_1 arithmetical formula. Without loss of generality we assume that $\varphi(\vec{y}, z)$ is provably equivalent to $\exists x \psi(x, \vec{y}, z)$, for some provably Δ_1-formula $\psi(x, \vec{y}, z)$. By $\mu z.\varphi(\vec{y}, z)$ we mean a function $z = f(\vec{y})$ that, given $\vec{y}$,

1. Calculates the first pair of natural numbers (k, l) such that $\psi(k, \vec{y}, l)$ holds;
2. Puts $z = l$.

It is clear that $\mu z.\varphi(\vec{y}, z)$ is computable (though not necessarily total).

By the fixed point argument we construct a formula $\mathsf{Prf}(x, y)$ such that PA proves the following *fixed point equation (FPE)*:

$$\begin{array}{lll}\mathsf{Prf}(x,y) & \leftrightarrow & \mathsf{Proof}(x,y) \quad \vee \\ & & (\text{“}\mathit{x = \ulcorner t \urcorner \ for\ some\ t\ and\ y = \ulcorner B^{\dagger} \urcorner\ for\ some\ B \in I(t)}\text{”})\end{array}$$

The arithmetical formula “...” above describes a primitive recursive procedure: given x and y recover t and B such that $x = \ulcorner t \urcorner$ and $y = \ulcorner B^{\dagger} \urcorner$, then verify $B \in I(t)$. From *FPE* it is immediate that Prf is a provably Δ_1-formula, since $\mathsf{Proof}(x, y)$ is provably Δ_1. It also follows from *FPE* that $\mathsf{PA} \vdash \psi$ yields $\mathsf{Prf}(k, \psi)$, for some $k \in \omega$.

We define the arithmetical formulas $M(x, y, z)$, $A(x, y, z)$, $C(x, z)$ as follows. Here s, t denote proof polynomials.

$$\begin{array}{ll}M(x,y,z) & \leftrightarrow (\text{“}\mathit{x = \ulcorner s \urcorner\ and\ y = \ulcorner t \urcorner\ for\ some\ s\ and\ t}\text{”} \wedge z = \ulcorner s \cdot t \urcorner) \\ & \quad \vee \\ & (\text{“}\mathit{x = \ulcorner s \urcorner\ for\ some\ s\ and\ y \neq \ulcorner t \urcorner\ for\ any\ t}\text{”} \wedge \\ & \exists v[\text{“}v = \mu w.(\bigwedge\{\mathsf{Proof}(w, B^{\dagger}) \mid B \in I(s)\})\text{”} \wedge z = v \otimes y]) \\ & \quad \vee \\ & (\text{“}\mathit{x \neq \ulcorner s \urcorner\ for\ any\ s\ and\ y = \ulcorner t \urcorner\ for\ some\ t}\text{”} \wedge \\ & \exists v[\text{“}v = \mu w.(\bigwedge\{\mathsf{Proof}(w, B^{\dagger}) \mid B \in I(t)\})\text{”} \wedge z = x \otimes v]) \\ & \quad \vee \\ & (\text{“}\mathit{x \neq \ulcorner s \urcorner\ and\ y \neq \ulcorner t \urcorner\ for\ any\ s\ and\ t}\text{”} \wedge z = x \otimes y)\end{array}$$

$$\begin{array}{ll}A(x,y,z) & \leftrightarrow (\text{“}\mathit{x = \ulcorner s \urcorner\ and\ y = \ulcorner t \urcorner\ for\ some\ s\ and\ t}\text{”} \wedge z = \ulcorner s + t \urcorner) \\ & \quad \vee \\ & (\text{“}\mathit{x = \ulcorner s \urcorner\ for\ some\ s\ and\ y \neq \ulcorner t \urcorner\ for\ any\ t}\text{”} \wedge \\ & \exists v[\text{“}v = \mu w.(\bigwedge\{\mathsf{Proof}(w, B^{\dagger}) \mid B \in I(s)\})\text{”} \wedge z = v \oplus y]) \\ & \quad \vee \\ & (\text{“}\mathit{x \neq \ulcorner s \urcorner\ for\ any\ s\ and\ y = \ulcorner t \urcorner\ for\ some\ t}\text{”} \wedge \\ & \exists v[\text{“}v = \mu w.(\bigwedge\{\mathsf{Proof}(w, B^{\dagger}) \mid B \in I(t)\})\text{”} \wedge z = x \oplus v]) \\ & \quad \vee \\ & (\text{“}\mathit{x \neq \ulcorner s \urcorner\ and\ y \neq \ulcorner t \urcorner\ for\ any\ s\ and\ t}\text{”} \wedge z = x \oplus y)\end{array}$$

$$\begin{array}{ll}C(x,z) \leftrightarrow & (\text{“}\mathit{x = \ulcorner t \urcorner\ for\ some\ t}\text{”} \wedge z = \ulcorner !t \urcorner) \\ & \quad \vee \\ & (\text{“}\mathit{x \neq \ulcorner t \urcorner\ for\ any\ t}\text{”} \wedge \\ & \exists v[\text{“}v = \mu w.(\bigwedge\{\mathsf{Proof}(w, \mathsf{Proof}(t, \varphi) \to \mathsf{Prf}(t, \varphi)) \mid \\ & \varphi \in T(t)\})\text{”} \wedge z = v \otimes \Uparrow x])\end{array}$$

Here each of "..." denotes a natural arithmetical formula representing in PA the corresponding condition. Note that in the definitions of $M(x, y, z)$, $A(x, y, z)$ and $C(x, z)$ above, all the functions of sort $\mu w.\varphi$ are computable, since all the corresponding φ's are Δ_1. Therefore, $M(x, y, z)$, $A(x, y, z)$ and $C(x, z)$ are provably Σ_1. Let

$$\mathbf{m}(x, y) := \mu z.M(x, y, z), \quad \mathbf{a}(x, y) := \mu z.A(x, y, z), \quad \mathbf{c}(x) := \mu z.C(x, z).$$

As follows from the above, the functions $\mathbf{m}(x, y)$, $\mathbf{a}(x, y)$ and $\mathbf{c}(x)$ are computable. Moreover, Lemma 217 below yields that these functions are total.

We continue defining the interpretation $*$. Let Prf for $*$ be the one from *FPE*, and the functions $\mathbf{m}(x, y)$, $\mathbf{a}(x, y)$ and $\mathbf{c}(x)$ are as above.

LEMMA 213. *a)* $t^* = t^\dagger$ *for any proof polynomial* t,
b) $B^* \equiv B^\dagger$ *for any* LP*-formula* B.

Proof. a) Induction on the construction of a proof polynomial. Base cases are covered by the definition of the interpretation $*$. For the induction step note that according to the definitions, the following equalities are provable in PA:

$$(s \cdot t)^* = \mathbf{m}(s^*, t^*) = \mathbf{m}(\ulcorner s \urcorner, \ulcorner t \urcorner) = \ulcorner s \cdot t \urcorner = (s \cdot t)^\dagger,$$

$$(s + t)^* = \mathbf{a}(s^*, t^*) = \mathbf{a}(\ulcorner s \urcorner, \ulcorner t \urcorner) = \ulcorner s + t \urcorner = (s + t)^\dagger,$$

$$(!t)^* = \mathbf{c}(t^*) = \mathbf{c}(\ulcorner t \urcorner) = \ulcorner !t \urcorner = (!t)^\dagger.$$

b) By induction on B. The atomic case when B is a propositional letter holds by the definitions. If B is $t : F$, then $(t{:}F)^* \equiv \mathsf{Prf}(t^*, F^*)$. By a), $t^* = t^\dagger$. By the induction hypothesis, $F^* \equiv F^\dagger$ which yields $\ulcorner F^* \urcorner = \ulcorner F^\dagger \urcorner$. Therefore $\mathsf{Prf}(t^*, F^*) \equiv \mathsf{Prf}(t^\dagger, F^\dagger) \equiv (t{:}F)^\dagger$. The inductive steps are trivial. ■

COROLLARY 214. *The mapping* $*$ *is injective on terms and formulas of* LP. *In particular, for all expressions* E_1 *and* E_2,

$$E_1{}^* = E_2{}^* \quad \Rightarrow \quad E_1 \equiv E_2 \ .$$

COROLLARY 215. X^* *is provably* Δ_1, *for any* LP*-formula* X .

Indeed, if X is atomic, then X^* is provably Δ_1 by the definition of $*$. If X is $t{:}Y$, then $(t{:}Y)^*$ is $\mathsf{Prf}(t^*, Y^*)$. By Lemma 213,

$$\mathsf{PA} \vdash \mathsf{Prf}(t^*, Y^*) \leftrightarrow \mathsf{Prf}(t, Y^*).$$

The latter formula is provably Δ_1, therefore $(t{:}Y)^*$ is provably Δ_1. Since the set of provably Δ_1-formulas is closed under boolean connectives, X^* is provably Δ_1 for each X.

LEMMA 216. *If* $X \in \widetilde{\Gamma}'$, *then* $\mathsf{PA} \vdash X^*$. *If* $X \in \Delta'$, *then* $\mathsf{PA} \vdash \neg X^*$.

Proof. By induction on the length of X. Base case, i.e. X is atomic or $X = t{:}Y$. Let X be atomic. By the definition of $*$, X^* is true iff $X \in \widetilde{\Gamma'}$. Let $X = t{:}Y$ and $t{:}Y \in \widetilde{\Gamma'}$. Then $\mathsf{PA} \vdash$ "$Y \in I(t)$". By *FPE*, $\mathsf{PA} \vdash \mathsf{Prf}(t, Y^\dagger)$. By Lemma 213, $\mathsf{PA} \vdash \mathsf{Prf}(t^*, Y^*)$. Therefore $\mathsf{PA} \vdash (t{:}Y)^*$.

If $t : Y \in \Delta'$, then $t : Y \notin \widetilde{\Gamma'}$ and "$Y \in I(t)$" is false. The formula $\mathsf{Proof}(t^*, Y^*)$ is also false, since t^* is $\ulcorner t \urcorner$ (by Lemma 213) and $\mathsf{Proof}(t, k)$ is false for any k by assumption. By *FPE*, $(t{:}Y)^*$ is false. Since $(t{:}Y)^*$ is provably Δ_1 (Lemma 215), $\mathsf{PA} \vdash \neg(t{:}Y)^*$.

The induction steps corresponding to boolean connectives are standard and based on the saturation properties of $\Gamma' \Rightarrow \Delta'$. For example, let $X = Y \to Z \in \widetilde{\Gamma'}$. Then $Y \to Z \in \Gamma'$, and, by Definition 209, $Y \in \Gamma'$ or $Z \in \Delta'$. By the induction hypothesis, Y^* is true or Z^* is false, thus, $(Y \to Z)^*$ is true, etc. ■

LEMMA 217. $\mathsf{PA} \vdash \varphi \quad \Leftrightarrow \quad \mathsf{Prf}(n, \varphi)$ *for some* $n \in \omega$.

Proof. It remains to establish ($\Leftarrow$). Let $\mathsf{Prf}(n, \varphi)$ hold, for some $n \in \omega$. By *FPE*, either $\mathsf{Proof}(n, \varphi)$ holds or $\ulcorner \varphi \urcorner = \ulcorner B^\dagger \urcorner$, for some B such that $t{:}B \in \widetilde{\Gamma'}$. In the latter case by the saturation property of $\widetilde{\Gamma'}$, $B \in \widetilde{\Gamma'}$. By Lemma 216, $\mathsf{PA} \vdash B^*$. By the injectivity of the Gödel numbering, $\varphi \equiv B^\dagger$. By Lemma 213, $\varphi \equiv B^*$. Therefore $\mathsf{PA} \vdash \varphi$. ■

LEMMA 218. *For all arithmetical formulas φ, ψ and natural numbers k, n, the following is true*

a) $\mathsf{Prf}(k, \varphi \to \psi) \wedge \mathsf{Prf}(n, \varphi) \to \mathsf{Prf}(\mathbf{m}(k, n), \psi)$;

b) $\mathsf{Prf}(k, \varphi) \to \mathsf{Prf}(\mathbf{a}(k, n), \varphi), \qquad \mathsf{Prf}(n, \varphi) \to \mathsf{Prf}(\mathbf{a}(k, n), \varphi)$;

c) $\mathsf{Prf}(k, \varphi) \to \mathsf{Prf}(\mathbf{c}(k), \mathsf{Prf}(k, \varphi))$.

Proof. *a)* Assume $\mathsf{Prf}(k, \varphi \to \psi)$ and $\mathsf{Prf}(n, \varphi)$. There are four possibilities.

i) Neither of k, n is a Gödel number of a proof polynomial. By *FPE*, both $\mathsf{Proof}(n, \varphi)$ and $\mathsf{Proof}(k, \varphi \to \psi)$ hold, so $\mathsf{Proof}(k \otimes n, \psi)$ also does.

ii) Both k and n are equal to Gödel numbers of some proof polynomials, say $k = \ulcorner s \urcorner$ and $n = \ulcorner t \urcorner$. By *FPE*, φ is F^* and ψ is G^* for some LP-formulas F, G such that $F \to G \in I(s)$ and $F \in I(t)$. By the closure property of $\widetilde{\Gamma'}$ (Lemma 211(4)), $G \in I(s \cdot t)$. By *FPE*, $\mathsf{Prf}(s \cdot t, G^*)$. By Lemma 213 and by definitions, PA proves that

$$\ulcorner s \cdot t \urcorner = (s \cdot t)^* = \mathbf{m}(s^*, t^*) = \mathbf{m}(\ulcorner s \urcorner, \ulcorner t \urcorner)\mathbf{m}(k, n).$$

Thus, $\mathbf{m}(k, n) = \ulcorner s \cdot t \urcorner$ and $\mathsf{Prf}(\mathbf{m}(k, n), \psi)$ is true.

iii) k is not equal to the Gödel number of a proof polynomial, $n = \ulcorner t \urcorner$ for some proof polynomial t. By *FPE*, $\mathsf{Proof}(k, \varphi \to \psi)$ and $\varphi \equiv F^\dagger$ for some LP-formula F such that $F \in I(t)$. Compute the number

$$l = \mu w. (\bigwedge \{\mathsf{Proof}(w, B^\dagger) \mid B \in I(t)\})$$

by the following method. Take $I(t) = \{B_1, \ldots, B_l\}$. By definition, $B_i \in \widetilde{\Gamma}'$, $i = 1, \ldots, l$. By Lemma 216, $\mathsf{PA} \vdash {B_i}^*$ for all $i = 1, \ldots, l$. By Lemma 213, $\mathsf{PA} \vdash {B_i}^\dagger$ for all $i = 1, \ldots, l$. By the conjoinability property of Proof there exists w such that $\mathsf{Proof}(w, {B_i}^\dagger)$ for all $i = 1, \ldots, l$. Let j be the least such w. In particular, $\mathsf{Proof}(j, F^\dagger)$. By the definition of $\otimes$, $\mathsf{Proof}(k \otimes j, \psi)$. By the definition of M, $\mathsf{PA} \vdash \mathbf{m}(k, n) = k \otimes j$, therefore $\mathsf{Proof}(\mathbf{m}(k, n), \psi)$ holds.

