

# Problems in the Logic of Provability

Lev D. Beklemishev\* & Albert Visser

May 10, 2005

## Abstract

In the first part of the paper we discuss some conceptual problems related to the notion of proof. In the second part we survey five major open problems in Provability Logic as well as possible directions for future research in this area.

## 1 Introduction

Provability logic was conceived by Kurt Gödel in 1933 [42], but it really took off in the seventies as a study of modal logics with provability interpretations. After about thirty years of fruitful development this area now finds itself in a transitional period. On the one hand, many of the problems originally perceived by the founders of the discipline have been successfully solved. On the other hand, new challenges are coming from the other areas of Logic and Computer Science. Presently, provability logic starts crossing the original borders of its domain and expands in several novel directions.

The original motivation for the study of provability as a modality was mainly philosophical in nature.<sup>1</sup> Provability logic for the first time provided a mathematically robust intended semantics of modality. Traditionally, modal logics were used to explicate informal and sometimes inherently vague notions such as necessity, belief, obligation, etc., often plagued by paradoxes. These logics were then provided with formal Kripke-style semantics along the chain

*intended semantics — logic axioms — Kripke semantics,*

---

\*Supported by the Russian Foundation for Basic Research and the Council for Support of Leading Scientific Schools.

<sup>1</sup>Therefore, it is not surprising that the axioms of provability logic first appeared in a treatise on ethics by T. Smiley [79].

in which only the second link was robust.<sup>2</sup>

In contrast, provability is a notion for which there is a formal as well as an informal understanding. Moreover, there is a widely accepted belief that the formal notion of provability in some respects adequately represents the informal one.<sup>3</sup> Therefore, philosophers such as Willard Van Orman Quine, Saul Kripke and George Boolos were for the first time in a position to study a modality for which all the three links above could be subject to a rigorous mathematical analysis.<sup>4</sup>

Very soon some mathematical logicians, with their own set of concerns, working in the USA, the Netherlands, Italy and the Soviet Union, started to get interested in the topic. They saw the potential usefulness of provability logic as a tool in the study of formal axiomatic theories. Perhaps the earliest example of such an application was the so-called de Jongh–Sambin fixed point theorem, which clarified the reason why certain fixed point equations in arithmetic had explicit and unique solutions. Some other applications followed (see [4, 82, 45]). However, most of the emphasis in the study of provability logics from the late 70s until the late 90s was on the arithmetical completeness results.

Robert Solovay proved, in a famous paper from 1976 [83], that the propositional Gödel–Löb logic **GL** is complete w.r.t. the provability semantics. Subsequent research mainly concerned with the possibility of extending Solovay’s theorems to more expressive languages, such as the language of predicate logic, propositional language with several modalities, the language of interpretability logic, second order propositional logic language, and some others. In this way, the natural borders of the discipline were roughly mapped. By now we have a reasonably good idea which of the above mentioned logics with provability semantics are manageable (are decidable, have nice axiomatizations, models, etc.) and which are not.<sup>5</sup> By the late 90s the time for autonomous development of provability logic was over. The time has come for new challenges and search for new applications.

Two such new directions of research emerged in the recent years. The first one is the so-called *logic of proofs* initiated by S. Artemov about 1994 and which since then has grown into a lively area. It was inspired by foundational concerns and the question of providing intuitionistic logic with a

---

<sup>2</sup>Sometimes also a direct link between intended semantics and Kripke semantics was established, but it had to be informal as well.

<sup>3</sup>For a more detailed discussion of this claim see below.

<sup>4</sup>This analysis was particularly important for discussing the semantics of quantification in modal logic.

<sup>5</sup>There are some notable open questions, though, which are discussed below.

robust provability semantics in the spirit of Brouwer–Heyting–Kolmogorov interpretation.<sup>6</sup> The logic of proofs also has connections with several important topics in Computer Science such as lambda calculus, logics of knowledge and belief, and formal verification. Discussing numerous open questions in this interesting area would exceed the limits of the present paper. However, we refer the reader to [3, 6] for a comprehensive exposition and a recent survey. A list of current problems in this field can be found at the homepage of S. Artemov.<sup>7</sup>

The second new development, the theory of the so-called *graded provability algebras* [11], aims to establish links and find applications of provability logic in the mainstream proof theory tradition. Graded provability algebras reveal surprising connections between provability logic and ordinal notation systems and provide a technically simple and clean treatment of proof-theoretic results such as consistency proofs, combinatorial independent principles, etc.

In this paper we want to present some open questions in this new developing area, as well as to record some long standing problems left in traditional provability logic.

In the first, mainly philosophical, part of the paper we discuss conceptual problems related to the notion of proof, in particular the relation between formal and informal proofs. We believe that these questions are very important and interesting in their own right, and provability logic may contribute to their study in the future.

In the second, mathematical, part we formulate five major open problems left in the area of traditional provability logic and discuss some related questions. These five problems are:

- (i) Provability logic of intuitionistic arithmetic.
- (ii) Provability logic of bounded arithmetic.
- (iii) Classification of bimodal provability logics.
- (iv) Decidability of the  $\forall^*\exists^*$ -fragment of the first order theory of Magari algebra of Peano arithmetic.
- (v) Interpretability logic of all reasonable theories.

---

<sup>6</sup>This problem was the main motivation for Gödel in 1933 to introduce a calculus of provability in the first place.

<sup>7</sup>URL <http://www.cs.gc.cuny.edu/~sartemov/>.

In the last section of the paper we present current problems in the area of graded provability algebras.

A compound list of the problems discussed in this paper is given in the Appendix. An online version with some additional questions is maintained by the first author.<sup>8</sup>

The authors would like to thank Sergei Artemov, Yuri Gurevich, Joost Joosten and Rostik Yavorsky and for useful discussions and comments.

## 2 Informal concepts of proof

The role of provability logic and its relationship with the other parts of proof theory are best to be explained by first discussing the general relations between formal and informal proofs. Our discussion will be of necessity one-sided: we concentrate on the phenomenology (representations) of proofs, but altogether ignore the questions such as validity (see e.g. [84] for a discussion). We are mainly interested in the aspects where a provability logic approach could be relevant.

### 2.1 Formal and informal provability and the problem of equivalence of proofs

The notion of axiomatic system and the associated formal notion of proof emerged at the beginning of 20-th century in the hands of G. Frege, G. Peano, and D. Hilbert. Behind these notions there was an implicitly accepted thesis, emphasized by D. Hilbert: every sufficiently developed area of mathematics (and perhaps not only of mathematics) can be axiomatized. Hence, every valid mathematical argument can be faithfully represented in a suitable formal axiomatic system.

The status, as well as the spirit, of this statement is similar to that of the Church–Turing Thesis in the theory of computation: it is a conjecture relating a robust mathematical notion (formal proof) and informal one (informal proof) that can only be verified by practice. Soon enough universal ‘in practice’ axiomatic systems, such as Zermelo–Fraenkel set theory, were formulated, which were apparently sufficient to formalize all current mathematics. Of course, the universality of these axiom systems had to be later qualified by Gödel: no single axiom system can be universal, in an absolute

---

<sup>8</sup>URL <http://www.phil.uu.nl/~lev/>.

sense.<sup>9</sup> From this point on, proof theory developed in the course of the 20-th century as a study of sufficiently universal axiomatic systems and the associated concepts of proof.

What is apparent from the historical perspective is a predominant interest in one particular model of proofs (aka deductive axiomatic proofs). Of course, this is almost the only model which had a clear mathematical formulation. However, one should not forget that deductive axiomatic proofs are not the only kind of proofs around. Compare, e.g., the commonsense notion of proof in natural sciences, with experiment as a major means of proof, or a well-developed concept of proof in jurisprudence having many non-traditional features such as defeasibility (see e.g. [66]). In contrast with these important but informal concepts of proof, probabilistic proof-like concepts encountered, in particular, in cryptography (see [44, 43]), do have rigorous models.<sup>10</sup>

On the other hand, even if one restricts attention to standard deductive axiomatic proofs only, there is a notable discrepancy between conventional, informal mathematical proofs and their formalized representations. Fully formalized proofs have become a reality with the advent of automatic provers and interactive proof assistants such as `NuPr1` and `Coq`. The difference with informal proofs becomes evident if one compares, say, a textbook proof of the main theorem of algebra with its formalization (proof-object) in `Coq`. The appearance, as well as the possible uses, of both proofs are quite different. In what sense are they actually the same?

The situation is analogous with the one around the Church–Turing Thesis: there is a difference between the high-level notion of algorithm and the low-level notion of program code (or Turing machine). Therefore, both in proof theory and in computer science the problem of equivalence of proofs (respectively, of programs) arises:

**Problem 1** *Which derivations/programs are essentially the same, that is, represent the same informal proof (respectively, algorithm)?*

---

<sup>9</sup>This contrasts with the existence of truly universal models of computation, such as Turing machines.

<sup>10</sup>There have also been some discussions within the logic community of the so-called ‘visual’ proofs in geometry. It can be argued that proofs directly appealing to visual intuition form a separate class of proofs. However, it is worth remembering that getting rid of such ‘visual’ intuitions was one of the main purposes of Hilbert’s program of axiomatizing geometry, which made a strong impression on his contemporaries. There also were various interesting attempts to look at some visual kind of proofs from the point of view of logic, see e.g. [7, 68].

Needless to say, this question is a notoriously difficult one and it may not have a unique answer. The problem of equivalence of proofs is known in proof theory for quite some time, and was advocated by Georg Kreisel and Dag Prawitz (see [60, 67]). However, remarkably little has been done on it — partly because it fell, and still falls, outside the mainstream proof theory, partly because it is a conceptual rather than a strictly mathematical problem. See Došen [29] for an interesting recent discussion.

We believe that, in the coming years, the importance of this and related questions will become more obvious to the wider community of logicians under the influences coming from Computer Science and the development of the automated deduction systems.

In principle there are two possible approaches to this problem, which we can call *bottom-up* and *top-down*. The bottom-up approach starts with a low-level notion of proof, tries to obtain more canonical representation of such proofs and looks for meaningful equivalence relations.

This is how this problem is usually perceived within the context of structural proof theory and within a related categorial proof theory approach. Attempts were made to find mathematically attractive and sufficiently broad equivalence relations on formal proofs. The first significant contribution to this problem came from Prawitz [67] who isolated the following notion of equivalence: two (natural deduction style) proofs are equivalent, if they normalize to the same proof. This equivalence relation is certainly very interesting, however it behaves well only for rather restricted kinds of formalisms. Already for classical propositional logic it does not really work — in fact, it identifies all proofs of a contradiction from a given hypothesis (see [29]).

Since the 70's the structural proof theory underwent a rapid development with the popularization of proofs-as-programs paradigm [41], Girard's linear logic [38], game semantics [2] and culminating in Girard's *ludics* [40]. It certainly falls outside the scope of this paper to discuss possible bearings of these broad doctrines on the problem of equivalence of proofs.

Instead we would like to concentrate on the opposite, top-down approach, which is how the question “What is an algorithm?” was handled with some success in Computer Science. Following the top-down methodology, researchers formulated more and more general models of computational processes, so that to make these abstract descriptions eventually fit the desired intuitive concept of algorithm.

## 2.2 Strengthening Hilbert’s Thesis

It is instructive, from the point of view of the above mentioned problem in proof theory, to gain some wisdom from the debate around Church–Turing Thesis in the theory of computation.<sup>11</sup> Some relevant issues have been recently raised by Yuri Gurevich in connection with his Abstract State Machines (ASM) (see e.g. [19, 18]) and Yiannis Moschovakis in connection with general recursive algorithms (see [63]).

Although according to the Church–Turing Thesis every computable function can be represented by a Turing machine, such a representation will not in general be faithful w.r.t. the algorithm’s data and elementary steps, which, from the point of view of computational practice, is a major drawback.<sup>12</sup> In contrast, Blass and Gurevich in [19] and elsewhere convincingly argue that every algorithm can be faithfully represented *on its own level of abstraction* by a suitable ASM. Of course, this does not yet settle the equivalence of programs problem — the question is simply being translated into a similar one about ASMs. However, it reveals additional information hidden in the informal notion of algorithm, such as its ‘abstraction level’, which narrows the gap between the algorithm’s formal and informal presentations.<sup>13</sup>

Gurevich’s Thesis, as opposed to the Church–Turing Thesis, is noticeably non-uniform: there cannot be a single ASM which could simulate any algorithm on its own level of abstraction. This is the price we pay for representing the algorithms more faithfully. Again, the situation is parallel to the non-uniform version of Hilbert’s Thesis — stating that every proof can be represented in a suitable axiomatic system — as opposed to a uniform version related to, say, a fixed system of set theory ZFC. Uniformity presupposes some kind of coding.

A close connection between Gurevich strengthening of the Church–Turing Thesis and what we dubbed Hilbert’s thesis is not really accidental. They both rely on the same basic presupposition that *the data for algorithms, as well as for mathematical proofs, are faithfully representable by first order logic structures*. Blass and Gurevich [19] call this claim (for the case of algorithms) the *abstract state postulate*.

---

<sup>11</sup>The fact that there are challenges to the Church–Turing Thesis from various directions, including e.g. physics, shows a healthy attitude developed in Computer Science towards these matters.

<sup>12</sup>Some early investigations in computation theory dealt with attempts to challenge the Church–Turing Thesis by inventing computing devices working with more complex kinds of data such as labeled complexes (Kolmogorov–Uspensky machines [58], Schönhage storage modification machines [73]).

<sup>13</sup>What exactly is a level of abstraction remains a bit unclear. The idea is intuitive.

A further issue addressed by the ASM approach is what kind of action constitutes a possible computation step (with roughly the answer: almost anything goes). A parallel question, what constitutes an admissible proof step, received some attention in proof theory. The initial answer — only logical inference rules modus ponens and generalization are sufficient — is not really satisfactory as these steps are too restricted.

One of the reasons for G. Gentzen to introduce his *natural deduction* proof system was to provide a model that better fitted the actual (‘natural’) form of mathematical arguments. However, despite the fundamental significance and various useful applications of this approach, that particular goal was not really achieved. Partly this is due to the fact that notions such as abstraction levels (of proofs) did not play any role in his analysis. He might have uncovered a natural form of logical steps for purely propositional (in a sense, lowest level) proofs, though.

Our next problem asks whether an analog of Gurevich Thesis holds for proof systems.

**Problem 2** *Find a reasonable proof system, or a class of proof systems, such that every (informal) mathematical proof admits a faithful formalization on its own level of abstraction and complexity in a suitable system from that class.*

We think that a positive answer to this question would be a significant step towards a better understanding of the problem of equivalence of proofs and similar questions relating formal and informal notions of proof.<sup>14</sup>

A priori it is not really clear if such a system can be formulated. The ASM approach showed its practical usefulness for the task of program specification. Similarly, what we are looking for is a language suitable for *proof specification*, with formalized proofs playing the role of implementations of informal proofs. To have convenient proof-specification tools based on clear principles is of obvious importance for the development of automated deduction systems.

### 2.3 Coordinate-free proof theory

As we have indicated before, standard proof systems formalizing logic, be it Hilbert-style, Gentzen sequent-style or natural deduction-style, suffer from

---

<sup>14</sup>According to R. Thiele (“Hilbert’s Twenty-Fourth Problem”, American Mathematical Monthly, 110(1), 1–24) Hilbert wanted to formulate the following problem as Problem 24 in his famous list: what is the simplest possible proof of a theorem? This question was found in his notes but never made it to the final list.



the same drawback: they are in a sense too concrete, that is, they depend on a lot of arbitrary, and irrelevant — from the point of view of the informal proof — details of syntax. This makes methods of structural proof theory — the part of proof theory dealing with the study of concrete proof systems by essentially syntactical methods — mathematically inelegant and *non-modular* (see similar remarks, e.g., in [39], although developments in linear logic, such as proof nets, do offer a real improvement in terms of elegance).

Modularity in this context means the ability to fruitfully apply a single proof-theoretic result to various formalisms. De facto, even such a basic result as the cut-elimination theorem has to be proved for every formalism anew, even if the modifications are relatively minor. In this sense, cut-elimination plays the role of a useful method rather than that of a single important result. Stating such a general cut-elimination result seems to be quite difficult, for it presupposes that one is able to formulate some kind of general conditions under which ‘cut-elimination’ holds. Such conditions are very elusive: meaningful classifications of syntactic formalisms have not been really developed.

The problem of non-canonicity of syntax and related non-modularity problem are major methodological drawbacks of current structural proof theory — they make the methods unattractive which ultimately results in technical difficulties and lack of progress. It might be the case that these problems are caused in part by the same phenomenon: having introduced the concept of formal proof with all the non-canonical and irrelevant for the content syntactical details we are then no longer able to state proof-theoretic results in a clear, modular way.

Whatever the cause, a way out is to provide a more general, or more abstract, notion of proof. In mathematics a way to generalizations is often pointed out by an axiomatic approach. The standard notion of formal proof is a *genetic* one — proofs are the objects constructed by certain rules from basic symbols.<sup>15</sup> Provability logic emerges from the idea of treating the notions of provability and proof *axiomatically* rather than genetically.

The situation when one and the same notion can be defined genetically as well as axiomatically is quite common in mathematics. A well-known example is the concept of real number, where the standard Dedekind definition is, in this sense, genetic. A well-known axiomatic definition would be the second order categorical axiomatization of the structure of reals. In an ideal situation, as in this one, axiomatic description is categorical and

---

<sup>15</sup>The term ‘genetic’ as opposed to axiomatic was used by Hilbert in 1899 in his paper “On the concept of number”.

one obtains a perfect match between both approaches. However, this ideal cannot be always achieved.

The main advantage of axiomatic approach compared to a genetic one is that it allows more easily for generalizations of notions in question. Hence, it widens the range of applicability of the theory and clarifies its logical structure. On the other hand, genetic approach is better when it comes to questions of explicit representation and computation. A good illustration of these two different roles is the axiomatic treatment of vector algebra versus genetic matrix calculus.

As far as the study of proofs is concerned, an abstract axiomatic approach has not yet been really developed. Could there be such a thing as *coordinate-free proof theory*? We share the spirit of Hilbertian optimism and formulate a problem, which is in fact a broad program of research rather than a single question, in the form of an imperative.

**Problem 3** *Develop the theory of proofs on a sufficiently abstract axiomatic basis.*

In particular, we hope that such a theory could potentially help to elucidate the informal notion of deductive proof and the other, non-deductive notions of proof.

The main line of development of provability logic strived to characterize axiomatically an already existing (genetically defined) notion of provability in a sufficiently strong arithmetical theory. The goal was to obtain a sound and complete system of axioms for provability. This goal was achieved by R. Solovay for the case of a very weak propositional language with provability modality. However, the success of propositional provability logic was undermined by non-axiomatizability results by Artemov, Vardanyan and Boolos–McGee on predicate provability logic (see [20]).

In hindsight it appears that the preoccupation with arithmetical completeness results in provability logic — which naturally came from the motivations discussed at the beginning of this paper and was further enhanced by the fascination with the beauty of Solovay’s Theorem — actually lead the researchers away from the other relevant questions such as those related to the informal concepts of proof. Hence, the idea of axiomatic reconstruction of proof theory was never pursued or posed as a problem.

Although from the very beginning there were hopes to find applications of modal logic methods in the study of formal arithmetic, yet it was never acknowledged that serious applications would require to some extent the reconstruction of the standard proof-theoretic results and that the current

modal languages were way too weak for that task. An abstract approach to proof theory based on provability logic ideas would require a development of this discipline in a new direction. Of course, technical experience accumulated in this area for so many years will still be highly relevant for this program.

From this point of view, both recent developments in provability logic mentioned before — logic of proofs and provability algebras — can be seen as attempts to approach various aspects of this general problem. The logic of proofs made the first steps in characterizing axiomatically the notion of proof rather than that of provability. Provability algebraic approach, in contrast, aims at reconstructing those results in classical proof theory which are expressible in terms of the more abstract notion of provability and treating them at the same level of generality.

Having this philosophy in mind we now turn to a different topic — the long standing problems in traditional provability logic.

### 3 Basics of Provability Logic

In this section we briefly formulate the basic facts concerning provability logic needed to read the rest of the paper. The section is more intended to fix the notations than as a real introduction. The reader is referred to one of the introductory texts [82, 81, 20, 21, 25] and a survey [6].