Case iv): "s is a Gödel number of a proof polynomial, but t is not a Gödel number of any proof polynomial" is similar to (iii).

Part (b) can be checked in the same way as (a).

$c)$ Given $\mathsf{Prf}(k, \varphi)$ there are two possibilities.

i) $k = \ulcorner t \urcorner$ for some proof polynomial t. By *FPE*, $\varphi \equiv F^\dagger$ for some F such that $F \in I(t)$. By the closure property from Lemma 211(5) of $\widetilde{\Gamma}'$, $!t\!:\!t\!:\!F \in \widetilde{\Gamma}'$. By Lemma 216, $(!t\!:\!t\!:\!F)^*$ holds. By definitions,

$$(!t\!:\!t\!:\!F)^* \equiv \mathsf{Prf}(\mathbf{c}(t^*), \mathsf{Prf}(t^*, F^*)).$$

By Lemma 213, $t^* = \ulcorner t \urcorner$ and $F^* \equiv F^\dagger$. Therefore $t^* = k$, $F^* \equiv \varphi$ and

$$\mathsf{Prf}(\mathbf{c}(k), \mathsf{Prf}(k, \varphi)).$$

ii) $k \neq \ulcorner t \urcorner$ for any proof polynomial t. By *FPE*, $\mathsf{Proof}(k, \varphi)$ holds. By definition of the proof checking operation $\Uparrow$ for Proof,

$$\mathsf{Proof}(\Uparrow\! k, \mathsf{Proof}(k, \varphi)).$$

By the definition of C, in this case $\mathsf{PA} \vdash \mathbf{c}(k) = l \otimes \Uparrow\! k$ where l equals

$$\mu w. \bigwedge\{\mathsf{Proof}(w, \mathsf{Proof}(k, \psi) \to \mathsf{Prf}(k, \psi)) \mid \mathsf{Proof}(k, \psi)\}.$$

By the definition of l,

$$\mathsf{Proof}(l, \mathsf{Proof}(k, \varphi) \to \mathsf{Prf}(k, \varphi)).$$

Therefore

$$\mathsf{Proof}(l \otimes \Uparrow\! k, \mathsf{Prf}(k, \varphi)).$$

By *FPE*,

$$\mathsf{Prf}(l \otimes \Uparrow\! k, \mathsf{Prf}(k, \varphi)),$$

therefore

$$\mathsf{Prf}(\mathbf{c}(k), \mathsf{Prf}(k, \varphi)).$$

■

LEMMA 219. *The normality conditions for* Prf *are fulfilled.*

Proof. By *FPE*, Prf is provably Δ_1. It follows from *FPE* and Lemma 217 that for any arithmetical sentence φ

$$\mathsf{PA} \vdash \varphi \text{ if and only if } \mathsf{Prf}(n, \varphi) \text{ holds for some } n.$$

Finiteness of proofs. For each k, the set

$$T(k) = \{l \mid \mathsf{Prf}(k, l)\}$$

is finite. Indeed, if k is a Gödel number of a proof polynomial, we can use the finiteness of $I(t)$; otherwise we use the normality of Proof. An algorithm for the function from k to the code of $T(k)$ for Prf can be easily constructed from those for Proof, and from the decision algorithm for $I(t)$, Lemma 211(1).

Conjoinability of proofs for Prf is realized by the function $\mathbf{a}(x, y)$, since by Lemma 218,

$$T(k) \cup T(n) \subseteq T(\mathbf{a}(k, n)).$$

■

Let us finish the proof of the final "not 1 implies not 5" part of Theorem 212. Given a sequent $\Gamma \Rightarrow \Delta$ not provable in LP_0^{G-} we have constructed an interpretation $*$ such that Γ^* are all true, and Δ^* are all false in the standard model of arithmetic (Lemma 216). Therefore, $(\bigwedge \Gamma \to \bigvee \Delta)^*$ is false. ■

COROLLARY 220 (Completeness of LP).

LP*(CS)* $\vdash F$ iff F *is (provably) valid under constant specification CS* .

COROLLARY 221.

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

COROLLARY 222. LP_0 *is decidable.*

Given an LP-formula F run the saturation algorithm $\mathcal{A}$ on a sequent $\Rightarrow F$. If $\mathcal{A}$ fails, then $\mathsf{LP}_0 \vdash F$. Otherwise, $\mathsf{LP}_0 \not\vdash F$.

COROLLARY 223 (Cut-elimination in LP_0). *Every sequent derivable in* LP_0^G *can be derived without the cut rule.*

Proof. By Theorem 212, $\mathsf{LP}_0^{G-} \vdash \Gamma \Rightarrow \Delta$ iff $\mathsf{LP}_0^G \vdash \Gamma \Rightarrow \Delta$. ■

Cut-elimination in the intuitionistic version of LP was established by a syntactical method in [Artemov, 1998; Artemov, 2002]. An idea of a semantical proof of cut-elimination for the whole LP was presented in [Artemov, 2001], a detailed proof was given in [Renne, 2004].

Decidability of LP follows from the results of [Mkrtychev, 1997]. This fact can also be easily obtained from the cut-elimination property of LP.

COROLLARY 224 (Non-emptiness of provability semantics for LP).

For any constant specification CS there exists a CS-interpretation $*$.

Proof. An easy inspection of the rules in LP_0^G shows that the sequent $CS \Rightarrow$ is not derivable in LP_0^{G-}, thus, $\mathsf{LP}_0^G \nvdash CS \Rightarrow$. Indeed, if $\mathsf{LP}_0^{G-} \vdash c\!:\!A \Rightarrow$, then $c\!:\!A$ is introduced by the rule $(: \ \Rightarrow)$ from a previously derived sequent $A \Rightarrow$. This is impossible, since A is an axiom of LP_0, thus, $\mathsf{LP}_0^G \vdash \ \Rightarrow A$: should $\mathsf{LP}_0^G \vdash A \Rightarrow$, we would have $\mathsf{LP}_0^G \vdash \ \Rightarrow$, which is impossible, e.g. because $\mathsf{LP}_0^{G-} \nvdash \ \Rightarrow$.

From $\mathsf{LP}_0^G \nvdash CS \Rightarrow$ it follows that $\mathsf{LP}_0^G \nvdash \ \Rightarrow \neg CS$. By Theorem 212, there exists an interpretation $*$ such that $(\neg CS)^*$ is false, i.e. CS^* is true. ∎

COROLLARY 225 (Arithmetical completeness of $\mathsf{S4}$).

$$\mathsf{S4} \vdash F \quad \Leftrightarrow \quad F^r \textit{ is (provably) valid for some realization } r.$$

DEFINITION 226. A propositional formula F is *proof realizable* if $(t(F))^r$ is valid under some realization r.

THEOREM 227 (Provability completeness of IPC). *For any formula F*

$$\mathsf{IPC} \vdash F \quad \Leftrightarrow \quad F \textit{ is proof realizable.}$$

Proof. A straightforward combination of Gödel-McKinsey-Tarksi reduction of IPC to $\mathsf{S4}$:

$$\mathsf{IPC} \vdash F \quad \Leftrightarrow \quad \mathsf{S4} \vdash t(F)$$

([Gödel, 1933; McKinsey and Tarski, 1948], cf. also [Chagrov and Zakharyaschev, 1997] Section 3.9, [Troelstra and Schwichtenberg, 1996] Sections 10.2 and 10.6), and the arithmetical completeness of $\mathsf{S4}$, (Corollary 225). ∎

Theorem 227 provides an exact specification of IPC by means of classical notion of proof consistent with *BHK* semantics.

Why, despite a steady interest in the subject, did the problem of finding the intended provability semantics for intuitionistic and modal logic $\mathsf{S4}$ evade the solution for so long? Here are some challenges that were to be met.

1. The search for the right provability format for modal logic $\mathsf{S4}$ proved long and difficult. Gödel's work of 1938 where he suggested such a format was not published until the actual solution to the problem, so it did not help the process. Solving the problem in a purely modal language could not be achieved, as was noticed in [Montague, 1963].

2. The technical part turned out not to be easy either. Many approaches coming from the related areas, such as λ-calculus or combinatory logic, had to be reconsidered. Turning from quantifiers over proofs to functions representing proofs "á la Skolem" gives rise to the phenomenon of self-referentiality. A question had to be answered, how to handle expressions of type $t\!:\!F(t)$, where a proof term t may occur in the very formula it proves. The necessary experience in dealing with self-referentiality in a similar context was gathered during more than 20 years of research in provability logic, where modality $\Box F$ was interpreted as $\mathsf{Prov}(F)$ (cf. the first part of this article).

12 MODELS FOR THE LOGIC OF PROOFS

The logic of proofs LP is complete with respect to the natural provability semantics. Still, having a family of convenient artificial models for LP could be very important for a successful study of LP and its applications.

The first artificial models for LP were introduced in [Mkrtychev, 1997]. Let us fix a constant specification $CS = \{c_1\!:\!A_1, c_2\!:\!A_2, \dots\}$, where each c_i is a proof constant and each A_i is an axiom of LP. A Mkrtychev model (M-model; in [Mkrtychev, 1997] these structures are called *pre-models*) for LP, corresponding to the constant specification CS, is a pair of mappings $(*, \Vdash)$. Here a *witness function* $*$ maps every proof polynomial t into a set $*(t)$ of formulas, which admit t as an "acceptable witness." $\Vdash$, in turn, is a truth assignment on formulas. The witness function is consistent with the constant specification CS, i.e. if $c\!:\!A \in CS$ then $A \in *(c)$. In addition, $*$ is consistent with the operations of LP, i.e.

$$\textit{if } (F \to G) \in *(s) \textit{ and } F \in *(t) \textit{ then } G \in *(s \cdot t)$$

$$\textit{if } F \in *(t) \textit{ then } t\!:\!F \in *(!\ t)$$

$$*(s) \cup *(t) \subseteq *(s+t)$$

The truth assignment $\Vdash$ is defined by an (arbitrarily) assignment of truth values to propositional letters; it is then inductively distributed over all LP-formulas according to the usual laws for Boolean connectives and the following condition for modalized formulas:

$$\Vdash t\!:\!F \quad \Leftrightarrow \quad F \in *(t) \ \text{ and } \Vdash F$$

In other words, "$t\!:\!F$ is true" means that "t is an acceptable witness for F and that F is true". In principle for any given M-model, it is possible to construct another model equivalent to M by restricting the witness function $*$ in a certain way so that the truth value of formulas $t\!:\!F$ would depend solely upon the witness function:

$$\Vdash t\!:\!F \quad \Leftrightarrow \quad F \in *(t)$$

In [Mkrtychev, 1997], soundness and completeness of LP with respect to M-models established. This result was obtained via the classical method of maximal consistent sets of formulas. M-models proved to be a convenient tool for studying the logic of proofs. For instance, they helped to establish in [Mkrtychev, 1997] the decidability of LP.

[Kuznets, 2000] obtained an upper bound Σ_2^p on the satisfiability problem for LP-formulas in M-models. This bound was lower than the known upper bound PSpace on the satisfiability problem in $\mathsf{S4}$. One of the possible explanations, why LP wins in complexity over closely related to it $\mathsf{S4}$, is that the satisfiability test for LP is somewhat similar to the type checking, i.e. checking the correctness of assigning types (formulas) to terms (proofs), which is known to be relatively easy in classical cases.

M-models were further explored in N. Krupski's paper [Krupski(jr.), 2003], where he constructs the minimal model of LP, which completely describes derivability in LP of "modalized" formulas (i.e. formulas of type $t{:}F$), namely

$$\Vdash t{:}F \quad \Leftrightarrow \quad \mathsf{LP} \vdash t{:}F$$

This yielded a better upper bound (NP) for the "modalized" fragment of the logic of proofs ([Krupski(jr.), 2003]). The minimal model is also used in [Krupski(jr.), 2003] to answer a well-known question about the disjunctive property of the logic of proofs:

$$\mathsf{LP} \vdash s{:}F \vee t{:}G \quad \Leftrightarrow \quad \mathsf{LP} \vdash s{:}F \ \text{ or } \ \mathsf{LP} \vdash t{:}G$$

M. Fitting in [Fitting, 2003a; Fitting, 2003b] gave a description of the canonical model for LP as a Kripke style model and established the fundamental *fully explanatory* property of the model: *if F is true in all the worlds reachable from Γ, then $t{:}F$ must be true in Γ for some polynomial t.* Fitting gives an application of the canonical model by providing an alternative "semantical" proof for the realizability theorem of $\mathsf{S4}$ in LP, whereby clarifying the role of operation "+" in this realization.

In [Fitting, 2003b; Fitting, 2005] a general definition of Kripke-style models for LP was also developed. A *frame* is a structure $(\mathcal{G}, \mathcal{R})$, where $\mathcal{G}$ is a non-empty set of *states* or *possible worlds*, and $\mathcal{R}$ is a binary relation on $\mathcal{G}$, called *accessibility.* Given a frame $(\mathcal{G}, \mathcal{R})$, a *possible evidence* function $\mathcal{E}$ is a mapping from states and proof polynomials to sets of formulas. We can read $X \in \mathcal{E}(\Gamma, t)$ as "X is one of the formulas that t serves as possible evidence for in state Γ." An evidence function must obey conditions that respect the intended meanings of the operations on proof polynomials. Except for monotonicity, the following have their origins in [Mkrtychev, 1997].

DEFINITION 228. $\mathcal{E}$ is an evidence function on $(\mathcal{G}, \mathcal{R})$ if, for all proof polynomials s and t, for all formulas X and Y, and for all $\Gamma, \Delta \in \mathcal{G}$:

1. *Application* $X \to Y \in \mathcal{E}(\Gamma, s)$ and $X \in \mathcal{E}(\Gamma, t)$ implies $Y \in \mathcal{E}(\Gamma, s \cdot t)$.

2. *Monotonicity* $\Gamma\mathcal{R}\Delta$ implies $\mathcal{E}(\Gamma, t) \subseteq \mathcal{E}(\Delta, t)$.

3. *Proof Checker* $X \in \mathcal{E}(\Gamma, t)$ implies $t{:}X \in \mathcal{E}(\Gamma, !t)$.

4. *Sum* $\mathcal{E}(\Gamma, s) \cup \mathcal{E}(\Gamma, t) \subseteq \mathcal{E}(\Gamma, s+t)$.

As usual in Kripke semantics, truth of atomic formulas at possible worlds is specified arbitrarily.

A structure $\mathcal{M} = (\mathcal{G}, \mathcal{R}, \mathcal{E}, \mathcal{V})$ is a *weak* LP-model provided $(\mathcal{G}, \mathcal{R})$ is a frame with $\mathcal{R}$ reflexive and transitive, $\mathcal{E}$ is an evidence function on $(\mathcal{G}, \mathcal{R})$, and $\mathcal{V}$ is a mapping from propositional variables to subsets of $\mathcal{G}$.

Given a weak model $\mathcal{M} = (\mathcal{G}, \mathcal{R}, \mathcal{E}, \mathcal{V})$, a forcing relation is defined by the following rules. For each $\Gamma \in \mathcal{G}$:

1. $\Gamma \Vdash S$ for a propositional variable S provided $\Gamma \in \mathcal{V}(S)$.