The basic system of provability logic is the modal propositional Gödel-Löb logic **GL**. On top of the classical propositional calculus the modal axioms and rules of **GL** are as follows:

$$\mathbf{L1} \vdash \varphi \Rightarrow \vdash \Box \varphi$$

$$\mathbf{L2} \vdash \Box(\varphi \rightarrow \psi) \rightarrow (\Box \varphi \rightarrow \Box \psi)$$

$$\mathbf{L3} \vdash \Box \varphi \rightarrow \Box \Box \varphi$$

$$\mathbf{L4} \vdash \Box(\Box \varphi \rightarrow \varphi) \rightarrow \Box \varphi$$

The principle **L4** (Löb's Principle) is interderivable with Löb's Rule:

$$\mathbf{LR} \vdash \Box \varphi \rightarrow \varphi \Rightarrow \vdash \varphi$$

Löb's Rule works too when we add assumptions of the form  $\Box \chi$ , where  $\Box \chi = (\chi \wedge \Box \chi)$ . The principle **L3** follows from **L1**, **L2** and **L4**.

Consider a theory  $T$  into which a sufficiently strong arithmetical theory  $S$  is interpretable.<sup>16</sup> Specifically, we want  $S$  to be an extension of Buss's theory  $S_2^1$  (see [23] or [47]). We assume that the axioms of  $T$  are given by a  $\Delta_1^b$ -formula. We employ a fixed efficient arithmetization of arithmetical concepts like the provability predicate  $\text{Prov}_T(x)$ . These arithmetizations are employed in  $T$  via the interpreted theory  $S$ .<sup>17</sup>

We define a  $T$ -realization  $f_T$  of the modal language into the language of  $T$  as follows:

- $f_T(p)$  is a sentence of the language of  $T$ ,
- $f_T$  commutes with the propositional connectives,
- $f_T(\Box\varphi) := \text{Prov}_T(\ulcorner f_T(\varphi) \urcorner)$ .

We say that a formula  $\varphi$  is *arithmetically valid* in  $T$  iff, for all  $T$ -realizations  $f$ ,  $T \vdash f_T(\varphi)$ . The *provability logic of  $T$* , denoted  $\mathbf{PL}_T$ , is the set of all modal propositional formulas arithmetically valid in  $T$ .

It is easy to see that all theorems of  $\mathbf{GL}$  are arithmetically valid in  $T$ , that is,  $\mathbf{GL}$  is arithmetically sound. For a wide class of theories we also have arithmetical completeness as was shown by Robert Solovay [83].

Let EA denote the Elementary Arithmetic (or  $I\Delta_0 + \text{exp}$ ) [47].

**Theorem 1 (Solovay)** *Suppose that  $T$  interprets EA and that  $T$  is  $\Sigma_1$ -sound w.r.t. this interpretation. Then,  $\mathbf{PL}_T = \mathbf{GL}$ .*

The strength of Solovay's theorem can also be seen as a disadvantage: the provability logic of a theory gives very little information about a theory. We will see that the situation is different if we change the underlying logic, e.g. to constructive logic (see Section 4), or if we extend the modal language, e.g. to the language of interpretability logic (see Section 8). Also we may build in extra variation in our notion 'the logic of'. E.g. we may consider the principles for provability in  $T$  verifiable in a theory  $U$ . The appropriate notion here is  $\mathbf{PL}_T(U)$ , the set of all  $\varphi$  such that, for all  $T$ -realizations  $f$ ,  $U \vdash f_T(\varphi)$ .  $\mathbf{PL}_T(U)$  is called the *provability logic of  $T$  relative to a metatheory  $U$* .

---

<sup>16</sup>The formulation employing an interpretation takes care of cases like  $\mathbf{ZF}$  which are not 'really' about numbers. We need an interpretation like the von Neumann interpretation to have access to number theory.

<sup>17</sup>Note that e.g. in  $\mathbf{ZF}$  we could treat syntax directly without the detour via arithmetic. Our strategy of treating syntax by *composing* a fixed arithmetization in a basic arithmetical theory with an interpretation of that theory is just a convenient design choice that goes back to Feferman's classical paper [30].

Solovay has also shown (the so-called *Solovay's Second Theorem*) that  $PL_T(\mathbf{TA}) = \mathbf{S}$ , where  $\mathbf{TA}$  is the set of all true arithmetical sentences,  $T$  is a sound arithmetical theory, and  $\mathbf{S}$  is the extension of all theorems of  $\mathbf{GL}$  by the axiom  $\Box\varphi \rightarrow \varphi$  and modus ponens as the sole inference rule. A complete classification of relative provability logics (for  $T$  containing  $\mathbf{EA}$ ) was given by Lev Beklemishev in [8], see also Section 6.

## 4 Provability Logic for Intuitionistic Arithmetic

Whereas provability logic for classical arithmetical theories turns out to be remarkably stable, as long as we restrict ourselves to the usual unimodal language, the situation for constructive arithmetical theories is spectacularly different. Different constructive theories may have different logics. Moreover, many of the principles of different logics of this kind are mutually incompatible in the sense that together they imply an iterated inconsistency statement over the minimal constructive provability logic  $i\mathbf{GL}$ , i.e. the Gödel–Löb logic over intuitionistic propositional logic instead of the classical one.

In our exposition we will assume that the reader is familiar to some extent with constructive logic and constructive arithmetic. The reader is referred to the excellent textbooks [87, 88].

The main problem we want to describe is the problem of axiomatization and decidability of the provability logic of Heyting arithmetic  $\mathbf{HA}$ . The theory  $\mathbf{HA}$  is defined exactly as the first order Peano arithmetic  $\mathbf{PA}$ , with its underlying logic changed to the intuitionistic predicate logic.<sup>18</sup>

**Problem 4** *Give an axiomatization of  $PL_{\mathbf{HA}}$ . Is  $PL_{\mathbf{HA}}$  decidable?*

To get all the pieces on the board in a systematic way, we will backtrack a bit, to treat the simpler question: *what is the ordinary propositional logic of an arithmetical theory?* The question is, in the intuitionistic case, not trivial.

### 4.1 Propositional Logics of Arithmetical Theories

The propositional logic of an arithmetical theory  $T$  is, of course, just the box-free part of  $PL_T$ .

---

<sup>18</sup>The least element principle intuitionistically implies the law of excluded middle. So, we should employ the standard version of induction.

In 1969, Dick de Jongh shows in an unpublished paper that the propositional logic of HA is precisely IPC. He uses substitutions of formulas of a complicated form. See the extended abstract [26]. Subsequently, the same result has been proved for many theories other than HA, see e.g. [80, 34, 91].

Harvey Friedman [32] improves de Jongh's result for propositional logic, showing that there is a substitution of  $\Pi_2$ -sentences  $\sigma$  such that, for all propositional formulas  $\varphi$ ,  $\text{HA} \vdash \sigma(\varphi) \iff \text{IPC} \vdash \varphi$ . Thus, IPC is *uniformly complete* for  $\Pi_2$ -realizations in HA. From the algebraic point of view the result tells us that the free Heyting algebra on countably many generators can be embedded in the Lindenbaum Algebra of HA. Moreover, we may take as generators (equivalence classes of)  $\Pi_2$ -sentences. Albert Visser [93] improves Friedman's result, employing a realization by  $\Sigma_1$ -sentences. The proof is verifiable in  $\text{HA} + \text{Con}(\text{HA})$ . (Note that de Jongh's theorem *implies*  $\text{Con}(\text{HA})$ , so the result is, in a sense, optimal.) The proof is based on the *NNIL-algorithm*, an algorithm that is used to characterize the admissible rules for  $\Sigma_1$ -realizations. This last result also holds for a number of other theories, see [92, 28].

Let MP be Markov's Principle and let  $\text{ECT}_0$  be extended Church's Thesis. Craig Smoryński has shown that the logic of  $\text{HA} + \text{MP}$  is precisely IPC and Yuri Gavrilenko has shown that the logic of  $\text{HA} + \text{ECT}_0$  is precisely IPC. Surprisingly, the logic of  $\text{HA} + \text{MP} + \text{ECT}_0$  turns out to be a proper extension of IPC. Consider the following formulas  $\chi$  and  $\rho$ :

- $\chi := (\neg p \vee \neg q),$
- $\rho := [(\neg\neg\chi \rightarrow \chi) \rightarrow (\neg\neg\chi \vee \neg\chi)] \rightarrow (\neg\neg\chi \vee \neg\chi)$

Clearly,  $\rho$  is IPC-invalid. A minor adaptation of the arguments of G.F. Rose, see [70], shows that  $\rho$  is in the logic of  $\text{HA} + \text{MP} + \text{ECT}_0$ .

**Problem 5** *What is the propositional logic of  $\text{HA} + \text{MP} + \text{ECT}_0$ ?*

Let  $\varphi$  be an arithmetical sentence. We write  $x \mathbf{r} \varphi$  for ' $x$  realizes  $\varphi$ ' in the sense of Kleene, see [87, 88]. The formula ' $x \mathbf{r} \varphi$ ' is itself an arithmetical formula. We write  $\varphi^r$  for  $\exists x x \mathbf{r} \varphi$ . By a result of Anne Troelstra, we have that the theorems of  $\text{HA} + \text{MP} + \text{ECT}_0$  coincide with the set of  $\varphi$  such that  $\text{HA} + \text{MP} \vdash \varphi^r$ . Thus, our Problem 5 can be viewed as the question what the propositional logic of the principles realized in  $\text{HA} + \text{MP}$  is. What happens if we replace  $\text{HA} + \text{MP}$  in this rephrased question by another theory?

Perhaps the most interesting theory for this question is the full true arithmetic TA. So, we may ask: what is the propositional logic of the set

of principles realized in TA? The answer is somewhat disappointing: it is precisely classical propositional logic, since sentential excluded middle is realized over TA.<sup>19</sup> Upon reflection, the reason of this disappointing outcome is the fact that we did not demand a sufficiently effective connection between the realizations of propositional formulas and their realizers. To obtain more effective connections, we introduce the following notions.

- *The propositional logic of a theory  $T$  for open realizations* is the set of propositional  $\varphi$  such that, for all realizations  $f$  in (possibly open) arithmetical formulas, we have  $T \vdash \forall \vec{x} f(\varphi)$ , where  $\vec{x}$  consists of the free variables of  $f(\varphi)$ . Clearly, the propositional logic for open realizations of a theory is a sublogic of the propositional logic of that theory for sentential realizations.
- IR, the set of *identically realizable propositional formulas* is the logic for open realizations of the set of realizable sentences of TA. I.o.w., it is the set of  $\varphi$  such that, for all open  $f$ , there is an  $n$  such that  $n \mathbf{r} \forall \vec{x} f(\varphi)$ .
- ER, the set of *effectively realizable propositional formulas*, is the set of  $\varphi$ , such that there is a recursive function  $F$  on (finite representations of) realizations such that, for all  $f$ ,  $F(f) \mathbf{r} f(\varphi)$ .
- IER, the set of *effectively indentially realizable propositional formulas*, is the set of  $\varphi$ , such that there is a recursive function  $F$  on (finite representations of) open realizations such that, for all open realizations  $f$ ,  $F(f) \mathbf{r} \forall \vec{x} f(\varphi)$ .
- UR, the set of *uniformly realizable propositional formulas*, is the set of  $\varphi$ , such that there is an  $n$  such that, for all  $f$ ,  $n \mathbf{r} f(\varphi)$ .

It is not difficult to show that  $\text{UR} \subseteq \text{IER} = \text{ER} \subseteq \text{IR}$ . We are led to the following questions.

**Problem 6** (Markov) *Give a characterization of UR, ER and IR. Is UR equal to ER? Is ER equal to IR? Are these logics decidable?*

We can ask analogous questions as in Problem 6 replacing Kleene's realizability by Gödel's *Dialectica* interpretation, so we arrive at the following problem, first formulated by Valery Plisko.

---

<sup>19</sup>Despite this fact, the set of sentences realized over TA is a constructive theory inconsistent with classical logic!

**Problem 7** (Plisko) *Characterize the propositional logics of Gödel's Dialectica interpretation.*

It is well-known that  $\text{HA} + \text{MP} + \text{ECT}_0$  is finitely axiomatizable over the theory  $\text{HA} + \text{ECT}_0$ , to wit by primitive recursive Markov's Principle  $\text{MP}_{\text{PR}}$  (see [87, 88]). By Gavrilenko's result, the propositional logic of  $\text{HA} + \text{ECT}_0$  is IPC. By Rose's result the propositional logic of  $\text{HA} + \text{MP} + \text{ECT}_0$  is not IPC. Thus, consistent addition of one sentence may change the propositional logic of a theory. This suggests the following problem.

**Problem 8** *Suppose  $\text{HA} + A$  is consistent. Is it always the case that the propositional logic of  $\text{HA} + A$  is IPC?*<sup>20</sup>

We can ask the same question for a consistent recursively enumerable  $T$  which extends  $\text{HA}$  by axioms of restricted complexity.<sup>21</sup>

The problems concerning the propositional logic of a theory extend to similar problems concerning the predicate logic of a theory. We will not pursue that direction of questioning in this paper.

Another extension of the questions is to ask for a description of the Lindenbaum algebra of a theory, which in this case will be a Heyting algebra. We will mention some of these questions in Section 7. Notice that a positive answer to Problem 8 implies that the Lindenbaum algebras of  $\text{HA}$  and  $\text{HA} + \text{ECT}_0$  are not isomorphic. The problem of the admissible rules of a theory, discussed in the next subsection, is in fact a subproblem of the problem of characterizing the Lindenbaum algebra, since the admissible rules of a theory only depend on the Lindenbaum algebra.

## 4.2 Admissible Rules

Recall that a propositional inference rule  $\varphi/\psi$  is admissible in a logic  $L$ , if for every substitution  $\sigma$  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  $f$ ,

$$T \vdash f(\varphi) \Rightarrow T \vdash f(\psi).$$

<sup>20</sup>In this spirit, Smoryński has shown that excluded middle is not finitely axiomatizable over  $\text{HA}$ .

<sup>21</sup>See [22] for a treatment of logical complexity measures over  $\text{HA}$ .



The simplest example of a (nontrivial) admissible rule in IPC is the *independence of premise* rule:

$$\text{IP: } \text{IPC} \vdash \neg\varphi \rightarrow (\psi \vee \theta) \Rightarrow \text{IPC} \vdash (\neg\varphi \rightarrow \psi) \vee (\neg\varphi \rightarrow \theta).$$

A well-known result obtained by Vladimir Rybakov [71, 72] is that the property of a rule being admissible in IPC is decidable. Albert Visser [97] showed that the propositional admissible rules for HA are the same as those for IPC.<sup>22</sup>

It is clear that any admissible propositional inference rule  $\varphi/\psi$  in HA (or, equivalently, IPC) delivers a principle of the provability logic  $\mathbf{PL}_{\text{HA}}(\text{TA})$  of the form  $\vdash \Box\varphi \rightarrow \Box\psi$ . Here TA is true arithmetic. Is the principle also in  $\mathbf{PL}_{\text{HA}}$ ? The answer is *yes*. Rosalie Iemhoff [49] proved, building on work of Silvio Ghilardi [37], that the set of all pairs of propositional formulas  $\varphi/\psi$  such that  $\Box\varphi \rightarrow \Box\psi$  is in  $\mathbf{PL}_{\text{HA}}$  is precisely the set of admissible rules of IPC.

Note that it follows that, corresponding to IP, we have the following principle of the provability logic of HA:

$$\vdash \Box(\neg\varphi \rightarrow (\psi \vee \theta)) \rightarrow \Box((\neg\varphi \rightarrow \psi) \vee (\neg\varphi \rightarrow \theta)).$$

This principle is in  $\mathbf{GL}$ , but not in  $i\mathbf{GL}$ , the version of  $\mathbf{GL}$  with IPC as underlying logic.

**Problem 9** *What are the propositional admissible rules of  $\text{HA} + \text{MP}$  and of  $\text{HA} + \text{ECT}_0$ ?*

For the study of the provability logic of HA it is interesting to study admissible rules for  $\Sigma_1$ -realizations over HA. Specifically the characterization we give below is a lemma to characterize the closed fragment of the provability logic of HA. To formulate the result we need a few preliminaries.

A NNIL-formula is a formula with no nestings of implications to the left. We take  $\neg p$  to be an abbreviation of  $(p \rightarrow \perp)$ . So  $(p \rightarrow (q \vee \neg q))$  and  $\neg p$  are NNIL-formulas, and  $((p \rightarrow q) \rightarrow q)$  is not a NNIL-formula. Van Benthem and Visser have shown (independently) that the NNIL-formulas are precisely

---

<sup>22</sup>One can easily show that, if  $L$  is the propositional logic of a theory  $T$ , then the rules admissible in  $T$  are a subset of the rules admissible in  $L$ . Thus, Visser's result shows that the rules admissible for HA are the maximum set possible, given de Jongh's Theorem. In contrast one can produce an example of an arithmetical theory for which the logic is IPC and for which the admissible rules are the *derivable rules* of IPC (see [28]). Thus, the admissible rules of theories  $T$  with propositional logic IPC can both be the maximal possible set and the minimal possible one.

the formulas preserved under taking sub-Kripke models (modulo provable equivalence). Here a submodel is a full submodel given by an arbitrary subset of the nodes, see [93, 98, 100]. Note that the NNIL-formulas form an analogue of the universal formulas in ordinary model theory [100].

The result connecting NNIL-formulas and admissible rules is as follows (see Visser [98]).

**Theorem 2** *There is an effectively computable function  $(.)^*$  from propositional formulas to NNIL-formulas such that*

- (i)  $(.)^*$  is, modulo IPC-provable equivalence, surjective;
- (ii)  $\varphi/\psi$  is admissible for HA w.r.t.  $\Sigma_1$ -realizations iff  $\text{IPC} \vdash \varphi^* \rightarrow \psi$ .

If we write  $\varphi \vdash_{\text{HA}}^{\Sigma_1} \psi$ , for  $\varphi/\psi$  is admissible in HA w.r.t.  $\Sigma_1$ -realizations, the second part of the above result can be symbolized as follows.

$$\varphi^* \vdash_{\text{IPC}} \psi \Leftrightarrow \varphi \vdash_{\text{HA}}^{\Sigma_1} \psi.$$

In other words, the embedding  $\vdash_{\text{IPC}} \hookrightarrow \vdash_{\text{HA}}^{\Sigma_1}$  has left adjoint  $(.)^*$ . It follows that  $\varphi^*$  is the best NNIL-approximation from below (in the preorder of IPC-provability) of  $\varphi$ . Thus, admissibility in HA w.r.t.  $\Sigma_1$ -realizations is, in a sense, completely characterized by the class NNIL.<sup>23</sup> It would be interesting to combine the results on the admissible rules of HA for arbitrary realizations and those on  $\Sigma_1$ -realizations.

**Problem 10** *Extend the language of propositional logic with a second sort of propositional variables  $s_1, s_2, \dots$ . Realizations will send ordinary variables to arithmetical sentences and the new variables to  $\Sigma_1$ -sentences. Characterize the rules for this language admissible in HA.*

### 4.3 The Provability Logic of HA and related theories

The usual process of arithmetization of syntax is constructive and therefore can be carried out in  $i\text{EA}$ . Here  $i\text{EA}$  is the intuitionistic counterpart of elementary arithmetic ( $\Delta_0$ -induction plus the axiom stating that the exponentiation function is total). In particular, the provability predicate for any elementarily presented theory  $T$  can be formulated as a  $\Sigma_1$ -formula. Moreover, this formula satisfies the usual Löb's derivability conditions within  $i\text{EA}$ .

---

<sup>23</sup>Ordinary admissibility for HA, and, thus, IPC, also has a left adjoint. However, we know of no simple formula class characterizing this adjoint.

The definitions of provability interpretation and of provability logic of a theory relative to a metatheory carry over without any change.  $\mathbf{PL}_{\mathbf{HA}}$  will denote the provability logic of Heyting arithmetic that we are particularly interested in.

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  $i\mathbf{EA}$ . Consequently, the logic  $\mathbf{PL}_T(i\mathbf{EA})$  contains the axioms and rules of  $\mathbf{GL}$  formulated over the intuitionistic propositional logic  $\mathbf{IPC}$ . We denote this basic system by  $i\mathbf{GL}$ .