2. $\Gamma \Vdash \bot$ never holds—written $\Gamma \nVdash \bot$.

3. $\Gamma \Vdash X \to Y$ if and only if $\Gamma \nVdash X$ or $\Gamma \Vdash Y$.

4. $\Gamma \Vdash t{:}X$ if and only if $X \in \mathcal{E}(\Gamma, t)$ and, for every $\Delta \in \mathcal{G}$ with $\Gamma\mathcal{R}\Delta$, $\Delta \Vdash X$.

We say X is *true at world* Γ if $\Gamma \Vdash X$, and otherwise X is *false at* Γ.

A weak LP-model $\mathcal{M}$ is *Fully Explanatory* provided that, whenever $\Delta \Vdash X$ for every $\Delta \in \mathcal{G}$ such that $\Gamma\mathcal{R}\Delta$, then for some proof polynomial t we have $\Gamma \Vdash t{:}X$. If $\mathcal{M}$ is a weak LP-model, and if the Fully Explanatory condition is also met, then $\mathcal{M}$ is a *strong* LP-*model*.

The following completeness theorems have been established in [Fitting, 2003b; Fitting, 2005].

THEOREM 229. LP *is complete with respect to weak* LP-*models and with respect to strong* LP-*models.*

As it was noted by V. Krupski, LP is complete with respect to the class of Mkrtychev models with the fully explanatory property. Thus, completeness of LP with respect to strong Fitting models is reached on one-element models (i.e. on Mkrtychev models which are fully explanatory). The power of Fitting models became apparent in epistemic logics containing both proof polynomials and the usual S4-modality, since the accessibility relation does not degenerate in these logics (cf. Section 13).

A tableau system for the logic of proofs was developed in [Renne, 2004]. B. Renne also established a completeness theorem with respect to M-models and cut-elimination in the entirely of LP, though cut-elimination in LP with empty constant specifications was shown in [Artemov, 2001].

The interpolation property of the Logics of Proofs was studied by Tatiana Sidon (who later became Tatiana Yavorskaya) in [Sidon, 1998]. The standard formulation of the Craig interpolation property for given logic I

is that, if $I \vdash A \to B$, then there exists a formula C (an interpolant of A and B) in the *intersection of languages* of formulas A and B such that $I \vdash A \to C$ and $I \vdash C \to B$. For LP, one can consider two types of interpolation: a *weak* interpolation, where only propositional variables of C need be common to A and B, and a *strong* interpolation, where both propositional and proof variables of C have to be common to A and B. As was shown in [Sidon, 1998], the fragment of LP without functional symbols (also known as P or BLP) enjoys the strong interpolation property. Derivations in LP with an empty constant specification enjoy the weak (but not the strong) interpolation property. If CS is not empty, then even the weak interpolation property in $\mathsf{LP}(CS)$ may be violated.

13 FROM LOGICS OF PROOFS AND PROVABILITY TO EPISTEMIC LOGIC WITH JUSTIFICATIONS

The results of this subsection were taken from [Artemov and Nogina, 2004], unless stated otherwise. We describe a joint logic LPGL of provability and proofs, which is the arithmetically complete closure of the logic of provability GL and the logic of proofs LP. LPGL is a refinement of an earlier system LPP from [Sidon, 1997; Yavorskaya (Sidon), 2002]. We then describe systems $\mathsf{LPS4}$ and $\mathsf{LPS4}^-$, which may be regarded as basic epistemic logics with justifications.

DEFINITION 230. *Proof polynomials* for LPGL are the same as for the logic of proofs LP (cf. Definition 190).

Note, that there are more axioms and hence more choices to specify proof constants in LPGL, which makes LPGL-polynomials more expressive than the standard LP-polynomials.

DEFINITION 231. Using t to stand for any proof polynomial and S for any sentence variable, the formulas are defined by the grammar

$$A = S \mid A_1 \to A_2 \mid A_1 \wedge A_2 \mid A_1 \vee A_2 \mid \neg A \mid \Box A \mid t{:}A.$$

We assume also that "$t{:}$", "$\Box$" and "$\neg$" bind stronger than "$\wedge, \vee$", which bind stronger than "$\to$".

The logic of proofs and provability LPGL has axioms and rules as follows.

I. **Classical propositional logic**

A standard set of classical propositional axioms;
R1. *Modus Ponens.*

II. **Provability Logic** GL

GL1. $\Box(F \to G) \to (\Box F \to \Box G)$;
GL2. $\Box F \to \Box\Box F$;

GL3. $\Box(\Box F \to F) \to \Box F$;
R2. $\vdash F \;\Rightarrow \vdash \Box F$;
R3. $\vdash \Box F \Rightarrow \vdash F$ (*reflexivity rule*).

III. **Logic of Proofs** LP

LP1. $s\!:\!(F \to G) \;\to\; (t\!:\!F \to (s\cdot t)\!:\!G)$;
LP2. $t\!:\!F \;\to\, !t\!:\!(t\!:\!F)$;
LP3. $s\!:\!F \to (s{+}t)\!:\!F, \quad t\!:\!F \to (s{+}t)\!:\!F$;
LP4. $t\!:\!F \to F$;
R4. $\vdash c\!:\!A$, where A is an axiom and c is a proof constant.

IV. **New principles connecting explicit and formal provability**

C1. $t\!:\!F \to \Box F$ (*explicit-implicit connection*);
C2. $\neg(t\!:\!F) \to \Box\neg(t\!:\!F)$ (*negative introspection*);
C3. $t\!:\!\Box F \to F$ (*weak reflexivity*).

As in Subsection 11.1, a *constant specification CS* is a finite set $\{c_1 : A_1, \ldots, c_n\!:\!A_n\}$ of formulas, where each A_i is an axiom from I-IV and each c_i is a proof constant. By LPGL*(CS)* we mean a subsystem of LPGL where R4 is restricted to producing formulas from a given *CS* only. In particular, LPGL($\varnothing$) is a subsystem of LPGL without R4.

The reflexivity rule R3 is usually omitted in the standard formulation of GL, since it is an admissible rule of the latter (cf. [Boolos, 1979b; Boolos, 1993]). The same holds here: the rule R3 is derivable from the rest of LPGL (below). However, we need R3 as a postulated rule of the system to guarantee a good behavior of LPGL*(CS)*'s. Another curious feature is the fact that the axiom C3 is derivable from the rest of LPGL($\varnothing$).

LEMMA 232. LPGL($\varnothing$) $\vdash t\!:\!\Box F \to F$.

Proof.
1. $\neg\Box F \to \neg t\!:\!\Box F$ (contrapositive of LP4);
2. $\neg t\!:\!\Box F \to \Box(\neg t\!:\!\Box F)$ axiom C2;
3. $\Box(\neg t\!:\!\Box F) \to \Box(t\!:\!\Box F \to F)$, by reasoning in GL;
4. $\neg\Box F \to \Box(t\!:\!\Box F \to F)$, from 1, 2 and 3;
5. $\Box F \to \Box(t\!:\!\Box F \to F)$, by reasoning in GL;
6. $\Box(t\!:\!\Box F \to F)$, from 4 and 5, by propositional reasoning;
7. $t\!:\!\Box F \to F$, by R3. ■

However, proof constants corresponding to C3 are needed to guarantee the internalization property of LPGL. There are other innocent redundancies in the above formulation of LPGL, for example, GL2 is derivable from the rest of the system (cf. [Boolos, 1979b]).

LEMMA 233. *The following are provable in* LPGL($\varnothing$) *(hence in* LPGL *and in* LPGL*(CS) for any constant specification CS).*

1. $x{:}F \to \Box x{:}F$ *(positive introspection)*
2. $\Box x{:}F \vee \Box\neg x{:}F$ *(decidability of proof assertions)*

Proof.

1. $x{:}F \to\ !x{:}x{:}F$, by LP2;
$!x{:}x{:}F \to \Box x{:}F$, by C1;
$x{:}F \to \Box x{:}F$, by propositional logic .

2. $x{:}F \to \Box x{:}F$, by the previous item of this lemma;
$\neg(x{:}F) \to \Box\neg(x{:}F)$, by C2; ■
$\Box x{:}F \vee \Box\neg x{:}F$, by propositional logic.

LEMMA 234. $\mathsf{LPGL}(CS) \vdash\ F \quad\Leftrightarrow\quad \mathsf{LPGL}(\varnothing)\ \vdash\ \bigwedge CS \to F.$

Proof. Similar to Lemma 2.1 from [Yavorskaya (Sidon), 2002], by induction on a derivation of F in $\mathsf{LPGL}(CS)$. The only nontrivial cases are the rules of necessitation and reflection.

If F is obtained by the necessitation rule, i.e. F is $\Box G$ and $\mathsf{LPGL}(CS) \vdash G$, then, by the induction hypothesis, $\mathsf{LPGL}(\varnothing) \vdash \bigwedge CS \to G$. By GL reasoning,

$$\mathsf{LPGL}(\varnothing) \vdash \Box\bigwedge CS \to \Box G.$$

By positive introspection (Lemma 233.1) and some trivial GL reasoning,

$$\mathsf{LPGL}(\varnothing) \vdash \bigwedge CS \to \Box\bigwedge CS,$$

hence, $\mathsf{LPGL}(\varnothing) \vdash \bigwedge CS \to F$.

If F is obtained by the reflection rule, then $\mathsf{LPGL}(CS) \vdash \Box F$ and, by the induction hypothesis, $\mathsf{LPGL}(\varnothing) \vdash \bigwedge CS \to \Box F$, hence $\mathsf{LPGL}(\varnothing) \vdash \neg\bigwedge CS \vee \Box F$. By the negative introspection (axiom C2) and some GL reasoning,

$$\mathsf{LPGL}(\varnothing) \vdash \neg\bigwedge CS \to \Box\neg\bigwedge CS.$$

Therefore, $\mathsf{LPGL}(\varnothing) \vdash \Box\neg\bigwedge CS \vee \Box F$ and $\mathsf{LPGL}(\varnothing) \vdash \Box(\neg\bigwedge CS \vee F)$. By the reflection rule, $\mathsf{LPGL}(\varnothing) \vdash \neg\bigwedge CS \vee F$, hence, $\mathsf{LPGL}(\varnothing) \vdash \bigwedge CS \to F$. ■

Note, that both positive and negative introspection are needed to reduce the whole of LPGL to its fragment with unspecified constants $\mathsf{LPGL}(\varnothing)$.

LEMMA 235. *For any formula F there are proof polynomials $\mathtt{up}_F(x)$ and $\mathtt{down}_F(x)$ such that LPGL proves*

1. $x{:}F \to \mathtt{up}_F(x){:}\Box F;$
2. $x{:}\Box F \to \mathtt{down}_F(x){:}F$.

Proof.

1. $x{:}F \to \Box F$, by C1;
   $a{:}(x{:}F \to \Box F)$, specifying constant a, by R4;
   $!x{:}x{:}F \to (a{\cdot}!x){:}\Box F$, by LP1 and propositional logic;
   $x{:}F \to !x{:}x{:}F$, by LP2;
   $x{:}F \to (a{\cdot}!x){:}\Box F$, by propositional logic .

It suffices now to put $\mathtt{up}_F(x)$ equal to $a{\cdot}!x$ such that $a{:}(x{:}F \to \Box F)$.

2. $x{:}\Box F \to F$, by C3;
   $b{:}(x{:}\Box F \to F)$, specifying constant b, by R4;
   $!x{:}x{:}\Box F \to (b{\cdot}!x){:}F$, by LP1 and propositional logic;
   $x{:}\Box F \to !x{:}x{:}\Box F$, by LP2;
   $x{:}\Box F \to (b{\cdot}!x){:}F$, by propositional logic .

It suffices now to put $\mathtt{down}_F(x)$ equal to $b{\cdot}!x$ such that $b{:}(x{:}\Box F \to F)$ ■

PROPOSITION 236 (Constructive necessitation in LPGL).
If LPGL $\vdash F$ *then* LPGL $\vdash p{:}F$ *for some proof polynomial* p.

Proof. Induction on a derivation of F. The only interesting cases are rules R2 and R3. If F is obtained by R2, then $F = \Box G$ and $\vdash G$. By the induction hypothesis, $\vdash t{:}G$ for some proof polynomial t. Use lemma 235.1 to conclude that $\vdash \mathtt{up}_G(t){:}\Box G$ and put $p = \mathtt{up}_G(t)$. If F is obtained by R3, then $\vdash \Box F$. By the induction hypothesis, $\vdash t{:}\Box F$ for some proof polynomial t. Use lemma 235.2 to conclude that $\vdash \mathtt{down}_F(t){:}F$ and put $p = \mathtt{down}_F(t)$. Note that the presented derivation of $p{:}F$ does not use R2. ■

An easy modification of the above argument shows that LPGL enjoys the internalization property.

COROLLARY 237. *The necessitation rule* R2 *is derivable from the rest of* LPGL.

Indeed, if $\vdash F$ then, by proposition 236, $\vdash p{:}F$ for some proof polynomial p. By C1, $\vdash \Box F$.

Note that R2 is not redundant in LPGL*(CS)* for any specific *CS*. Indeed, to emulate R2 one needs to apply constructive necessitation to the unbounded set of theorems, which requires an unbounded set of constant specifications.

13.1 Kripke models for LPGL

DEFINITION 238. An LPGL-*model* is a triple $(K, \prec, \Vdash)$ where $(K, \prec)$ is a finite irreflexive tree with a unique root node, $\Vdash$ is a forcing relation between nodes of K and LPGL-formulas satisfying the following *forcing conditions*:

1. usual modal conditions for $\Box$, i.e. $\Vdash$ respects boolean connectives at each node, $a \Vdash \Box F$ iff $b \Vdash F$, for all $b \succ a$;

2. *stability* for every formula $t\!:\!F$ either all nodes of K force $t\!:\!F$ or all nodes of K force $\neg t\!:\!F$;

3. LP1-LP4 hold at each node.

A formula F holds in a model M if F is forced at each node of M. The root node of a model is called *root*. Put

$$S(F) = \{\Box G \to G \mid \Box G \textit{ is a subformula of } F\}.$$

We call a model F-sound, if $root \Vdash S(F)$. Similar conditions on a Kripke model could be found in [Artemov, 1994; Yavorskaya (Sidon), 2002]. Moreover, each LPP-model is also a LPGL-model. A model M is a *CS-model*, for a given constant specification $CS = \{c_1 : A_1, c_2 : A_2 \ldots c_n : A_n\}$, if M is X-sound and X holds in M for each $X = c_i\!:\!A_i$ from this *CS*. A formula F is *CS-valid* if it holds in each F-sound *CS*-model.

THEOREM 239 (Completeness). LPGL*(CS)* $\vdash F$ *iff* F *is CS-valid.*

13.2 Provability semantics for LPGL

The provability semantics for LPGL in Peano Arithmetic PA is the natural blend of those for the provability logic GL and the logic of proofs LP.