It was immediately clear that  $\mathbf{PL}_{\mathbf{HA}}$  satisfies some additional principles. A number of such independent principles were found in [91]. For example  $\mathbf{HA}$  is closed under the so-called *Markov's rule* (see [87]):

$$\mathbf{HA} \vdash \neg\neg\pi \Rightarrow \mathbf{HA} \vdash \pi,$$

where  $\pi$  is a  $\Pi_2$ -formula. This can be proved constructively using the so-called *Friedman–Dragalin translation*. Thus, a proof of this fact can be formalized in  $\mathbf{HA}$  itself, therefore  $\mathbf{PL}_{\mathbf{HA}}$  contains the principle

$$\Box\neg\neg\Box\varphi \rightarrow \Box\Box\varphi.$$

A more general provable form of the same principle is as follows:

$$\text{Ma : } \Box\neg\neg(\Box\psi \rightarrow \bigvee_{i=1}^n \Box\varphi_i) \rightarrow \Box(\Box\psi \rightarrow \bigvee_{i=1}^n \Box\varphi_i).$$

The *disjunction property* for  $\mathbf{HA}$  is the statement that, whenever  $\mathbf{HA} \vdash \varphi \vee \psi$ , one has  $\mathbf{HA} \vdash \varphi$  or  $\mathbf{HA} \vdash \psi$ . This can be written down as

$$\text{Dis : } \Box(\varphi \vee \psi) \rightarrow \Box\varphi \vee \Box\psi.$$

However, Friedman [33] has shown that the proof of disjunction property cannot be formalized in  $\mathbf{HA}$  itself.

Daniel Leivant found a nice weakening of the disjunction property that is already provable in  $\mathbf{HA}$ :

$$\text{Le : } \Box(\varphi \vee \psi) \rightarrow \Box(\Box\varphi \vee \psi).$$

This principle was formulated by Leivant in his Ph.D. thesis. For a proof of this fact see [98].

Leivant's principle is weakly inconsistent with excluded middle in the sense that these principles combined prove a formula of the form  $\Box^n \perp$  over  $i\mathbf{GL}$ . We reason in  $\mathbf{GL} + \text{Le}$ . Clearly, we have  $\Box(\Box\perp \vee \neg\Box\perp)$ . Hence, by Le, we have that  $\Box(\Box\perp \vee \Box\neg\Box\perp)$ . So, by Löb's principle, we obtain  $\Box^2\perp$ .

The validity of Leivant's principle illustrates that  $\mathbf{PL}_T$  is *not* monotonically increasing in  $T$ .

These sample principles do not exhaust the list of all principles valid over HA. For all we know, the  $\mathbf{PL}_{\text{HA}}$  could be complete  $\Pi_2$ . A list of all the principles we know at the moment of writing can be found in [98] or [50] or [6].

Next to the great open problem of the provability logic of HA, we may ask after the logic of all extensions of HA.

**Problem 11** *What is the intersection of all  $\mathbf{PL}_T$  for recursively enumerable extensions  $T$  of HA in the language of HA?*

We have precise knowledge concerning two fragments of  $\mathbf{PL}_{\text{HA}}$ . In the first place, as described above, we have a precise description of all principles of the form  $\Box\varphi \rightarrow \Box\psi$ , for box-free  $\varphi$  and  $\psi$ , in  $\mathbf{PL}_{\text{HA}}$ . The pairs  $\varphi, \psi$  occurring in these principles are precisely the pairs  $\varphi/\psi$  admissible in IPC. Secondly we have a precise characterization of the closed or letterless fragment of  $\mathbf{PL}_{\text{HA}}$ . This characterization was given by Albert Visser in [93] (see also [98]). Visser's proof uses in an essential way the characterization given for the admissible rules of HA for  $\Sigma_1$ -realizations (and the fact that this result is HA-verifiable). Not much is known about other closed fragments — except in the classical case of course.

**Problem 12** *What are the closed fragments of the logic  $\mathbf{PL}_{\text{HA}+\text{MP}}$  and of the logic  $\mathbf{PL}_{\text{HA}+\text{MP}+\text{ECT}_0}$ ?*

## 5 Provability logic and bounded arithmetic

Bounded arithmetic theories were introduced and developed by Sam Buss and others in order to capture the informal notion of feasible proof and to clarify the relationships between proof theory and computation complexity theory [23, 47, 59]. The most important among these theories is the system  $S_2^1$ , which corresponds to the class of polytime computable functions, and whose principal axiom schema is the induction over binary words for  $\Sigma_1^b$ -formulas. ( $\Sigma_1^b$ -formulas in the language of bounded arithmetic naturally represent NP-predicates.) There were suggestions to identify the notion of provability in  $S_2^1$  with *feasible provability*.

Bounded arithmetic systems are important, because they allow to approach such questions as provability or unprovability of  $P \neq NP$  conjecture and are related to the study of complexity of proofs. Questions of separation

and axiomatizability of various bounded arithmetic theories are often highly non-trivial and depend on difficult open problems in complexity theory, for a source book see [59].

Several questions related to the study of provability principles in bounded arithmetic are open. The most well-known problem is whether Solovay's arithmetical completeness theorem holds for bounded arithmetic.

**Problem 13** *Characterize the propositional provability logic of bounded arithmetic theories such as  $S_2^1$  and  $S_2$ .*

It is known that all principles of the modal logic **GL** are valid under the arithmetical interpretation w.r.t. bounded arithmetic theories. Hence, it seems natural to conjecture that the provability logic of  $S_2^1$  and  $S_2$  is precisely **GL**. However, this conjecture appears to be difficult to prove.

The reason is that the standard proof of Solovay's theorem — and this is essentially the only currently known proof of that result — relies on the property of, at least sentential, *provable  $\exists\Pi_1^b$ -completeness*. Given a theory  $T$ , this property states that, for every  $\Pi_1^b$ -formula  $\varphi(x)$ ,

$$T \vdash \exists x \varphi(x) \rightarrow \text{Prov}_T(\ulcorner \exists x \varphi(x) \urcorner).$$

The standard formalization of the proof predicate is  $\Sigma_1^b$  in the bounded arithmetic hierarchy (in fact, it is polytime computable). Hence,  $\text{Prov}_T(x)$  is a  $\exists\Sigma_1^b$ -formula. It is known [23] that  $S_2^1$  satisfies provable  $\exists\Sigma_1^b$ -completeness and, in particular, Löb's third derivability condition

$$T \vdash \text{Prov}_T(\ulcorner \varphi \urcorner) \rightarrow \text{Prov}_T(\ulcorner \text{Prov}_T(\ulcorner \varphi \urcorner) \urcorner).$$

With the principle of  $\exists\Pi_1^b$ -completeness the situation is different. Theories proving the totality of exponentiation function, such as the elementary arithmetic **EA**, do satisfy this property. It is unknown whether this property holds for bounded arithmetic theories such as  $S_2^1$  and  $S_2$ . However, there is a reason to believe that it fails, because according to a result by Alexander Razborov and Rineke Verbrugge [90], if it holds then  $P = NP \cap \text{co-NP}$ . The latter statement is one of the difficult open questions in complexity theory, but it is believed to be false.

Thus, the question whether **GL** is the logic of provability for  $S_2$  may actually depend on complexity-theoretic assumptions. For one thing, if the answer is *no*, then  $S_2$  does not prove the Matiyasevich–Davis–Robinson–Putnam (MDRP) theorem, that is, that every r.e. set is diophantine. This would settle a well-known difficult problem in bounded arithmetic which

is open since the 80s. Indeed, MDRP theorem implies that every  $\exists\Pi_1^b$ -sentence is equivalent to a purely existential one, and for such sentences we have provable completeness even in  $S_2^1$ . Of course, we do not really believe that this is a plausible way to solve the MDRP problem.

A possibility remains that one can give a relatively easy proof of the arithmetical completeness theorem for **GL** without using provable  $\exists\Pi_1^b$ -completeness. Alessandro Berarducci and Rineke Verbrugge investigated this option in their paper [16]. Although they involved some ingenious modifications of Solovay construction, they only succeeded in embedding very simple kinds of Kripke models into bounded arithmetic. The main question remains open.

Apart from this intriguing problem, there are some other related open questions between bounded arithmetic and provability logic.

**Problem 14** *Does the Friedman–Goldfarb–Harrington principle hold in  $S_2^1$ ?*

The *FGH principle* for an arithmetical theory  $T$ , proved by Harvey Friedman, Leo Harrington and Warren Goldfarb (see [81, 99]), states that for any  $\Sigma_1$ -sentence  $S$  there is a sentence  $R$  such that  $\text{Prov}_T(\ulcorner R \urcorner)$  is  $T$ -equivalent to  $S \vee \text{Prov}_T(\ulcorner 0 = 1 \urcorner)$ . In particular, for any  $\Sigma_1$ -sentence  $S$  such that  $T \vdash \text{Prov}_T(\ulcorner 0 = 1 \urcorner) \rightarrow S$ ,  $S$  is equivalent to  $\text{Prov}_T(\ulcorner R \urcorner)$ , for some  $R$ .

For  $T = S_2^1$  we deal with sentences  $S$  of the form  $\exists\Sigma_1^b$ , because such is the complexity of the provability predicate, and ask whether the corresponding statement holds.

The problem arises in  $S_2^1$ , because the standard proof uses a Rosser-type fixed point construction of  $R$ :

$$R \leftrightarrow \text{“}S \text{ is witnessed before } \text{Prov}_T(\ulcorner R \urcorner)\text{”}.$$

Such an  $R$  is  $\exists\Pi_1^b$  and one then would want to apply the familiar  $\exists\Pi_1^b$ -completeness principle, which is not available.

A related argument was used by Harvey Friedman to prove the equivalence of the  $\Sigma_1$ -disjunction property and the  $\Sigma_1$ -reflection principle (modulo consistency). This yields a similar kind of problem in bounded arithmetic. Recall that (sentential)  $\Gamma$ -*reflection principle* is the schema

$$\text{Rfn}_\Gamma(T) : \quad \text{Prov}_T(\ulcorner \varphi \urcorner) \rightarrow \varphi,$$

for all sentences from a class  $\Gamma$ , whereas (sentential)  $\Gamma$ -*disjunction property* can be formalized as the schema

$$\text{Dis}_\Gamma(T) : \quad \text{Prov}_T(\ulcorner \varphi \vee \psi \urcorner) \rightarrow (\text{Prov}_T(\ulcorner \varphi \urcorner) \vee \text{Prov}_T(\ulcorner \psi \urcorner)),$$

for all sentences  $\varphi, \psi$  from  $\Gamma$ .

By provable  $\exists\Sigma_1^b$ -completeness,  $S_2^1 + \text{Rfn}_{\exists\Sigma_1^b}(T)$  proves  $\text{Dis}_{\exists\Sigma_1^b}(T)$ . It is also clear that the reflection schema proves  $\text{Con}(T)$ , the consistency assertion for  $T$ . However, the opposite implication relies on  $\exists\Pi_1^b$ -completeness and is, thus, an open question.

**Problem 15** *Does  $\exists\Sigma_1^b$ -disjunction property for a  $\Sigma_1^b$ -presented theory  $T$  imply its  $\exists\Sigma_1^b$ -reflection principle in  $S_2^1 + \text{Con}(T)$ ?*

A similar question also makes sense for natural versions of the reflection and disjunction principles with free variables.

## 6 Classification of bimodal provability logics

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*.

Perhaps, the most natural provability interpretation of the polymodal language is the understanding of modalities as provability predicates in some r.e. arithmetical 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. The following question is one of the main open problems in provability logic.

**Problem 16** *Characterize within the lattice of bimodal logics the propositional provability logics for pairs of r.e. arithmetical theories containing a sufficiently strong fragment of arithmetic such as **EA**.*

Although at present this problem appears to be out of reach, a lot of partial results have been obtained and it seems that a positive solution of this problem is possible. To present some details, let us first define the notion of bimodal provability logic more precisely.

The language  $\mathcal{L}(\Box, \Delta)$  of bimodal provability logic is obtained from that of propositional calculus by adding two unary modal operators  $\Box$  and  $\Delta$ . Let  $(T, U)$  be a pair of sufficiently strong r.e. arithmetical theories. An *arithmetical realization*  $f_{T,U}(\varphi)$  of a formula  $\varphi$  w.r.t.  $(T, U)$  translates  $\Box$  as

provability in  $T$  and  $\Delta$  as that in  $U$  while commuting with all the boolean connectives:

$$f_{T,U}(\Box\varphi) = \text{Prov}_T(\ulcorner f_{T,U}(\varphi) \urcorner), \quad f_{T,U}(\Delta\varphi) = \text{Prov}_U(\ulcorner f_{T,U}(\varphi) \urcorner).$$

The *provability logic for  $(T, U)$*  is the collection of all  $\mathcal{L}(\Box, \Delta)$ -formulas  $\varphi$  such that  $T \cap U \vdash f_{T,U}(\varphi)$ , for every arithmetical realization  $f$ . It is denoted  $\mathbf{PL}_{T,U}$ . In general, one can consider bimodal provability logics for  $(T, U)$  relative to an arbitrary metatheory  $V$ .  $\mathbf{PL}_{T,U}(V)$  is the set of all formulas  $\varphi$  such that  $V \vdash f_{T,U}(\varphi)$ , for every arithmetical realization  $f$ . Thus,  $\mathbf{PL}_{T,U}$  corresponds to  $V = T \cap U$ .

Not too much can a priori be said about  $\mathbf{PL}_{T,U}$ , for arbitrary  $T$  and  $U$ . Firstly,  $\mathbf{PL}_{T,U}$  is closed under *modus ponens*, substitution,  $\Box$ - and  $\Delta$ -necessitation rules.<sup>24</sup> Secondly,  $\mathbf{PL}_{T,U}$  has to be an extension of the following bimodal logic  $\mathbf{CS}$ , whose axioms and rules (on top of the principles of  $\mathbf{GL}$  formulated for  $\Box$  and for  $\Delta$ ) are as follows:

$$\text{CS1 } \vdash \Box\varphi \rightarrow \Delta\Box\varphi$$

$$\text{CS2 } \vdash \Delta\varphi \rightarrow \Box\Delta\varphi$$

$$\text{CS3 } \vdash \varphi \Rightarrow \vdash \Box\varphi$$

$$\text{CS4 } \vdash \varphi \Rightarrow \vdash \Delta\varphi.$$

In fact, applying the so-called uniform version of Solovay's theorem Craig Smoryński showed that  $\mathbf{CS}$  is the provability logic of a particular pair of finite extensions of  $\text{PA}$  (see [82]).

Deeper structural information on bimodal provability logics is provided by the Classification theorem for arithmetically complete unimodal logics [8, 13]. With every (normal) bimodal logic  $L$  containing  $\mathbf{CS}$  we can associate its  $\Box$ -*projection* or *type*:

$$(L)^\Box := \{\varphi \in \mathcal{L}(\Box) : L \vdash \Delta\varphi\}.$$

Notice that  $(L)^\Box$  contains  $\mathbf{GL}$  and is closed under modus ponens and substitution rules, but not necessarily under the necessitation.

Under the assumption of  $\Sigma_1$ -soundness of  $V$  the unimodal provability logic of  $T$  relative to  $U$  coincides with the type of  $\mathbf{PL}_{T,U}(V)$ :

$$\mathbf{PL}_T(U) = (\mathbf{PL}_{T,U}(V))^\Box.$$

---

<sup>24</sup>Notice that neither  $\mathbf{PL}_{T,U}(T)$  nor  $\mathbf{PL}_{T,U}(U)$  will in general be closed under both necessitation rules.



The Classification theorem shows that not every extension of **GL** is materialized as the projection of a bimodal provability logic and gives us a description of all such possible projections: **GL** $_{\alpha}$ , **GL** $_{\beta}^{-}$ , **S** $_{\beta}$ , **D** $_{\beta}$ ,  $\alpha, \beta \subseteq \omega$ ,  $\alpha$  r.e. and  $\omega \setminus \beta$  finite (see [13]).

This already excludes a lot of bimodal non-provability logics and provides the first step towards a general classification. Indeed, now the problem amounts to classifying bimodal provability logics of each of these types. However, such a classification is only known for logics of type **S**: in this case a theorem due to Tim Carlson [24] tells us that there is only one such provability logic. For types **D** and **GL** $_{\omega}$  we know that there is more than one logic. Some further partial results in this direction were obtained by Beklemishev in [9, 10].

Theorem 119 of [6] due to the authors of this paper characterizes all possible letterless fragments of bimodal provability logics. This result is based on a related characterization of r.e. subalgebras of Magari algebras of theories due to Vladimir Shavrukov (see [76] and [6] for a relationship between these problems).

Related to the general Classification problem for bimodal provability logics are the question of characterizing the bimodal logics for specific ‘natural’ pairs of theories. A number of results of this kind have been obtained, see [6] for an overview. However, some of the more exotic cases found among fragments of PA have not yet been treated. We mention the following questions.

**Problem 17** *Characterize the bimodal provability logics of any natural pair of theories  $(T, U)$  such that  $U$  is a  $\Pi_1$ -conservative extension of  $T$ , but  $U$  is not conservative over  $T$  w.r.t. boolean combinations of  $\Sigma_1$ -sentences.*

An example of such a pair of theories  $(T, U)$  is

$$T = \text{EA}_{\omega} = \text{EA} + \text{Con}(\text{EA}) + \text{Con}(\text{EA} + \text{Con}(\text{EA})) + \dots$$

and  $U = \text{EA} + \text{RFN}_{\Sigma_1}(\text{EA}) = I\Delta_0 + \text{supexp}$ .

**Problem 18** *Characterize the bimodal provability logics of  $(I\Sigma_1, I\Pi_2^{-})$  and the other natural pairs of incomparable fragments of PA.*

## 7 Magari algebras

The notion of Magari algebra was introduced by Roberto Magari [62] under the name *diagonalizable algebra*. Given an arithmetical theory  $T$  we consider its Lindenbaum boolean algebra  $\mathcal{B}_T$ , that is, the set of all  $T$ -sentences

modulo the equivalence relation

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

The usual logical operations provide this set with the structure of a boolean algebra, in particular, the ordering relation can be defined by:

$$[\varphi]_T \leq [\psi]_T \iff T \vdash \varphi \rightarrow \psi,$$

where  $[\varphi]_T$  denotes the  $\sim_T$  equivalence class of  $\varphi$ .

Gödel's provability formula  $\text{Prov}_T$  correctly defines an operator

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

acting on the Lindenbaum algebra  $\mathcal{B}_T$ . Indeed, if  $T \vdash \varphi \leftrightarrow \psi$ , then  $T \vdash \text{Prov}_T(\ulcorner \varphi \urcorner) \leftrightarrow \text{Prov}_T(\ulcorner \psi \urcorner)$ , by Löb's derivability conditions. The enriched structure  $\mathcal{M}_T = (\mathcal{B}_T, \Box_T)$  is called the *provability algebra* or the *Magari algebra of  $T$* .

The language of Magari algebras generalizes that of purely propositional provability logic. In particular, by Solovay's theorem the provability logic **GL** describes the set of all identities of  $\mathcal{M}_T$ , for  $\Sigma_1$ -sound  $T$ . Solovay's second theorem implies the decidability of the purely universal theory of  $\mathcal{M}_T$ , under the assumption of soundness of  $T$ .

Indeed, any quantifier-free formula  $A(\vec{x})$  in the language of Magari algebras can be understood as a propositional modal formula: One preserves the boolean connectives and translates equalities of terms  $s = t$  as formulas  $\Box(\varphi_s \leftrightarrow \varphi_t)$ , where  $\varphi_s$  denotes a formula corresponding to the term  $t$ . Validity of the universal closure of  $A(\vec{x})$  in  $\mathcal{M}_T$  is, thus, equivalent to the validity of the arithmetical interpretation of  $A(\vec{x})$ , under every substitution of  $T$ -sentences for variables  $\vec{x}$ . Hence, by Solovay's theorem, the question of validity of the universal closure of  $A$  is reducible to the one whether  $A$  is provable in the logic **S**.

The problem of decidability of the full first order theory of Magari algebra of PA stood open for some time, until it was answered negatively in an important paper by Vladimir Shavrukov [78]. In fact, for  $\Sigma_1$ -sound theories  $T$  the first order theory of  $\mathcal{M}_T$  happens to be mutually interpretable with the theory axiomatized by all true arithmetical sentences. Hence, it is not even arithmetical.