DEFINITION 240. Let $*$ be an interpretation from Definition 204. We define an arithmetical translation of proof terms and LPGL-formulas by induction as in Definition 204 with one extra clause:

$$(\Box F)^* = \exists x \mathsf{Prf}(x, \ulcorner F^* \urcorner).$$

THEOREM 241 (Arithmetical completeness). *For any given constant specification CS,* LPGL*(CS)* $\vdash F$ *iff* $\mathsf{PA} \vdash F^*$ *for all CS-interpretations* $*$.

13.3 Basic logics of knowledge with justifications

As we have already mentioned in Introduction, there is a well developed approach due to [Kuznetsov and Muravitsky, 1977; Goldblatt, 1978; Boolos, 1979b] of emulating $\mathsf{S4}$ in GL via strong provability modality $\boxdot F = F \wedge \Box F$. This translation could be applied to derive basic epistemic logics with justifications, $\mathsf{LPS4}$ and $\mathsf{LPS4}^-$, from LPGL.

DEFINITION 242. Polynomials and formulas in $\mathsf{LPS4}$ and $\mathsf{LPS4}^-$ are the same as in LPGL. The system $\mathsf{LPS4}$ has the following axioms and rules

I. **Classical propositional logic**.

II. **Basic Epistemic Logic** $\mathsf{S4}$ (as in Introduction).

III. **Logic of Proofs** LP (as in LPGL).

IV. **Principle connecting explicit and implicit knowledge**

C1. $t{:}F \to \Box F$.

$\mathsf{LPS4}$ contains both LP and $\mathsf{S4}$, enjoys the deduction theorem, analogues of the lifting lemma, internalization, constructive necessitation. The principle of *weak reflexivity* $t : \Box F \to F$ is derivable in $\mathsf{LPS4}$, *explicit reflexivity* $t{:}F \to F$ is redundant but we keep it listed for convenience.

DEFINITION 243. $\mathsf{LPS4}^- = \mathsf{LPS4}$ together with the negative introspection

C2. $\neg(t{:}F) \to \Box\neg(t{:}F)$.

For an alternative formulation of $\mathsf{LPS4}^-$ one could replace C1 and C2 by one principle of decidability of evidences.

As usual, by $\mathsf{LPS4}$*(CS)* and $\mathsf{LPS4}^-$*(CS)* we understand the corresponding systems where specifications of constants are taken from a given set *CS*.

13.4 Models for LPS4 *and* LPS4$^-$

The logics of proofs and provability gave us a clear idea how to build Kripke models for $\mathsf{LPS4}$ and $\mathsf{LPS4}^-$. We will consider here the case of $\mathsf{LPS4}^-$.

DEFINITION 244. An $\mathsf{LPS4}^-$-*model* is a triple $(K, \preceq, \Vdash)$ where $(K, \preceq)$ is a connected $\mathsf{S4}$-frame (transitive and reflexive), $\Vdash$ is a forcing relation between nodes of K and $\mathsf{LPS4}^-$-formulas satisfying the *forcing conditions.*

1. usual modal conditions for $\Box$, i.e. $\Vdash$ respects boolean connectives at each node, $a \Vdash \Box F$ iff $b \Vdash F$ for all $b \succeq a$;

2. *stability of explicit knowledge* every formula $t : F$ either holds at all nodes of K or does not hold at all nodes of K;

3. LP1-LP4 hold at each node.

A formula F holds in a model M if F holds at each node of M. M is an *CS*-model, for a given constant specification $CS = \{c_1{:}A_1, c_2{:}A_2 \ldots c_n{:}A_n\}$ if all $c_i : A_i \in CS$ hold in M. A formula F is *CS-valid* if it holds in each *CS*-model.

From this definition it follows that whereas $x \Vdash \Box F$ is understood in the conventional manner as *F holds in all worlds accessible from a given x*, an assertion $x \Vdash t{:}F$ says that *F holds in all worlds of the model.*

THEOREM 245 (Completeness). $\mathsf{LPS4}^-$*(CS)* $\vdash F$ *iff A is CS-valid.*

It follows from the proof of the above theorem that $\mathsf{LPS4}^-$*(CS)* is decidable for each constant specification *CS*.

The logic $\mathsf{LPS4}$ is, perhaps, the minimal epistemic logic with justifications when no specific assumptions were made concerning the character of the explicit knowledge operators. As a formal system $\mathsf{LPS4}$ behaves normally: it is closed under substitutions, enjoys the deduction theorem, internalization, has Kripke-style models, is sound with respect to the arithmetical semantics of the strong provability. Furthermore, $\mathsf{LPS4}$ has some interesting features not present in $\mathsf{LPS4}$. In particular, $\mathsf{LPS4}$ enjoyed the Fitting semantics (Definition 228) with the standard understanding of $x \Vdash \Box F$ ([Artemov and Nogina, 2004]). Furthermore, [Fitting, 2004] established the completeness of $\mathsf{LPS4}$ with respect to weak LP-models, Definition 228.

THEOREM 246. $\mathsf{LPS4}$ *is complete with respect to weak* LP*-models.*

[Fitting, 2004] also built a natural cut-free tableaux system for $\mathsf{LPS4}$.

13.5 Arithmetical semantics for LPS4 *and* LPS4$^-$

Arithmetical semantics for $\mathsf{LPS4}$ and $\mathsf{LPS4}^-$ is provided by the strong provability operator defined in the provability logic as $\boxdot F := F \wedge \Box F$.

DEFINITION 247. Define a translation $^+$ of $\mathsf{LPS4}$ formulas into the language of LPGL: $S^+ = S,\ \ (A \to B)^+ = (A \to B),\ \ (\neg A)^+ = \neg A,$
$(t\!:\!A)^+ = t\!:\!A,\ \ (\Box A)^+ = A \wedge \Box A.$

LEMMA 248. *If* $\mathsf{LPS4}, \mathsf{LPS4}^- \vdash F$*, then* $\mathsf{LPGL} \vdash F^+$*.*

An arithmetical provability semantics of $\mathsf{LPS4}$ and $\mathsf{LPS4}^-$ is inherited from the one of LPGL: in Definition 240 the item corresponding to the modality should be altered to the strong provability reading:

$$(\Box F)^* = F^* \wedge \exists x \mathsf{Prf}(x, \ulcorner F^* \urcorner).$$

THEOREM 249 (Arithmetical soundness of $\mathsf{LPS4}$ and $\mathsf{LPS4}^-$).
If $\mathsf{LPS4}, \mathsf{LPS4}^- \vdash F$*, then* $\mathsf{PA} \vdash F^*$ *for any arithmetical interpretation* $*$*.*

An arithmetically complete system $\mathsf{LPS4Grz}^-$ of the strong provability with proofs can be axiomatized by adding to $\mathsf{LPS4}^-$ the modal axiom by Grzegorczyk $\Box(\Box(F \to \Box F) \to F) \to F$. Kripke models for $\mathsf{LPS4Grz}^-$ are the special sort of $\mathsf{LPS4}^-$-models when the frame is a reflexive partial order.

14 THE LOGIC OF SINGLE-CONCLUSION PROOFS

By definition, each single-conclusion proof, also known as *functional proof*, proves a unique formula. As it was noticed earlier, the modality of provability corresponds to multi-conclusion proofs, whereas single-conclusion proofs lead to modal identities inconsistent with any normal modal logic.

In the functional logic of proofs, a formula $t\!:\!F$ still has the meaning "t is a proof of formula F," but the class of its possible interpretations is limited to functional proof systems only. The mathematical problem here was to give a full axiomatization of all resulting tautologies in the language of LP (without the operation "+" that is inconsistent with functional proofs).

The first step in finding such a complete system was to give a propositional description of the functionality property of proofs, which states that if $p{:}F \wedge p{:}G$ then F and G must coincide syntactically. The adequate description of this property was found in [Artemov and Strassen, 1992b] using a so-called *conditional unification.* It was then generalized in [Krupski, 1997; Krupski, 2002] to the full language of the logic of proofs.

Each formula C of type $t_1 : F_1 \wedge \ldots \wedge t_n : F_n$ generates a set of quasi-equations of type $S_C := \{\, t_i = t_j \Rightarrow F_i = F_j \mid 1 \leq i, j \leq n\,\}$. A *unifier* σ of a system S_C is a substitution σ such that either $t_i\sigma \not\equiv t_j\sigma$ or $F_i\sigma \equiv F_j\sigma$ holds for any i, j. Here and below "$X \equiv Y$" denotes the syntactic equality of X and Y.

DEFINITION 250. (Conditional unification.) $A = B\,(mod\, S)$ means that for each unifier σ of system S the property $A\sigma \equiv B\sigma$ holds.

LEMMA 251 (Decidability of conditional unification). *$A = B\,(mod\, S)$ is a decidable relation over A, B, S.*

DEFINITION 252. (Unification axiom) $t_1{:}F_1 \wedge \ldots \wedge t_n{:}F_n \rightarrow (A \leftrightarrow B)$, for each condition C of type $t_1{:}F_1 \wedge \ldots \wedge t_n{:}F_n$ and each A, B such that $A = B\,(mod\, S_C)$.

Logic FLP of functional proofs was introduced in [Krupski, 1997]. The language of FLP is the language of LP without the operation "+" and without proof constants. The axioms and rules of FLP are

Axiom schemes:

A0. Finite set of axiom schemes of classical propositional logic;

A1. $t{:}F \rightarrow F$;

A2. $t{:}(F \rightarrow G) \rightarrow (s{:}F \rightarrow (t{\cdot}s){:}G)$;

A3. $t{:}F \rightarrow\, !t{:}(t{:}F)$;

A4. Unification axiom.

Rule of inference:

R1. *modus ponens.*

The following result was obtained in [Krupski, 1997; Krupski, 2002].

THEOREM 253. FLP *is decidable, sound, and complete with respect to the arithmetical provability interpretation based on single-conclusion proof predicates.*

The logic of functional proofs was further developed in [Krupski, 2005], where a logic of proofs with references FLP_{ref} was introduced. System FLP_{ref} extends FLP with second-order variables which denote the operation of reconstructing an object from its reference, e.g., determining a formula proven by a given derivation. FLP_{ref} may also be viewed as a natural formal system for admissible propositional rules in arithmetic.

15 REFLECTION IN TYPED λ-CALCULUS AND COMBINATORY LOGIC

The logic of proofs naturally bridges two domains: the epistemic one, represented by modal logic and the computational one, represented by typed λ-calculi and combinatory logic. LP may be considered both as a realized modal logic S4 and as a typed combinatory logic with added expressive power. There is a natural analogy with the Curry-Howard isomorphism which states the identity of intuitionistic proofs and typed λ-terms understood as computational programs. I this section we will discuss what kind of new principles in λ-calculi and combinatory logic are prompted by this new development in the logic of proofs.

The original formulation of the logic of proofs LP was close to the typed combinatory logic format (cf. commentary in section 11) rather than in a more common format of λ-calculus. LP can emulate the λ-abstraction operator, e.g. in Curry-style as is standard in combinatory logic $\mathsf{CL}_{\rightarrow}$ (see, for example, [Troelstra and Schwichtenberg, 1996], p. 17). LP has many features not typical of λ-calculi, such as polymorphism, self-referentiality, capability to internalize its own derivations, etc. Polymorphism is not compatible with normalization property, and we have to give it up along with self-referentiality, which cannot be formulated in a language where each term has its own fixed type. On the other hand, the internalization property has had a prototype in λ-calculi, namely, the Curry-Howard isomorphism stated in form: if $A_1, \ldots, A_n \vdash B$ in intuitionistic logic, then $x_1{:}A_1, \ldots, x_n{:}A_n \vdash t(x_1, \ldots, x_n){:}B$ in λ-calculus for some λ-term t.

In [Alt and Artemov, 2001], some version $\boldsymbol{\lambda}^{\infty}$ of a reflexive λ-calculus was suggested that has an unrestricted internalization property. $\boldsymbol{\lambda}^{\infty}$ has the implicative intuitionistic (minimal) logic as a type system, a role of atoms is played by both usual propositional variables (atomic types) and statements of form $t{:}F$, where t is a term and F is a type. The "term:type" correspondence is a rigid typing "á la Church", i.e. each term has a unique type. $\boldsymbol{\lambda}^{\infty}$ has several countable series of term-building operations. Admissibility of the internalization rule for $\boldsymbol{\lambda}^{\infty}$ is proven as a metatheorem. The existence and the uniqueness of normal forms in $\boldsymbol{\lambda}^{\infty}$ all appear to be connected with the depth of type preservation during reductions, which is a new phenomenon compared to ordinary λ-calculus. This theory is currently under development.

15.1 Reflexive Combinatory Logic

Systems of combinatory logic are usually more compact than their corresponding λ-calculi. Reflexive combinatory logic RCL introduced in [Artemov, 2004] is no exception. RCL has the implicative intuitionistic (minimal) logic as a type system, a rigid typing. Reflexive combinatory terms are built from variables, "old" combinators $\mathbf{k}$ and $\mathbf{s}$, and new combinators $\mathbf{d}$, $\mathbf{o}$, and $\mathbf{c}$. The principles of RCL are

A1. $t{:}A \rightarrow A$;
A2. $\mathbf{k}{:}(A \rightarrow (B \rightarrow A))$;
A3. $\mathbf{s}{:}[(A \rightarrow (B \rightarrow C)) \rightarrow ((A \rightarrow B) \rightarrow (A \rightarrow C))]$;
A4. $\mathbf{d}{:}(t{:}A \rightarrow A)$;
A5. $\mathbf{o}{:}[u{:}(A \rightarrow B) \rightarrow (v{:}A \rightarrow (u \cdot v){:}B)]$;
A6. $\mathbf{c}{:}(t{:}A \rightarrow\ !t{:}t{:}A)$;

Modus Ponens rule. RCL has a natural provability semantics inherited from LP. Combinatory terms stand for proofs in PA or in intuitionistic arithmetic HA. Formulas $t : F$ are interpreted as arithmetical statements about provability, $\mathsf{Proof}(t, F)$, combinators $\mathbf{k}$, $\mathbf{s}$, $\mathbf{d}$, $\mathbf{o}$, and $\mathbf{c}$ denote terms corresponding to proofs of arithmetical translations of axioms A2–A6.

RCL evidently contains implicative intuitionistic logic, ordinary combinatory logic $\mathsf{CL}_{\rightarrow}$ and is closed under the combinatory application rule

$$\frac{u{:}(A \rightarrow B) \quad v{:}A}{(u{\cdot}v){:}B}\ .$$

The next theorem was established in [Artemov, 2004].

THEOREM 254. RCL *enjoys the internalization property: if* $A_1, \ldots, A_n \vdash B$*, then for any set of fresh variables* $x_1, \ldots, x_n$ *of respective types it is possible to construct a term* $t(x_1, \ldots, x_n)$ *such that*

$$x_1{:}A_1, \ldots, x_n{:}A_n \vdash t(x_1, \ldots, x_n){:}B\ .$$

One of the goals of RCL was to introduce a more expressive system of types and terms intended for programming language applications. Thus, it is interesting to consider the following natural (though informal) computational semantics for combinators of RCL. This semantics is based on the standard set-theoretic semantics of types, i.e. a type is a set and the implication type $U \rightarrow V$ is a set of functions from U to V. Some elements of a given type may be constructive objects which have *names*, i.e. computational programs. Terms of RCL are names of constructive objects, which are either specific (e.g. combinators $\mathbf{k}$, $\mathbf{s}$, $\mathbf{d}$, $\mathbf{o}$, or $\mathbf{c}$) or variable. The type $t{:}F$ is interpreted as a set, consisting of the object corresponding to term t. Basic combinators of RCL are understood as follows:

Combinators $\mathbf{k}$ and $\mathbf{s}$ are borrowed from the combinatory logic $\mathsf{CL}_{\rightarrow}$ along with their standard functional semantics. For example, $\mathbf{k}$ maps an element $x \in A$ into the constant function $\lambda y.x$ with y ranging over B.