The question remains, where the border between decidable and undecidable fragments of that theory exactly goes. The proof of Shavrukov's theorem shows that four quantifier alternations are enough to get undecidability. The most interesting question then concerns the fragment for which

there is still a considerable hope for a positive result: the  $\forall^*\exists^*$ -fragment, in other words, the set of prenex formulas with a block of universal quantifiers followed by a block of existential ones.

**Problem 19** *Is the set of  $\forall^*\exists^*$ -formulas valid in the Magari algebra of PA decidable?*

Examples of meaningful valid arithmetical principles expressed by  $\forall^*\exists^*$ -formulas are plentiful. The two most prominent ones are *Rosser's theorem* and the familiar *FGH-principle*.

Rosser's theorem states that for every consistent theory there is an independent sentence. Applying this to an arbitrary finite extension of  $T$  of the form  $T + p$  yields the following principle:

$$\forall p (\Diamond p \rightarrow \exists q (\neg \Box(p \rightarrow q) \wedge \neg \Box(p \rightarrow \neg q))).$$

The FGH-principle implies that the set of sentences of the form  $\Box_T \psi$  coincides, modulo equivalence in  $T$ , with the set of all  $\Sigma_1$ -sentences above  $\Box_T \perp$ . Whereas the set of all  $\Sigma_1$ -sentences does not seem to be definable in the language of Magari algebras<sup>25</sup>, we can still formally state some nontrivial consequences of the FGH-principle, for example:

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

It is worth noticing that neither in this case, nor in the case of Rosser's principle, is the Skolem function implicitly defined by these  $\forall^*\exists^*$ -formulas expressible by a term in  $\mathcal{M}_T$ . In this sense these principles are nontrivial, i.e., not expressible by  $\forall^*$ -formulas.

As we mentioned before, we really expect a positive solution of Problem 19. Not much is known towards a possible solution of this difficult problem. There is an interesting connection, observed in [6], between this problem and the problem of classification of the so-called propositional *provability logics with constants*.

Consider a tuple of arithmetical sentences  $\vec{A}$ . Provability logic with the constants for  $\vec{A}$  is, essentially, the set of all universal formulas  $\forall \vec{x} \varphi(\vec{c}, \vec{x})$  such that  $\mathcal{M}_T \models \forall \vec{x} \varphi(\vec{A}, \vec{x})$ , in other words, the *universal type* of the tuple  $\vec{A}$  in  $\mathcal{M}_T$ . A universal type is *realizable* in  $\mathcal{M}_T$ , if

$$\mathcal{M}_T \models \exists \vec{c} \forall \vec{x} \varphi(\vec{c}, \vec{x}).$$

---

<sup>25</sup>In fact, it is an open question.

We conjecture that there is an effective description of all universal types realizable in  $\mathcal{M}_T$ , hence the  $\forall^*\exists^*$ -fragment of the first order theory of  $\mathcal{M}_T$  is decidable.

Notice that the classification of universal types is basically the same kind of problem as the classification of propositional polymodal provability logics. In fact, it is very close to the classification of provability logics for a tuple of finite extensions of  $T$ . At the moment, even the simplest variant of this question — the classification of bimodal provability logics for pairs of theories of the form  $(T, T + A)$ , for a single sentence  $A$  — is wide open. Therefore, we formulate the following meaningful particular case as a separate problem.

**Problem 20** *Give an effective description of all possible provability logics with a constant for a single sentence over PA.*

It is worth mentioning that Shavrukov [76] also essentially gave a description of all possible closed fragments of provability logics with constants for an arbitrary tuple of sentences  $\vec{A}$  (or *open types* of elements  $\vec{A}$  of  $\mathcal{M}_T$ ). They can be viewed as the propositional theories in the language with variables  $\vec{c}$  over **GL** which are r.e. and satisfy the so-called *strong disjunction property*, in the case  $T$  is  $\Sigma_1$ -sound. It is decidable whether a finitely axiomatized propositional theory satisfies this property. However, arithmetically realizable propositional theories may, in general, be infinitely axiomatized.

Apart from that, a number of particular logics for ‘natural’ constants have been characterized [9, 10]. For example, if a constant  $c$  corresponds to a true  $\Pi_1$ -sentence that implies all finite iterations of the consistency assertion for  $T$ , then the logic of such a constant has a nice axiomatization and is decidable. Examples are the consistency statement  $\text{Con}(\text{ZF})$  over PA or  $\text{Con}(I\Sigma_1)$  over EA.

Additional information on logics with constants can be extracted from the Classification theorem for unimodal provability logics [8, 13]. This can be done in a manner similar to our comments on the classification problem for bimodal logics of provability.

Apart from the important Problem 19 a number of other natural questions about Magari algebras remains open.

One group of questions concerns the isomorphism problem for Magari algebras. Ideally, one would want to have a complete classification of Magari algebras of theories modulo isomorphism. However, nobody believes such a sweeping classification to be possible. Rather, one is looking for interesting criteria of isomorphism and non-isomorphism of algebras. Some such criteria

have been formulated by Shavrukov [75, 77], who has proved, roughly, that  $\mathcal{M}_T$  is recursively isomorphic to  $\mathcal{M}_U$ , if  $T$  and  $U$  are (effectively) conservative over each other for boolean combinations of  $\Sigma_1$ -sentences (for example  $T = I\Sigma_1$  and  $U = \text{PRA}$ ). On the other hand, the algebras will be non-isomorphic, if one of the theories proves the uniform  $\Sigma_1$ -reflection principle for the other (for example,  $T = \text{PA}$  and  $U = \text{ZF}$ ). Between these two opposite classes there are still many meaningful examples of pairs of theories, such as  $\text{PA}$  and  $\text{PA} + \text{Con}(\text{PA})$ . We formulate the general question and its most obvious particular case.

**Problem 21** (*Isomorphism*) *Give sharp necessary and sufficient conditions for the isomorphism of Magari algebras of reasonable theories. In particular, are the algebras of  $\text{PA}$  and  $\text{PA} + \text{Con}(\text{PA})$  isomorphic?*

Another group of questions concerns definability in the provability algebras. Very little is known about it, in particular, about the definable elements of the algebra. Some non-definability results have been obtained by Shavrukov [77], however the general question remains open.

**Problem 22** *Characterize the first order definable elements of  $\mathcal{M}_{\text{PA}}$ .*

A natural conjecture is that these are only the elements of the 0-generated subalgebra of  $\mathcal{M}_{\text{PA}}$ , that is, the elements already definable by ground (or variable-free) terms of the structure.

As the reader may have noticed, all the questions formulated in this section, except for Problem 20, come from the deep work of Shavrukov on the theory of Magari algebras and have been formulated by him.

## 8 Interpretability Logic

One of the most interesting extensions of the modal language of provability logic is the extension with a binary modality which can be arithmetically interpreted as interpretability or as conservativity. The first one to consider such extensions was Vítěslav Švejdar in his pioneering paper [85]. In this paper he showed that some substantial reasoning can be represented in such logics.

The project of interpretability logic was subsequently taken up by Albert Visser who formulated two conjectures concerning arithmetical completeness. Visser proposed a system  $\text{ILM}$  as a candidate for the interpretability logic of Peano Arithmetic and a system  $\text{ILP}$  as a candidate for the interpretability logic of Gödel–Bernays set theory. Frank Veltman found a Kripke

style semantics for these logics, which was studied in [27].<sup>26</sup> Visser's first conjecture about arithmetical completeness was proved independently by Volodya Shavrukov [74] and by Alessandro Berarducci [15]. Visser himself proved the second conjecture, see [94]. The main open problem (on the arithmetical side) left open by the work of the 90's of the previous century is the question of the interpretability logic of all reasonable arithmetical theories. For a survey of the status questionis, see [56].<sup>27</sup> For surveys of the whole area, the reader might consult [53, 96] and [55].

Interpretations are ubiquitous in mathematics. To mention a few examples, think of the Poincaré interpretation of two dimensional hyperbolic geometry in two dimensional Euclidean geometry, the von Neumann interpretation of number theory in set theory, the Ackermann interpretation of the theory of finite sets in arithmetic and Tarski's interpretation of arithmetic in an extension of the theory of groups with one extra constant.

The interpretations we are interested in are *relative interpretations* in the sense of Tarski, Mostowski and Robinson (see [86]). Consider theories  $U$  with language  $L_U$  and  $T$  with language  $L_T$ . For simplicity, we assume that  $L_U$  is a relational language. An interpretation  $K$  of  $U$  in  $T$  is given by a pair  $\langle \delta(x), F \rangle$ . Here  $\delta(x)$  is an  $L_T$ -formula representing the *domain* of the interpretation.  $F$  is a mapping that associates to each relation symbol  $R$  of  $\mathcal{L}_U$  with arity  $n$  an  $\mathcal{L}_T$ -formula  $F(R)(x_1, \dots, x_n)$ . Here  $x_1, \dots, x_n$  are suitably chosen free variables. We translate the formulas of  $L_U$  to the formulas of  $L_T$  as follows:

- $K(R(y_1, \dots, y_n)) := F(R)(y_1, \dots, y_n)$ ,  
(We do not demand that identity is translated as identity.)
- $K$  commutes with the propositional connectives,
- $K(\forall y \varphi) := \forall y (\delta(y) \rightarrow K(\varphi))$ ,

---

<sup>26</sup>The history contains an instance of the sometimes almost mystical quality of mathematico-logical research. The system for which Albert Visser asked Frank Veltman to produce a semantics was  $\text{IL}(\text{KM1})$ , a system that is weaker than  $\text{ILM}$ . During the time Veltman was inventing the semantics, Visser realized the validity of  $\text{M}$ , when analysing an argument by Franco Montagna in a letter. (Later it would turn out that Per Lindström already knew the principle.) The miracle was that the semantics Veltman found *did* validate  $\text{M}$ , even if he didn't know this principle. It turned out that  $\text{F}$  does not imply  $\text{M}$ , even if both principles correspond to the same set of Veltman frames — an example of modal incompleteness. Later Joost Joosten and Evan Goris discovered the principle  $\text{R}$  when they were looking at a class of Veltman frames satisfying other principles.  $\text{R}$  turned out to be arithmetically valid.