The *denotate* combinator $\mathbf{d}{:}[t{:}F \rightarrow F]$ realizes the fundamental denotate function which maps a name (program) into the object with the given name. A primary example is the correspondence between indexes of computable functions and functions themselves.

The *interpreter* combinator $\mathbf{o}\!:\![u\!:\!(F \to G) \to (v\!:\!F \to (u{\cdot}v)\!:\!G)]$ realizes the interpreter program which maps program u and input v into the result of applying u to v.

The *coding* combinator $\mathbf{c} : [t : F \to !t : (t : F)]$ maps program t into its code $!t$ (alias, specific key in a database, etc.).

16 NEW APPROACH TO THE FOUNDATIONS OF VERIFICATION.

In the following example the provable principle of explicit reflection $t\!:\!F \to F$ from LP is applied to the verification theory. This example shows how the main idea of the logic of proofs, i.e., replacing quantifiers over proofs by explicit functions that realize actual proofs, may result in a more adequate mathematical model.

A metatheory for formal (possibly computerized) proof-checking systems, called here *verification systems*, was developed by Davis and Schwartz in [Davis and Schwartz, 1979]. They consider the following common scheme of building a formal verification system V. First the kernel V_0 of the system is chosen so that this kernel is elementary enough to declare its consistency evident and thus postulate it; V_0 is assumed to be expressive enough to be able to formalize proofs and check their correctness. The system is then extended with verified inference rules, e.g., with proven facts. Unlike what is typical in the foundations of mathematics where a theory is extended with new axioms to enhance the theory, here the extension of V is aimed at increasing effectiveness of the deductive apparatus of V, thereby maintaining the metamathematical power of V at the same level as that of V_0 if possible.

The inference rule Γ/F is considered to be verified in V, if

$$V \vdash \Box\Gamma \to \Box F \ ,$$

where $\Box\Gamma$, $\Box F$ are formal statements about provability in V of all formulas from Γ and of F, respectively. Adding the rule Γ/F to V yields $V' = V + \Gamma/F$. An extension V' of the system V is called *stable*, if V' is conservative over V. The main metamathematical question concerning this extension scheme is the following: is the theory V always capable of proving its own stability? In [Davis and Schwartz, 1979], this question was answered negatively. The reason can be explained by the following argument. Suppose one wishes to prove that $V' \vdash F$ implies $V \vdash F$ by an induction on the derivation in V'. The essential case of the induction step corresponds to the transition from $V \vdash \Gamma$ to $V \vdash F$ for an arbitrary instance of the inference rule under consideration, Γ/F. Internalization of $V \vdash \Gamma$ yields $V \vdash \Box\Gamma$; evidently, this step may be formalized in V. Using the fact that $V \vdash \Box\Gamma \to \Box F$ one gets $V \vdash \Box F$, whence it follows that $V \vdash F$. The last transition, from $V \vdash \Box F$ to $V \vdash F$, cannot (in general) be justified by means of V because reflection $\Box A \to A$ is not provable in V (Löb's theorem). The conclusion in [Davis and Schwartz, 1979] is that the considered process of building verification systems cannot be justified by initial assumptions of a system's consistency.

The paper [Artemov, 1999] analyzes the growth of metamathematical assumptions about a theory in a process of its extension according to the described scheme and suggests a new extension scheme free from the limitations of Davis and Schwartz's. According to this new scheme, a verification of inference rule Γ/F is done in an explicit manner by constructing a computable term $t(x)$ and a proof of $V \vdash x\!:\!\Gamma \to t(x)\!:\!F$. The advantage is that the stability of the resulting extensions is provable inside the system itself, which allows us to get rid of additional metamathematical assumptions about the given verification system. Here is the general argument showing the provable stability of the explicit verification. As before, the aim is to show that $V' \vdash F$ implies $V \vdash F$. The essential step of the induction on the derivation in V' is again the transition from $V \vdash \Gamma$ to $V \vdash F$ for the rule Γ/F. Internalizing a given derivation of $V \vdash \Gamma$ in the form of a term s yields $V \vdash s\!:\!\Gamma$. Using the fact that $V \vdash x\!:\!\Gamma \to t(x)\!:\!F$ one gets $V \vdash s\!:\!\Gamma \to t(s)\!:\!F$ and therefore $V \vdash t(s)\!:\!F$. The provability of explicit reflection $V \vdash t(s)\!:\!F \to F$ leads to a conclusion that $V \vdash F$.

An analysis of examples of implicit verification á la Davis-Schwartz shows that to prove "there exists a proof of Γ" implies in V that "there exists a proof of F," one usually first shows that $V \vdash x\!:\!\Gamma \to t(x)\!:\!F$ and then explicit terms x, $t(x)$ are replaced by existential quantifiers to satisfy the format of implicit verification. Thereby a knowledge of the term $t(x)$ was not used at all because of a lack of a theoretically justified mechanism for its utilization. Thus, in spite of the more restrictive format of explicit verification, it is reasonable to assume that in practice explicit verification is applicable if and only if implicit verification is.

Acknowledgments

We are deeply grateful to all our colleagues whose scientific contributions have shaped this wonderful field. Selecting the topics and papers to be represented in this Chapter was the hardest part of the job. There are plenty of first class results left outside the scope of the article. We refer the reader to excellent monographs [Boolos, 1979b; Smoryński, 1985; Boolos, 1993] and articles [Smoryński, 1981; Smoryński, 2004; de Jongh and Japaridze, 1998; Visser, 1998] which cover those results.

We would like to thank A. Visser for substantial help with Section 9 on the provability logic of HA and inspiring discussions. We are indebted to T. Yavorskaya (Sidon) who made valuable suggestions concerning Section 13 on logics of proofs and provability. Special thanks to J. Joosten, V. Krupski, R. Kuznets, E. Nogina, V. Shavrukov, B. Renne for numerous constructive comments and proofreading.

BIBLIOGRAPHY

[Ackermann, 1940] W. Ackermann. Zur Wiederspruchsfreiheit der reinen Zahlentheorie. *Math. Ann.*, 117:162–194, 1940.

[Adamovicz and Bigorajska, 1989] Z. Adamovicz and T. Bigorajska. Functions provably total in $I^{-}\Sigma_1$. *Fundamenta mathematicae*, 132:189–194, 1989.

[Allen *et al.*, 1990] S. Allen, R. Constable, D. Howe, and W. Aitken. The semantics of reflected proofs. In *Proceedings of the Fifth Annual IEEE Symposium on Logic in Computer Science*, pages 95–107, Los Alamitos, CA, USA, 1990. IEEE Computer Society Press.

[Alt and Artemov, 2001] J. Alt and S. Artemov. Reflective λ-calculus. In *Proceedings of the Dagstuhl-Seminar on Proof Theory in Computer Science*, volume 2183 of *Lecture Notes in Computer Science*, pages 22–37. Springer, 2001.

[Artemov and Beklemishev, 1993] S.N. Artemov and L.D. Beklemishev. On propositional quantifiers in provability logic. *Notre Dame Journal of Formal Logic*, 34:401–419, 1993.

[Artemov and Krupski, 1996] S. Artemov and V. Krupski. Data storage interpretation of labeled modal logic. *Annals of Pure and Applied Logic*, 78(1):57–71, 1996.

[Artemov and Nogina, 2004] S. Artemov and E. Nogina. Logic of knowledge with justifications from the provability perspective. Technical Report TR-2004011, CUNY Ph.D. Program in Computer Science, 2004.

[Artemov and Sidon-Yavorskaya, 2001] S. Artemov and T. Sidon-Yavorskaya. On the first order logic of proofs. *Moscow Mathematical Journal*, 1:475–490, 2001.

[Artemov and Strassen, 1992a] S. Artemov and T. Strassen. The basic logic of proofs. In E. Börger, G. Jäger, H. Kleine Büning, S. Martini, and M.M. Richter, editors, *Computer Science Logic. 6th Workshop, CSL'92. San Miniato, Italy, September/October 1992. Selected Papers*, volume 702 of *Lecture Notes in Computer Science*, pages 14–28. Springer, 1992.

[Artemov and Strassen, 1992b] S. Artemov and T. Strassen. Functionality in the basic logic of proofs. Technical Report IAM 92-004, Department of Computer Science, University of Bern, Switzerland, 1992.

[Artemov and Strassen, 1993] S. Artemov and T. Strassen. The logic of the Gödel proof predicate. In G. Gottlob, A. Leitsch, and D. Mundici, editors, *Computational Logic and Proof Theory. Third Kurt Gödel Colloquium, KGC'93. Brno, Chech Republic, August 1993. Proceedings*, volume 713 of *Lecture Notes in Computer Science*, pages 71–82. Springer, 1993.

[Artemov *et al.*, 1999] S. Artemov, E. Kazakov, and D. Shapiro. Epistemic logic with justifications. Technical Report CFIS 99-12, Cornell University, 1999.

[Artemov, 1979] S.N. Artemov. *Extensions of arithmetic and modal logics*. PhD thesis, Steklov Mathematical Insitute, Moscow, 1979. In Russian.

[Artemov, 1980] S.N. Artemov. Arithmetically complete modal theories. In *Semiotika i Informatika,* 14, pages 115–133. VINITI, Moscow, 1980. In Russian. English translation in: *Amer. Math. Soc. Transl.* (2), 135: 39–54, 1987.

[Artemov, 1982] S.N. Artemov. Applications of modal logic in proof theory. In *Problems of Cybernetics: Nonclassical logics and their applications*, pages 3–20. Nauka, Moscow, 1982. In Russian.

[Artemov, 1985a] S.N. Artemov. Nonarithmeticity of truth predicate logics of provability. *Doklady Akad. Nauk SSSR*, 284(2):270–271, 1985. In Russian. English translation in *Soviet Mathematics Doklady* 33:403–405, 1985.

[Artemov, 1985b] S.N. Artemov. On modal logics axiomatizing provability. *Izvestiya Akad. Nauk SSSR, ser. mat.*, 49(6):1123–1154, 1985. In Russian. English translation in: *Math. USSR Izvestiya* 27(3).

[Artemov, 1986] S.N. Artemov. Numerically correct provability logics. *Doklady Akad. Nauk SSSR*, 290(6):1289–1292, 1986. In Russian. English translation in *Soviet Mathematics Doklady* 34:384-387, 1987.

[Artemov, 1990] S. Artemov. Kolmogorov logic of problems and a provability interpretation of intuitionistic logic. In *Theoretical Aspects of Reasoning about Knowledge - III Proceedings*, pages 257–272. Morgan Kaufman Pbl., 1990.

[Artemov, 1994] S. Artemov. Logic of proofs. *Annals of Pure and Applied Logic*, 67(1):29–59, 1994.

[Artemov, 1995] S. Artemov. Operational modal logic. Technical Report MSI 95-29, Cornell University, 1995.

[Artemov, 1998] S. Artemov. Logic of proofs: a unified semantics for modality and λ-terms. Technical Report CFIS 98-06, Cornell University, 1998.

[Artemov, 1999] S. Artemov. On explicit reflection in theorem proving and formal verification. In *Automated Deduction - CADE-16. Proceedings of the 16th International Conference on Automated Deduction, Trento, Italy, July 1999*, volume 1632 of *Lecture Notes in Artificial Intelligence*, pages 267–281. Springer, 1999.

[Artemov, 2000] S. Artemov. Operations on proofs that can be specified by means of modal logic. In *Advances in Modal Logic. Volume 2*. CSLI Publications, Stanford University, 2000.

[Artemov, 2001] S. Artemov. Explicit provability and constructive semantics. *Bulletin of Symbolic Logic*, 7(1):1–36, 2001.

[Artemov, 2002] S. Artemov. Unified semantics for modality and λ-terms via proof polynomials. In K. Vermeulen and A. Copestake, editors, *Algebras, Diagrams and Decisions in Language, Logic and Computation*, pages 89–119. CSLI Publications, Stanford University, 2002.

[Artemov, 2004] S. Artemov. Kolmogorov and Gödel's approach to intuitionistic logic: current developments. *Russian Mathematical Surveys*, 59(2):203–229, 2004.

[Avigad and Feferman, 1998] J. Avigad and S. Feferman. Gödel's functional ("Dialectica") interpretation. In S. Buss, editor, *Handbook of Proof Theory*, pages 337–406. Elsevier, 1998.

[Avron, 1984] A. Avron. On modal systems having arithmetical interpretations. *The Journal of Symbolic Logic*, 49:935–942, 1984.

[Beeson, 1975] M. Beeson. The nonderivability in intuitionistic formal systems of theorems on the continuity of effective operations. *The Journal of Symbolic Logic*, 40:321–346, 1975.

[Beeson, 1980] M. Beeson. *Foundations of Constructive Mathematics*. Springer-Verlag, 1980.

[Beklemishev *et al.*, 1999] L.D. Beklemishev, M. Pentus, and N. Vereshchagin. Provability, complexity, grammars. *American Mathematical Society Translations, Series 2*, 192, 1999.

[Beklemishev, 1989a] L.D. Beklemishev. On the classification of propositional provability logics. *Izvestiya Akademii Nauk SSSR, ser. mat.*, 53(5):915–943, 1989. In Russian. English translation in *Math.USSR Izvestiya* 35 (1990) 247–275.

[Beklemishev, 1989b] L.D. Beklemishev. A provability logic without Craig's interpolation property. *Matematicheskie Zametki*, 45(6):12–22, 1989. In Russian. English translation in *Math. Notes* 45 (1989).

[Beklemishev, 1991] L.D. Beklemishev. Provability logics for natural Turing progressions of arithmetical theories. *Studia Logica*, 50(1):107–128, 1991.

[Beklemishev, 1992] L.D. Beklemishev. Independent numerations of theories and recursive progressions. *Sibirskii Matematichskii Zhurnal*, 33(5):22–46, 1992. In Russian. English translation in *Siberian Math. Journal*, 33 (1992).).

[Beklemishev, 1994] L.D. Beklemishev. On bimodal logics of provability. *Annals of Pure and Applied Logic*, 68(2):115–160, 1994.

[Beklemishev, 1996] L.D. Beklemishev. Bimodal logics for extensions of arithmetical theories. *The Journal of Symbolic Logic*, 61(1):91–124, 1996.

[Beklemishev, 1997a] L.D. Beklemishev. Induction rules, reflection principles, and provably recursive functions. *Annals of Pure and Applied Logic*, 85:193–242, 1997.

[Beklemishev, 1997b] L.D. Beklemishev. Notes on local reflection principles. *Theoria*, 63(3):139–146, 1997.

[Beklemishev, 1997c] L.D. Beklemishev. Parameter free induction and reflection. In G. Gottlob, A. Leitsch, and D. Mundici, editors, *Lecture Notes in Computer Science 1289. Computational Logic and Proof Theory, 5-th K.Gödel Colloquium KGC'97, Proceedings*, pages 103–113. Springer-Verlag, Berlin, 1997.

[Beklemishev, 1998a] L.D. Beklemishev. A proof-theoretic analysis of collection. *Archive for Mathematical Logic*, 37:275–296, 1998.

[Beklemishev, 1998b] L.D. Beklemishev. Reflection principles in formal arithmetic. Doctor of Sciences Dissertation, Steklov Math. Institute, Moscow. In Russian, 1998.

[Beklemishev, 1999a] L.D. Beklemishev. Open least element principle and bounded query computation. In J. Flum and M. Rodrigues-Artalejo, editors, *Lecture Notes in Computer Science 1683. Computer Science Logic, 13-th international workshop, CSL'99. Madrid, Spain, September 20–25, 1999. Proceedings*, pages 389–404. Springer-Verlag, Berlin, 1999.

[Beklemishev, 1999b] L.D. Beklemishev. Parameter free induction and provably total computable functions. *Theoretical Computer Science*, 224(1–2):13–33, 1999.

[Beklemishev, 2001] L.D. Beklemishev. Provability algebras and proof-theoretic ordinals, I. Logic Group Preprint Series 208, University of Utrecht, 2001. `http://preprints.phil.uu.nl/lgps/`.

[Beklemishev, 2003a] L.D. Beklemishev. Proof-theoretic analysis by iterated reflection. *Archive for Mathematical Logic*, 42:515–552, 2003. DOI: 10.1007/s00153-002-0158-7.

[Beklemishev, 2003b] L.D. Beklemishev. The Worm principle. Logic Group Preprint Series 219, University of Utrecht, 2003. `http://preprints.phil.uu.nl/lgps/`.

[Beklemishev, 2004] L.D. Beklemishev. Provability algebras and proof-theoretic ordinals, I. *Annals of Pure and Applied Logic*, 128:103–123, 2004.

[Bellin, 1985] G. Bellin. A system of natural deduction for GL. *Theoria*, 51:89–114, 1985.

[Berarducci and Verbrugge, 1993] A. Berarducci and R. Verbrugge. On the provability logic of bounded arithmetic. *Annals of Pure and Applied Logic*, 61:75–93, 1993.

[Berarducci, 1990] A. Berarducci. The interpretability logic of Peano Arithmetic. *The Journal of Symbolic Logic*, 55:1059–1089, 1990.

[Bernardi, 1976] C. Bernardi. The uniqueness of the fixed point in every diagonalizable algebra. *Studia Logica*, 35:335–343, 1976.

[Boolos and McGee, 1987] G. Boolos and V. McGee. The degree of the set of sentences of predicate provability logic that are true under every interpretation. *The Journal of Symbolic Logic*, 52(1):165–171, 1987.

[Boolos and Sambin, 1991] G. Boolos and G. Sambin. Provability: the emergence of a mathematical modality. *Studia Logica*, 50(1):1–23, 1991.

[Boolos, 1976] G. Boolos. On deciding the truth of certain statements involving the notion of consistency. *Journal of Symbolic Logic*, 41:779–781, 1976.

[Boolos, 1979a] G. Boolos. Reflection principles and iterated consistency assertions. *The Journal of Symbolic Logic*, 44:33–35, 1979.

[Boolos, 1979b] G. Boolos. *The Unprovability of Consistency: An Essay in Modal Logic.* Cambridge University Press, Cambridge, 1979.

[Boolos, 1980] G. Boolos. Omega-consistency and the diamond. *Studia Logica*, 39:237–243, 1980.

[Boolos, 1982] G. Boolos. Extremely undecidable sentences. *The Journal of Symbolic Logic*, 47:191–196, 1982.

[Boolos, 1993] G. Boolos. *The Logic of Provability.* Cambridge University Press, Cambridge, 1993.

[Brezhnev, 2000] V. Brezhnev. On explicit counterparts of modal logics. Technical Report CFIS 2000-05, Cornell University, 2000.

[Brezhnev, 2001] V. Brezhnev. On the logic of proofs. In *Proceedings of the Sixth ESSLLI Student Session, Helsinki*, pages 35–46, 2001. `http://www.helsinki.fi/esslli/`.

[Bull and Segerberg, 2001] R. Bull and K. Segerberg. Basic modal logic. In D. Gabbay and F. Guenthner, editors, *Handbook of Philosophical Logic, 2nd ed.*, volume 3, pages 1–83. Kluwer, Dordrecht, 2001.

[Burr, 2000] W. Burr. Fragments of Heyting arithmetic. *The Journal of Symbolic Logic*, 65(3):1223–1240, 2000.

[Buss, 1986] S. Buss. *Bounded arithmetic.* Bibliopolis, Napoli, 1986.

[Buss, 1990] S. Buss. The modal logic of pure provability. *Notre Dame Journal of Formal Logic*, 31(2):225–231, 1990.

[Buss, 1998] S.R. Buss. Introduction to Proof Theory. In S.R. Buss, editor, *Handbook of Proof Theory*, pages 1–78, Amsterdam, 1998. Elsevier, North-Holland.

[Carbone and Montagna, 1989] A. Carbone and F. Montagna. Rosser orderings in bimodal logics. *Zeitschrift f. math. Logik und Grundlagen d. Math.*, 35:343–358, 1989.

[Carbone and Montagna, 1990] A. Carbone and F. Montagna. Much shorter proofs: A bimodal investigation. *Zeitschrift f. math. Logik und Grundlagen d. Math.*, 36:47–66, 1990.

[Carlson, 1986] T. Carlson. Modal logics with several operators and provability interpretations. *Israel Journal of Mathematics*, 54:14–24, 1986.

[Chagrov and Zakharyaschev, 1997] A. Chagrov and M. Zakharyaschev. *Modal Logic.* Oxford Science Publications, 1997.

[Chagrov *et al.*, 2001] A.V. Chagrov, F. Wolter, and M. Zakhariashchev. Advanced modal logic. In D. Gabbay and F. Guenthner, editors, *Handbook of Philosophical Logic, 2nd ed.*, volume 3, pages 83–266. Kluwer, Dordrecht, 2001.

[Chagrov, 1985] A.V. Chagrov. On the complexity of propositional logics. In *Complexity problems in Mathematical Logic*, pages 80–90. Kalinin State University, Kalinin, 1985. In Russian.

[Constable, 1994] Robert L. Constable. Using reflection to explain and enhance type theory. In Helmut Schwichtenberg, editor, *Proof and Computation*, volume 139 of *NATO Advanced Study Institute, International Summer School held in Marktoberdorf, Germany, July 20-August 1, NATO Series F*, pages 65–100. Springer, Berlin, 1994.

[Constable, 1998] R. Constable. Types in logic, mathematics and programming. In S. Buss, editor, *Handbook of Proof Theory*, pages 683–786. Elsevier, 1998.

[Cutland, 1980] N. Cutland. *Computability. An introduction to recursive function theory.* Cambridge University Press, Cambridge, etc., 1980.

[Davis and Schwartz, 1979] M. Davis and J. Schwartz. Metamathematical extensibility for theorem verifiers and proof checkers. *Computers and Mathematics with Applications*, 5:217–230, 1979.

[de Jongh and Japaridze, 1998] D. de Jongh and G. Japaridze. The Logic of Provability. In S.R. Buss, editor, *Handbook of Proof Theory.* Studies in Logic and the Foundations of Mathematics, Vol.137., pages 475–546. Elsevier, Amsterdam, 1998.

[de Jongh and Montagna, 1989] D. de Jongh and F. Montagna. Much shorter proofs. *Z. Math. Logik Grundlagen Math.*, 35(3):247–260, 1989.

[de Jongh and Visser, 1996] D. de Jongh and A. Visser. Embeddings of Heyting algebras. In W. et al. Hodges, editor, *Logic: from foundations to applications. European Logic Colloquium, Keele, UK, July 20–29, 1993*, pages 187–213. Clarendon Press, Oxford, 1996.

[de Jongh, 1970] D. de Jongh. The maximality of the intuitionistic predicate calculus with respect to Heyting's Arithmetic. *The Journal of Symbolic Logic*, 36:606, 1970.

[Dzhaparidze, 1992] G. Dzhaparidze. The logic of linear tolerance. *Studia Logica*, 51(2):249–277, 1992.

[Dzhaparidze, 1993] G. Dzhaparidze. A generalized notion of weak interpretability and the corresponding modal logic. *Annals of Pure and Applied Logic*, 61(1-2):113–160, 1993.

[Fagin *et al.*, 1995] R. Fagin, J. Halpern, Y. Moses, and M. Vardi. *Reasoning About Knowledge.* MIT Press, 1995.

[Feferman, 1960] S. Feferman. Arithmetization of metamathematics in a general setting. *Fundamenta Mathematicae*, 49:35–92, 1960.

[Feferman, 1962] S. Feferman. Transfinite recursive progressions of axiomatic theories. *The Journal of Symbolic Logic*, 27:259–316, 1962.

[Fitting, 2003a] M. Fitting. A semantic proof of the realizability of modal logic in the logic of proofs. Technical Report TR-2003010, CUNY Ph.D. Program in Computer Science, 2003.

[Fitting, 2003b] M. Fitting. A semantics for the logic of proofs. Technical Report TR-2003012, CUNY Ph.D. Program in Computer Science, 2003.

[Fitting, 2004] M. Fitting. Semantics and tableaus for $\mathsf{LPS4}$. Technical Report TR-2004016, CUNY Ph.D. Program in Computer Science, 2004.

[Fitting, 2005] M. Fitting. The logic of proofs, semantically. To appear in *Annals of Pure and Applied Logic*. Available on `http://comet.lehman.cuny.edu/fitting`., 2005.

[Friedman, 1975a] H. Friedman. 102 problems in mathematical logic. *The Journal of Symbolic Logic*, 40:113–129, 1975.

[Friedman, 1975b] H. Friedman. The disjunction property implies the numerical existence property. *Proc. Nat. Acad. USA*, 72:2877–2878, 1975.

[Friedman, 1975c] H. Friedman. Some applications of Kleene's methods for intuitionistic systems. In A.R.D. Mathias and H. Rogers, editors, *Cambridge Summerschool in Mathematical Logic*, pages 113–170. Springer-Verlag, Berlin, 1975.

[Gabbay, 1994] D.M. Gabbay. *Labelled Deductive Systems.* Oxford University Press, 1994.

[Gavrilenko, 1981] Yu.V. Gavrilenko. Recursive realizability from the intuitionistic point of view. *Soviet Math. Doklady*, 23:9–14, 1981.

[Gentzen, 1936] G. Gentzen. Die Wiederspruchsfreiheit der reinen Zahlentheorie. *Math. Ann.*, 112:493–565, 1936.

[Gentzen, 1938] G. Gentzen. Neue Fassung des Wiederspruchsfreiheitsbeweises für die reine Zahlentheorie. *Forschungen zur Logik ung Grundlegung der exakten Wissenschaften*, 4:19–44, 1938.

[Ghilardi and Zawadowski, 1995] S. Ghilardi and M. Zawadowski. A sheaf representation and duality for finitely presented Heyting algebras. *The Journal of Symbolic Logic*, 60:911–939, 1995.

[Ghilardi, 1999] S. Ghilardi. Unification in intuitionistic logic. *The Journal of Symbolic Logic*, 64:859–880, 1999.

[Ghilardi, 2000] S. Ghilardi. Best solving modal equations. *Annals of Pure and Applied Logic*, 102:183–198, 2000.

[Girard *et al.*, 1989] J.-Y. Girard, Y. Lafont, and P. Taylor. *Proofs and Types.* Cambridge University Press, 1989.

[Gödel, 1933] K. Gödel. Eine Interpretation des intuitionistischen Aussagenkalkuls. *Ergebnisse Math. Kolloq.*, 4:39–40, 1933. English translation in: S. Feferman et al., editors, *Kurt Gödel Collected Works, Vol. 1*, pages 301–303. Oxford University Press, Oxford, Clarendon Press, New York, 1986.

[Gödel, 1995] K. Gödel. Vortrag bei Zilsel, 1938. In S. Feferman, editor, *Kurt Gödel Collected Works. Volume III*, pages 86–113. Oxford University Press, 1995.

[Goldblatt, 1978] R. Goldblatt. Arithmetical necessity, provability and intuitionistic logic. *Theoria*, 44:38–46, 1978.

[Goncharov, 1997] S.S. Goncharov. *Countable Boolean algebras and decidability, Siberian School of Algebra and Logic.* Plenum Press, New York, 1997. Russian original: Schetnye bulevy algebry i razreshimost'. Sibirskaya Shkola Algebry i Logiki. Novosibirsk: Nauchnaya Kniga. xii, 362 p. (1996).

[Goodman, 1970] N.D. Goodman. A theory of constructions is equivalent to arithmetic. In J. Myhill, A. Kino, and R.E. Vesley, editors, *Intuitionism and Proof Theory*, pages 101–120. North-Holland, 1970.

[Goryachev, 1986] S. Goryachev. On interpretability of some extensions of arithmetic. *Mat. Zametki*, 40:561–572, 1986. In Russian. English translation in *Math. Notes*, 40.

[Grigolia, 1987] R.Sh. Grigolia. *Free algebras of non-classical logics.* Metzniereba, Tbilissi, 1987. In Russian.

[Grzegorczyk, 1953] A. Grzegorczyk. Some classes of recursive functions. In *Rozprawy Matematiczne, IV.* Warszawa, 1953.

[Guaspari and Solovay, 1979] D. Guaspari and R. Solovay. Rosser sentences. *Annals of Math. Logic*, 16:81–99, 1979.

[Guaspari, 1979] D. Guaspari. Partially conservative sentences and interpretability. *Transactions of AMS*, 254:47–68, 1979.

[Hájek and Montagna, 1990] P. Hájek and F. Montagna. The logic of Π_1-conservativity. *Archive for Mathematical Logic*, 30(2):113–123, 1990.

[Hájek and Montagna, 1992] P. Hájek and F. Montagna. The logic of Π_1-conservativity continued. *Archive for Mathematical Logic*, 32:57–63, 1992.

[Hájek and Pudlák, 1993] P. Hájek and P. Pudlák. *Metamathematics of First Order Arithmetic.* Springer-Verlag, Berlin, Heidelberg, New York, 1993.

[Harrison, 1995] J. Harrison. Methatheory and reflection in theorem proving: A survey and critique. Technical report, University of Cambridge, 1995. URL http://www.dcs.glasgow.ac.uk/ tfm/hol-bib.html#H.

[Heyting, 1931] A. Heyting. Die intuitionistische grundlegung der mathematik. *Erkenntnis*, 2:106–115, 1931.

[Heyting, 1934] A. Heyting. *Mathematische Grundlagenforschung. Intuitionismus. Beweistheorie.* Springer, Berlin, 1934.

[Hilbert and Bernays, 1968] D. Hilbert and P. Bernays. *Grundlagen der Mathematik, Vols. I and II, 2d ed.* Springer-Verlag, Berlin, 1968.

[Iemhoff, 2001a] R. Iemhoff. A modal analysis of some principles of the provability logic of Heyting arithmetic. In *Advances in modal logic. Vol. 2. Selected papers from the 2nd international workshop (AiML'98), Uppsala, Sweden, October 16-18, 1998.* CSLI Lecture Notes 119, pages 301–336. CSLI Publications, Stanford, 2001.

[Iemhoff, 2001b] R. Iemhoff. On the admissible rules of intuitionistic propositional logic. *The Journal of Symbolic Logic*, 66(1):281–294, 2001.

[Iemhoff, 2001c] R. Iemhoff. *Provability logic and admissible rules.* PhD thesis, University of Amsterdam, Amsterdam, 2001.

[Ignatiev, 1991] K.N. Ignatiev. Partial conservativity and modal logics. ITLI Prepublication Series X–91–04, University of Amsterdam, 1991.

[Ignatiev, 1993a] K.N. Ignatiev. On strong provability predicates and the associated modal logics. *The Journal of Symbolic Logic*, 58:249–290, 1993.

[Ignatiev, 1993b] K.N. Ignatiev. The provability logic for Σ_1-interpolability. *Annals of Pure and Applied Logic*, 64:1–25, 1993.

[Japaridze, 1986] G.K. Japaridze. The modal logical means of investigation of provability. Thesis in Philosophy, in Russian, Moscow, 1986.

[Japaridze, 1988] G.K. Japaridze. The polymodal logic of provability. In *Intensional Logics and Logical Structure of Theories: Material from the fourth Soviet–Finnish Symposium on Logic, Telavi, May 20–24, 1985*, pages 16–48. Metsniereba, Tbilisi, 1988. In Russian.

[Japaridze, 1994] G. Japaridze. A simple proof of arithmetical completeness for Π_1-conservativity logic. *Notre Dame Journal of Formal Logic*, 35:346–354, 1994.

[Kaye *et al.*, 1988] R. Kaye, J. Paris, and C. Dimitracopoulos. On parameter free induction schemas. *The Journal of Symbolic Logic*, 53(4):1082–1097, 1988.

[Kirby and Paris, 1982] L.A.S. Kirby and J.B. Paris. Accessible independence results for Peano arithmetic. *Bull. London Math. Soc.*, 14:285–293, 1982.

[Kleene, 1945] S. Kleene. On the interpretation of intuitionistic number theory. *The Journal of Symbolic Logic*, 10(4):109–124, 1945.

[Kleene, 1952] S. Kleene. *Introduction to Metamathematics.* Van Norstrand, 1952.

[Kolmogoroff, 1932] A. Kolmogoroff. Zur Deutung der intuitionistischen logik. *Mathematische Zeitschrift*, 35:58–65, 1932. In German. English translation in V.M. Tikhomirov, editor, *Selected works of A.N. Kolmogorov. Volume I: Mathematics and Mechanics*, pages 151–158. Kluwer, Dordrecht 1991.

[Kolmogorov, 1985] A.N. Kolmogorov. About my papers on intuitionistic logic. In V.M. Tikhomirov, editor, *Selected works of A.N. Kolmogorov. Volume I: Mathematics and Mechanics*, page 393. Nauka, Moscow, 1985. In Russian, English translation in V.M. Tikhomirov, editor, *Selected works of A.N. Kolmogorov. Volume I: Mathematics and Mechanics*, pages 451–452. Kluwer, Dordrecht 1991.

[Kozen and Tiuryn, 1990] D. Kozen and J. Tiuryn. Logic of programs. In J. van Leeuwen, editor, *Handbook of Theoretical Computer Science. Volume B, Formal Models and Semantics*, pages 789–840. Elsevier, 1990.

[Kreisel and Lévy, 1968] G. Kreisel and A. Lévy. Reflection principles and their use for establishing the complexity of axiomatic systems. *Zeitschrift f. math. Logik und Grundlagen d. Math.*, 14:97–142, 1968.

[Kreisel, 1952] G. Kreisel. On the interpretation of non-finitist proofs, II. *The Journal of Symbolic Logic*, 17:43–58, 1952.

[Kreisel, 1962a] G. Kreisel. Foundations of intuitionistic logic. In E. Nagel, P. Suppes, and A. Tarski, editors, *Logic, Methodology and Philosophy of Science. Proceedings of the 1960 International Congress*, pages 198–210. Stanford University Press, 1962.

[Kreisel, 1962b] G. Kreisel. On weak completeness of intuitionistic predicate logic. *The Journal of Symbolic Logic*, 27:139–158, 1962.

[Kreisel, 1965] G. Kreisel. Mathematical logic. In T.L. Saaty, editor, *Lectures in Modern Mathematics III*, pages 95–195. Wiley and Sons, New York, 1965.

[Kripke, 1963] S. Kripke. Semantical considerations on modal logic. *Acta Philosophica Fennica*, 16:83–94, 1963.

[Krupski, 1997] V. Krupski. Operational logic of proofs with functionality condition on proof predicate. In S. Adian and A. Nerode, editors, *Logical Foundations of Computer Science' 97, Yaroslavl'*, volume 1234 of *Lecture Notes in Computer Science*, pages 167–177. Springer, 1997.

[Krupski, 2002] V. Krupski. The single-conclusion proof logic and inference rules specification. *Annals of Pure and Applied Logic*, 113(1-3):181–201, 2002.

[Krupski, 2005] V. Krupski. Referential logic of proofs. *Theoretical Computer Science*, (accepted), 2005.

[Krupski(jr.), 2003] N.V. Krupski(jr.). On the complexity of the reflected logic of proofs. Technical Report TR-2003007, CUNY Ph.D. Program in Computer Science, 2003.

[Kuznets, 2000] R. Kuznets. On the complexity of explicit modal logics. In *Computer Science Logic 2000*, volume 1862 of *Lecture Notes in Computer Science*, pages 371–383. Springer-Verlag, 2000.

[Kuznetsov and Muravitsky, 1976] A. Kuznetsov and A. Muravitsky. The logic of provability. In *Abstracts of the 4-th All-Union Conference on Mathematical Logic*, page 73, Kishinev, 1976. In Russian.

[Kuznetsov and Muravitsky, 1977] A.V. Kuznetsov and A.Yu. Muravitsky. Magari algebras. In *Fourteenth All-Union Algebra Conf., Abstract part 2: Rings, Algebraic Structures*, pages 105–106, 1977. In Russian.

[Kuznetsov and Muravitsky, 1986] A.V. Kuznetsov and A.Yu. Muravitsky. On superintuitionistic logics as fragments of proof logic. *Studia Logica*, XLV:76–99, 1986.

[Läuchli, 1970] H. Läuchli. An abstract notion of realizability for which intuitionistic predicate logic is complete. In J. Myhill, A. Kino, and R.E. Vesley, editors, *Intuitionism and Proof Theory*, pages 227–234. North-Holland, 1970.

[Leivant, 1981] D. Leivant. On the proof theory of the modal logic for arithmetic provability. *The Journal of Symbolic Logic*, 46:531–538, 1981.

[Leivant, 1983] D. Leivant. The optimality of induction as an axiomatization of arithmetic. *The Journal of Symbolic Logic*, 48:182–184, 1983.

[Lemmon, 1957] E. Lemmon. New foundations for Lewis's modal systems. *The Journal of Symbolic Logic*, 22:176–186, 1957.

[Lindström, 1984] P. Lindström. On partially conservative sentences and interpretability. *Proceedings of the AMS*, 91(3):436–443, 1984.

[Lindström, 1994] P. Lindström. *The modal logic of Parikh provability.* Tech. Rep. Filosofiska Meddelanden, Gröna Serien 5, Univ. Göteborg, 1994.

[Lindström, 1996] P. Lindström. Provability logic – a short introduction. *Theoria*, 62(1-2):19–61, 1996.

[Löb, 1955] M.H. Löb. Solution of a problem of Leon Henkin. *The Journal of Symbolic Logic*, 20:115–118, 1955.

[Magari, 1975a] R. Magari. The diagonalizable algebras (the algebraization of the theories which express Theor.:II). *Bollettino della Unione Matematica Italiana,* Serie 4, 12, 1975. Suppl. fasc. 3, 117–125.

[Magari, 1975b] R. Magari. Representation and duality theory for diagonalizable algebras (the algebraization of theories which express Theor.:IV). *Studia Logica*, 34:305–313, 1975.

[McCarthy, 2004] J. McCarthy. Notes on self-awareness. Internet posting by URL `http://www-formal.stanford.edu/jmc/selfaware/selfaware.html`, April 2004.

[McKinsey and Tarski, 1946] J.C.C. McKinsey and A. Tarski. On closed elements of closure algebras. *Annals of Mathematics*, 47:122–162, 1946.

[McKinsey and Tarski, 1948] J.C.C. McKinsey and A. Tarski. Some theorems about the sentential calculi of Lewis and Heyting. *The Journal of Symbolic Logic*, 13:1–15, 1948.

[Medvedev, 1962] Yu. Medvedev. Finite problems. *Soviet Mathematics Doklady*, 3:227–230, 1962.

[Meyer, 1975] A.R. Meyer. The inherent complexity of theories of ordered sets. In *Proceedings of the international congress of math., Vancouver, 1974*, pages 477–482. Canadian Math. Congress, 1975.

[Mints, 1971] G.E. Mints. Quantifier free and one quantifier systems. *Zapiski nauchnyh seminarov LOMI*, 20:115–133, 1971. In Russian.

[Mints, 1974] G. Mints. Lewis' systems and system T (a survey 1965-1973). In *Feys. Modal Logic (Russian translation)*, pages 422–509. Nauka, Moscow, 1974. In Russian, English translation in G. Mints, *Selected papers in proof theory*, Bibliopolis, Napoli, 1992.

[Mkrtychev, 1997] A. Mkrtychev. Models for the logic of proofs. In S. Adian and A. Nerode, editors, *Logical Foundations of Computer Science' 97, Yaroslavl'*, volume 1234 of *Lecture Notes in Computer Science*, pages 266–275. Springer, 1997.

[Montagna, 1978] F. Montagna. On the algebraization of a Feferman's predicate (the algebraiztion of theories which express Theor; X). *Studia Logica*, 37:221–236, 1978.

[Montagna, 1979] F. Montagna. On the diagonalizable algebra of Peano arithmetic. *Bollettino della Unione Matematica Italiana,* B (5), 16:795–812, 1979.

[Montagna, 1980] F. Montagna. Undecidability of the first order theory of diagonalizable algebras. *Studia Logica*, 39:347–354, 1980.

[Montagna, 1987a] F. Montagna. The predicate modal logic of provability. *Notre Dame Journal of Formal Logic*, 25:179–189, 1987.

[Montagna, 1987b] F. Montagna. Provability in finite subtheories of PA. *The Journal of Symbolic Logic*, 52(2):494–511, 1987.

[Montague, 1963] R. Montague. Syntactical treatments of modality with corollaries on reflection principles and finite axiomatizability. *Acta Philosophica Fennica*, 16:153–168, 1963.

[Moses, 1988] Y. Moses. Resource-bounded knowledge. In M. Vardi, editor, *Theoretical Aspects of Reasoning about Knowledge*, pages 261–276. Morgan Kaufman Pbl., 1988.

[Mostowski, 1953] A. Mostowski. On models of axiomatic systems. *Fundamenta Mathematicae*, 39:133–158, 1953.

[Myhill, 1960] J. Myhill. Some remarks on the notion of proof. *Journal of Philosophy*, 57:461–471, 1960.

[Myhill, 1985] J. Myhill. Intensional set theory. In S. Shapiro, editor, *Intensional Mathematics*, pages 47–61. North-Holland, 1985.

[Nogina, 1994] E. Nogina. Logic of proofs with the strong provability operator. Technical Report ILLC Prepublication Series ML-94-10, Institute for Logic, Language and Computation, University of Amsterdam, 1994.

[Nogina, 1996] E. Nogina. Grzegorczyk logic with arithmetical proof operators. *Fundamental and Applied Mathematics*, 2(2):483–499, 1996. In Russian, URL `http://mech.math.msu.su/~fpm/eng/96/962/96206.htm`.

[Ono, 1987] H. Ono. Reflection principles in fragments of Peano Arithmetic. *Zeitschrift f. math. Logik und Grundlagen d. Math.*, 33(4):317–333, 1987.

[Orevkov, 1979] V.P. Orevkov. Lower bounds for lengthening of proofs after cut-elimination. *Zapiski Nauchn. Semin. Leningr. Otd. Mat. Inst. Steklova*, 88:137–162, 1979. In Russian. English translation in: *Journal of Soviet Mathematics* 20, 2337-2350 (1982).

[Orlov, 1928] I.E. Orlov. The calculus of compatibility of propositions. *Matematicheskii Sbornik*, 35:263–286, 1928. In Russian.

[Parikh, 1971] R. Parikh. Existence and feasibility in arithmetic. *The Journal of Symbolic Logic*, 36:494–508, 1971.

[Parikh, 1987] R. Parikh. Knowledge and the problem of logical omniscience. In Z. Ras and M. Zemankova, editors, ISMIS-87 *(International Symposium on Methodolody for Intellectual Systems)*, pages 432–439. North-Holland, 1987.

[Parikh, 1995] R. Parikh. Logical omniscience. In D. Leivant, editor, *Logic and Computational Complexity*, pages 22–29. Springer Springer Lecture Notes in Computer Science No. 960, 1995.

[Parsons, 1970] C. Parsons. On a number-theoretic choice schema and its relation to induction. In A. Kino, J. Myhill, and R.E. Vessley, editors, *Intuitionism and Proof Theory*, pages 459–473. North Holland, Amsterdam, 1970.

[Parsons, 1972] C. Parsons. On n-quantifier induction. *The Journal of Symbolic Logic*, 37(3):466–482, 1972.

[Pitts, 1992] A. Pitts. On an interpretation of second-order quantification in first order intuitionistic propositional logic. *The Journal of Symbolic Logic*, 57:33–52, 1992.

[Plisko, 1977] V. Plisko. The nonarithmeticity of the class of realizable predicate formulas. *Soviet Mathematics Izvestia*, 11:453–471, 1977.

[Pohlers, 1998] W. Pohlers. Subsystems of set theory and second order number theory. In S.R. Buss, editor, *Handbook of Proof Theory*, pages 210–335. Elsevier, North-Holland, Amsterdam, 1998.

[Pour-El and Kripke, 1967] M.B. Pour-El and S. Kripke. Deduction-preserving "recursive isomorphisms" between theories. *Fundamenta Mathematicae*, 61:141–163, 1967.

[Rabin, 1961] M.O. Rabin. Non-standard models and independence of the induction axiom. In *Essays on the Foundations of Mathematics: Dedicated to A. Fraenkel on his 70th anniversary*, pages 287–299. North-Holland, Amsterdam, 1961.

[Rasiowa and Sikorski, 1963] H. Rasiowa and R. Sikorski. *The Mathematics of Metamathematics.* Polish Scientific Publishers, 1963.

[Ratajczyk, 1989] Z. Ratajczyk. Functions provably total in $I^{-}\Sigma_n$. *Fundamenta Mathematicae*, 133:81–95, 1989.

[Rathjen, 1994] M. Rathjen. Proof theory of reflection. *Annals of Pure and Applied Logic*, 68(2):181–224, 1994.

[Rathjen, 1999] M. Rathjen. The realm of ordinal analysis. In S.B. Cooper and J.K. Truss, editors, *Sets and proofs. London Math. Soc. Lect. Note Series 258*, pages 219–279. Cambridge University Press, Cambridge, 1999.

[Renne, 2004] B. Renne. Tableaux for the logic of proofs. Technical Report TR-2004001, CUNY Ph.D. Program in Computer Science, 2004.

[Rose, 1984] H.E. Rose. *Subrecursion: Functions and Hierarchies.* Clarendon Press, Oxford, 1984.

[Rosser, 1936] J.B. Rosser. Extensions of some theorems of Gödel and Church. *The Journal of Symbolic Logic*, 1:87–91, 1936.

[Rybakov, 1984] V.V. Rybakov. A criterion for admissibility of rules in the modal system $S4$ and intuitionistic logic. *Algebra and Logic*, 23:369–384, 1984.

[Rybakov, 1989] V.V. Rybakov. On admissibility of the inference rules in the modal system G. In Yu.L. Ershov, editor, *Trudy instituta matematiki*, volume 12, pages 120–138. Nauka, Novosibirsk, 1989. In Russian.

[Rybakov, 1997] V.V. Rybakov. *Admissibility of logical inference rules.* Elsevier, Amsterdam, 1997.

[Ryll-Nardzewski, 1953] C. Ryll-Nardzewski. The role of the axiom of induction in elementary arithmetic. *Fundamenta Mathematicae*, 39:239–263, 1953.

[Sambin and Valentini, 1982] G. Sambin and S. Valentini. The modal logic of provability, the sequential approach. *Journal of Philosophic Logic*, 11:311–342, 1982.

[Sambin and Valentini, 1983] G. Sambin and S. Valentini. The modal logic of provability: cut-elimination. *Journal of Philosophic Logic*, 12:471–476, 1983.

[Schmerl, 1979] U.R. Schmerl. A fine structure generated by reflection formulas over Primitive Recursive Arithmetic. In M. Boffa, D. van Dalen, and K. McAloon, editors, *Logic Colloquium'78*, pages 335–350. North Holland, Amsterdam, 1979.

[Segerberg, 1971] K. Segerberg. *An essay in classical modal logic.* Filosofiska Föreningen och Filosofiska Institutionen vid Uppsala Universitet, Uppsala, 1971.

[Shapiro, 1985a] S. Shapiro. Epistemic and intuitionistic arithmetic. In S. Shapiro, editor, *Intensional Mathematics*, pages 11–46. North-Holland, 1985.

[Shapiro, 1985b] S. Shapiro. Intensional mathematics and constructive mathematics. In S. Shapiro, editor, *Intensional Mathematics*, pages 1–10. North-Holland, 1985.

[Shavrukov, 1988] V.Yu. Shavrukov. The logic of relative interpretability over Peano arithmetic. Preprint, Steklov Mathematical Institute, Moscow, 1988. In Russian.

[Shavrukov, 1991] V.Yu. Shavrukov. On Rosser's provability predicate. *Zeitschrift f. math. Logik und Grundlagen d. Math.*, 37:317–330, 1991.

[Shavrukov, 1993a] V.Yu. Shavrukov. A note on the diagonalizable algebras of PA and ZF. *Annals of Pure and Applied Logic*, 61:161–173, 1993.

[Shavrukov, 1993b] V.Yu. Shavrukov. Subalgebras of diagonalizable algebras of theories containing arithmetic. *Dissertationes Mathematicae*, 323, 1993.

[Shavrukov, 1994] V.Yu. Shavrukov. A smart child of Peano's. *Notre Dame Journal of Formal Logic*, 35:161–185, 1994.

[Shavrukov, 1997a] V.Yu. Shavrukov. Isomorphisms of diagonalizable algebras. *Theoria*, 63(3):210–221, 1997.

[Shavrukov, 1997b] V.Yu. Shavrukov. Undecidability in diagonalizable algebras. *The Journal of Symbolic Logic*, 62(1):79–116, 1997.

[Sidon, 1997] T. Sidon. Provability logic with operations on proofs. In S. Adian and A. Nerode, editors, *Logical Foundations of Computer Science' 97, Yaroslavl'*, volume 1234 of *Lecture Notes in Computer Science*, pages 342–353. Springer, 1997.

[Sidon, 1998] T.L. Sidon. Craig interpolation property for operational logics of proofs. *Vestnik Moskovskogo Universiteta. Ser. 1 Mat., Mech.*, (2):34–38, 1998. In Russian. English translation in: *Moscow University Mathematics Bulletin*, v.53, n.2, pp.37–41, 1999.

[Simmons, 1988] H. Simmons. Large discrete parts of the E-tree. *The Journal of Symbolic Logic*, 53:980–984, 1988.

[Smiley, 1963] T. Smiley. The logical basis of ethics. *Acta Philosophica Fennica*, 16:237–246, 1963.

[Smoryński, 1973] C. Smoryński. Applications of Kripke models. In A. Troelstra, editor, *Metamathematical investigations of intuitionistic arithmetic and analysis.* Springer Lecture Notes 344, pages 324–391. Springer, Berlin, 1973.

[Smoryński, 1977a] C. Smoryński. ω-consistency and reflection. In *Colloque International de Logique (Colloq. Int. CNRS)*, pages 167–181. CNRS Inst. B. Pascal, Paris, 1977.

[Smoryński, 1977b] C. Smoryński. The incompleteness theorems. In J. Barwise, editor, *Handbook of Mathematical Logic*, pages 821–865. North Holland, Amsterdam, 1977.

[Smoryński, 1978] C. Smoryński. Beth's theorem and self-referential sentences. In A. Macintyre et al., editor, *Logic Colloquium'77.* North Holland, Amsterdam, 1978.

[Smoryński, 1981] C. Smoryński. Fifty years of self-reference. *Notre Dame Journal of Formal Logic*, 22:357–374, 1981.

[Smoryński, 1982] C. Smoryński. The finite inseparability of the first order theory of diagonalizable algebras. *Studia Logica*, 41:347–349, 1982.

[Smoryński, 1985] C. Smoryński. *Self-Reference and Modal Logic.* Springer-Verlag, Berlin, 1985.

[Smoryński, 1989] C. Smoryński. Arithmetical analogues of McAloon's unique Rosser sentences. *Archive for Mathematical Logic*, 28:1–21, 1989.

[Smoryński, 2004] C. Smoryński. Modal logic and self-reference. In D. Gabbay and F. Guenthner, editors, *Handbook of Philosophical Logic, 2nd ed.*, volume 11, page ?? Springer, Berlin, 2004.

[Solovay, 1976] R.M. Solovay. Provability interpretations of modal logic. *Israel Journal of Mathematics*, 28:33–71, 1976.

[Statman, 1978] R. Statman. Bounds for proof-search and speed-up in the predicate calculus. *Annals of Mathematical Logic*, 15:225–287, 1978.

[Takeuti, 1975] G. Takeuti. *Proof Theory.* North-Holland, 1975.

[Troelstra and Schwichtenberg, 1996] A. Troelstra and H. Schwichtenberg. *Basic Proof Theory.* Cambridge University Press, Amsterdam, 1996.

[Troelstra and van Dalen, 1988] A. Troelstra and D. van Dalen. *Constructivism in Mathematics, vols 1, 2.* North–Holland, Amsterdam, 1988.

[Troelstra, 1973] A. Troelstra. *Metamathematical investigations of intuitionistic arithmetic and analysis.* Springer Lecture Notes 344. Springer-Verlag, Berlin, 1973.

[Troelstra, 1998] A.S. Troelstra. Realizability. In S. Buss, editor, *Handbook of Proof Theory*, pages 407–474. Elsevier, 1998.

[Turing, 1939] A.M. Turing. System of logics based on ordinals. *Proc. London Math. Soc.*, ser. 2, 45:161–228, 1939.

[Uspensky and Plisko, 1985] V. Uspensky and V. Plisko. Intuitionistic Logic. Commentary on [Kolmogoroff, 1932] and [Kolmogorov, 1985]. In V.M. Tikhomirov, editor, *Selected works of A.N. Kolmogorov. Volume I: Mathematics and Mechanics*, pages 394–404. Nauka, Moscow, 1985. In Russian, English translation in V.M. Tikhomirov, editor, *Selected works of A.N. Kolmogorov. Volume I: Mathematics and Mechanics*, pages 452–466. Kluwer, Dordrecht 1991.

[Uspensky, 1992] V.A. Uspensky. Kolmogorov and mathematical logic. *Journal of Symbolic Logic*, 57(2):385–412, 1992.

[van Benthem, 1991] J. van Benthem. Reflections on epistemic logic. *Logique & Analyse*, 133-134:5–14, 1991.

[van Benthem, 2001] J. van Benthem. Correspondence theory. In D. Gabbay and F. Guenthner, editors, *Handbook of Philosophical Logic, 2nd ed.*, volume 3, pages 325–408. Kluwer, Dordrecht, 2001.

[van Dalen, 1986] D. van Dalen. Intuitionistic logic. In D. Gabbay and F. Guenther, editors, *Handbook of Philosophical Logic. Volume 3*, pages 225–340. Reidel, 1986.

[van Dalen, 1994] D. van Dalen. *Logic and Structure*. Springer-Verlag, 1994.

[Vardanyan, 1986] V.A. Vardanyan. Arithmetic comlexity of predicate logics of provability and their fragments. *Doklady Akad. Nauk SSSR*, 288(1):11–14, 1986. In Russian. English translation in *Soviet Mathematics Doklady* 33:569–572, 1986.

[Visser *et al.*, 1995] A. Visser, J. van Benthem, D. de Jongh, and G. Renardel de Lavalette. NNIL, a study in intuitionistic propositional logic. In A. Ponse, M. de Rijke, and Y. Venema, editors, *Modal Logic and Process Algebra, a bisimulation perspective. CSLI Lecture Notes, 53*, pages 289–326. CSLI Publications, Stanford, 1995.

[Visser, 1980] A. Visser. Numerations, λ-calculus and arithmetic. In J.P. Seldin and J.R. Hindley, editors, *To H.B. Curry. Essays on combinatory logic, lambda-calculus and formalism*, pages 259–284. Academic Press, London, 1980.

[Visser, 1981] A. Visser. *Aspects of Diagonalization and Provability.* PhD thesis, University of Utrecht, Utrecht, The Netherlands, 1981.

[Visser, 1982] A. Visser. On the completeness principle. *Annals of Mathematical Logic*, 22:263–295, 1982.

[Visser, 1984] A. Visser. The provability logics of recursively enumerable theories extending Peano Arithmetic at arbitrary theories extending Peano Arithmetic. *Journal of Philosophic Logic*, 13:97–113, 1984.

[Visser, 1985] A. Visser. Evaluation, provably deductive equivalence in Heyting arithmetic of substitution instances of propositional formulas. Logic Group Preprint Series 4, Department of Philosophy, University of Utrecht, 1985.

[Visser, 1989] A. Visser. Peano's smart children. A provability logical study of systems with built-in consistency. *Notre Dame Journal of Formal Logic*, 30:161–196, 1989.

[Visser, 1990] A. Visser. Interpretability logic. In P.P. Petkov, editor, *Mathematical Logic*, pages 175–208. Plenum Press, New York, 1990.

[Visser, 1991] A. Visser. The formalization of interpretability. *Studia Logica*, 50(1):81–106, 1991.

[Visser, 1992] A. Visser. An inside view of EXP. the closed fragment of the provability logic of $I\Delta_0 + \Omega_1$ with a propositional constant for EXP. *The Journal of Symbolic Logic*, 57(1):131–165, 1992.

[Visser, 1994] A. Visser. Propositional combinations of Σ_1-sentences in Heyting's arithmetic. Logic Group Preprint Series 117, Department of Philosophy, University of Utrecht, 1994.

[Visser, 1995] A. Visser. A course in bimodal provability logic. *Annals of Pure and Applied Logic*, 73:109–142, 1995.

[Visser, 1996] A. Visser. Uniform interpolation and layered bisimulation. In P. Hájek, editor, Lecture Notes in Logic 6. *Logical foundations of Mathematics, Computer Science and Physics – Kurt Gödel's Legacy, Gödel '96, Brno, Chech Republic, Proceedings*, pages 139–164. Springer-Verlag, Berlin, 1996.

[Visser, 1998] A. Visser. An overview of interpretability logic. In M. Kracht, M. de Rijke, H. Wansing, and M. Zakhariaschev, editors, *Advances in Modal Logic, v.1, CSLI Lecture Notes, No. 87*, pages 307–359. CSLI Publications, Stanford, 1998.

[Visser, 1999] A. Visser. Rules and arithmetics. *Notre Dame Journal of Formal Logic*, 40(1):116–140, 1999.

[Visser, 2002a] A. Visser. Faith and falsity. Logic Group Preprint Series 216, Department of Philosophy, University of Utrecht, 2002.

[Visser, 2002b] A. Visser. Substitutions of Σ^0_1-sentences: Explorations between intuitionistic propositional logic and intuitionistic arithmetic. *Annals of Pure and Applied Logic*, 114(1–3):227–271, 2002.

[Švejdar, 2003] V. Švejdar. The decision problem of provability logic with only one atom. *Archive for Mathematical Logic*, 42(8):763–768, 2003.

[Weinstein, 1983] S. Weinstein. The intended interpretation of intuitionistic logic. *Journal of Philosophical Logic*, 12:261–270, 1983.

[Wickline *et al.*, 1998] P. Wickline, P. Lee, F. Pfenning, and R. Davies. Modal types as staging specifications for run-time code generation. *ACM Computing Surveys*, 30(3es), 1998.

[Wilkie and Paris, 1987] A. Wilkie and J. Paris. On the scheme of induction for bounded arithmetic formulas. *Annals of Pure and Applied Logic*, 35:261–302, 1987.

[Wolter, 1998] F. Wolter. All finitely axiomatizable subframe logics containing CSM are decidable. *Archive for Mathematical Logic*, 37:167–182, 1998.

[Yavorskaya (Sidon), 2002] T. Yavorskaya (Sidon). Logic of proofs and provability. *Annals of Pure and Applied Logic*, 113(1-3):345–372, 2002.

[Yavorsky, 2000] R. Yavorsky. On the logic of the standard proof predicate. In *Computer Science Logic 2000*, volume 1862 of *Lecture Notes in Computer Science*, pages 527–541. Springer, 2000.

[Yavorsky, 2002] R. Yavorsky. Provability logics with quantifiers on proofs. *Annals of Pure and Applied Logic*, 113(1-3):373–387, 2002.

[Zambella, 1994] D. Zambella. Shavrukov's theorem on the subalgebras of diagonalizable algebras for theories containing $I\Delta_0 + \exp$. *Notre Dame Journal of Formal Logic*, 35:147–157, 1994.