<sup>27</sup>The survey is already slightly outdated since Joost Joosten and Evan Goris discovered new principles in the mean time, see [55].

- $K(\exists y \varphi) := \exists y (\delta(y) \wedge K(\varphi))$ ,

Finally, we demand that for all sentences  $\varphi$  which are universal closures of axioms of  $U$ , we have  $T \vdash K(\varphi)$ . We will write  $T \triangleright U$  for  $K$  is an interpretation of  $U$  in  $T$ . We write  $T \triangleright U$ , for  $K : T \triangleright U$ , for some  $K$ .

Interpretations are used for various purposes: to prove relative consistency, conservation results and undecidability results. The syntactical character of interpretations has the obvious advantage that it allows us to convert proofs of the interpreted theory in an efficient way into proofs of the interpreting theory.

In order to be able to reason modally about interpretability, we must ‘downtune’ the notion to relate sentences rather than theories. This can be realized as follows. We define  $\varphi \triangleright_T \psi$  as:  $(T + \varphi) \triangleright (T + \psi)$ .

To be able to iterate modalities, we must consider theories with sufficient coding potential. It is a delicate question to determine which theories are rich enough. In this section, we will not worry about finding the sharpest class. We simply will work with recursively enumerable theories which contain **EA** plus the  $\Sigma_1$ -collection principle (possibly via interpretation) and have a good coding of sequences of all objects of the domain (sequentiality). We will call these theories *reasonable*.

We consider the modal language of provability logic extended with a binary modality  $\triangleright$ . Let a reasonable theory  $T$  be given. It is easy to see that interpretability is formalizable in  $T$ . We define  $\mathbf{IL}_T$ , the interpretability logic of  $T$  in the same way as we defined the provability logic of  $T$ . The only new feature is the clause:

- $f_T(\varphi \triangleright \psi) := \ulcorner f_T(\varphi) \urcorner \triangleright_T \ulcorner f_T(\psi) \urcorner$ .

The following principles constitute a good starting point for both the modal and the arithmetical investigation. The logic **IL** is given on top of the principles of **GL** as follows:

- J1  $\vdash \Box(\varphi \rightarrow \psi) \rightarrow \varphi \triangleright \psi$
- J2  $\vdash (\varphi \triangleright \psi \wedge \psi \triangleright \chi) \rightarrow \varphi \triangleright \chi$
- J3  $\vdash (\varphi \triangleright \chi \wedge \psi \triangleright \chi) \rightarrow (\varphi \vee \psi) \triangleright \chi$
- J4  $\vdash \varphi \triangleright \psi \rightarrow (\Diamond \varphi \rightarrow \Diamond \psi)$
- J5  $\vdash \Diamond \varphi \triangleright \varphi$

The only surprising principle is J5. This principle is a ‘syntactification’ of the well-known model existence lemma: if a (first order) theory is consistent then it has a model. Here ‘model’ is replaced by ‘interpretation’.<sup>28</sup>

We will name further logics by appending the names of the further principles after IL. The first principle we consider is Montagna’s principle M.

$$\mathbf{M} \quad \vdash \varphi \triangleright \psi \rightarrow (\varphi \wedge \Box \chi) \triangleright (\psi \wedge \Box \chi)$$

This principle is valid in essentially reflexive theories. For our purposes we can simply describe these as all theories that contain Peano Arithmetic, possibly via interpretation, that have good coding of sequences and that satisfy induction *for the full language*. Examples of essentially reflexive theories are PA and ZF. Albert Visser conjectured that  $\mathbf{IL}_T = \mathbf{ILM}$ , for all essentially reflexive  $\Sigma_1$ -sound  $T$ . This conjecture was proved independently by Alessandro Berarducci [15] and Volodya Shavrukov [74].

The second principle we consider is the persistence principle P.

$$\mathbf{P} \quad \vdash \varphi \triangleright \psi \rightarrow \Box(\varphi \triangleright \psi).$$

The persistence principle is valid for interpretations in finitely axiomatized reasonable theories  $T$ . Examples of such theories are  $I\Sigma_1$ ,  $\text{ACA}_0$  and GB. The conjecture was proved in [94].

The most salient problem left by the history is the question concerning the interpretability principles valid in all reasonable theories.

**Problem 23** *Let  $\mathbf{IL}_{\text{all}}$  be the intersection of the  $\mathbf{IL}_T$  for all reasonable  $T$ . Characterize  $\mathbf{IL}_{\text{all}}$ .*

A number of further principles were discovered valid in all reasonable theories. Here they are.

$$\mathbf{W} \quad \vdash \varphi \triangleright \psi \rightarrow \varphi \triangleright (\psi \wedge \Box \neg \varphi)$$

$$\mathbf{M}_0 \quad \vdash \varphi \triangleright \psi \rightarrow (\Diamond \varphi \wedge \Box \chi) \triangleright (\psi \wedge \Box \chi)$$

$$\mathbf{W}^* \quad \vdash \varphi \triangleright \psi \rightarrow (\psi \wedge \Box \chi) \triangleright (\psi \wedge \Box \chi \wedge \Box \neg \psi)$$

$$\mathbf{P}_0 \quad \vdash \varphi \triangleright \Diamond \psi \rightarrow \Box(\varphi \triangleright \psi)$$

$$\mathbf{R} \quad \vdash \varphi \triangleright \psi \rightarrow \neg(\varphi \triangleright \neg \chi) \triangleright \psi \wedge \Box \chi$$

---

<sup>28</sup>To be able to verify the desired properties of the Henkin-style construction of the interpretation one needs the technology of *shortening cuts* to make up for possible lack of induction.



The principles  $W$ ,  $M_0$ ,  $W^*$  and  $P_0$  were discovered by Albert Visser, see [95] or [56]. The principle  $R$  was recently discovered by Joost Joosten and Evan Goris [55, 57].

There are many interesting particular theories which are neither reflexive nor finitely axiomatizable. For all of them the question of characterization of their interpretability logic is open. One such example is the *primitive recursive arithmetic* PRA.

**Problem 24** *Characterize the interpretability logic of PRA.*

Lev Beklemishev has observed that this logic is strictly stronger than  $IL_{all}$  (see [96]). Joost Joosten [55] found an appropriate frame condition for the principle found by Beklemishev and studied its relationship with the other known principles, including the so-called Zambella principle (see below).

There are various other interpretations of the  $\triangleright$  that give us interesting logics. We present two possibilities. First there is the notion of *local interpretability*. The theory  $U$  is locally interpretable in  $T$ , or  $T \triangleright^{loc} U$ , if every finite subtheory of  $U$  is interpretable in  $T$ . Both between finitely axiomatized reasonable theories and between essentially reflexive reasonable theories interpretability and local interpretability coincide. However, one may provide examples where the two kinds differ. We may ask the following question.

**Problem 25** *Let  $IL_{all}^{loc}$  be the logic of the principles for local interpretability valid in all reasonable theories. Characterize  $IL_{all}^{loc}$ .*

The second notion is  $\Pi_1$ -conservativity.  $U$  is  $\Pi_1$ -conservative over  $T$ , or  $T \triangleright_{\Pi_1} U$ , if for all  $\Pi_1$ -sentences  $\pi$ ,  $U \vdash \pi \Rightarrow T \vdash \pi$ . Petr Hájek and Franco Montagna show that ILM is complete for arithmetical interpretations for  $\Pi_1$ -conservativity in  $\Sigma_1$ -sound extensions of  $IS_1$ , see their paper [46]. This has been improved in [14] to extensions of the rather weak parameter-free induction schema  $III_1^-$ .

In the case of Primitive Recursive Arithmetic, PRA, we also get more principles. This insight is due to Domenico Zambella. Grigori Mints proved that Zambella's principle is verifiable in PRA itself, see [14]. So, a salient question is the following.

**Problem 26** *What is the  $\Pi_1$ -conservativity logic of PRA?*

Konstantin Ignatiev [51] studied conservativity notions for larger formula classes  $\Pi_n$  and  $\Sigma_n$ , for  $n \geq 1$ . He characterized the corresponding logics for almost all classes, with the exception of  $\Sigma_1$  and  $\Sigma_2$ .

**Problem 27** *Characterize the logics of  $\Sigma_1$ - and  $\Sigma_2$ -conservativity over a sufficiently strong fragment of arithmetic.*

## 9 Graded provability algebras

Graded provability algebras link provability logic with two other traditions in proof theory. The first one is the study of transfinite progressions of axiomatic theories by iterated reflection schemata, which goes back to Turing [89] and Feferman [31]. The second one is the ordinal analysis tradition that stems from the work of Gentzen [35, 36]. The goals are to gain insight into the results on proof-theoretic analysis and ordinal notation systems from a more abstract perspective, develop a framework into which different kinds of analyses naturally fit in, and ultimately approach the fundamental questions such as the problem of natural ordinal notations.

Main results in this direction achieved so far concern the analysis of Peano arithmetic and its fragments [11, 12]. In particular, using provability-algebraic methods a new consistency proof for Peano arithmetic by transfinite induction à la Gentzen was given. A characterization of provably total computable functions of PA and an interesting combinatorial independent principle were also derived from the graded provability algebra approach.

The main structure one deals with is the Lindenbaum algebra of a theory  $T$  enriched by operators  $\langle n \rangle$ , for each natural number  $n$ , which stand for  $n$ -consistency:  $\langle n \rangle \varphi$  is the arithmetization of the statement that the theory  $(T + \varphi + \text{all true } \Pi_n\text{-sentences})$  is consistent. This statement is also equivalent to the uniform  $\Pi_{n+1}$ -reflection principle for  $T + \varphi$ .

The structure  $\mathcal{M}_T^\infty = (\mathcal{B}_T, \langle 0 \rangle, \langle 1 \rangle, \dots)$  is called the *graded provability algebra of  $T$* .<sup>29</sup> Its identities constitute a polymodal provability logic **GLP** first formulated and studied by G. Japaridze [54] (see also [52] and [20]). The 0-generated subalgebra of this algebra, which can also be seen as the set of letterless formulas of **GLP** modulo provability in **GLP**, provides an ordinal notation system up to the ordinal  $\epsilon_0$ . The associated ordering

$$\varphi <_0 \psi \iff \mathbf{GLP} \vdash \psi \rightarrow \langle 0 \rangle \varphi$$

is isomorphic to the standard one for  $\epsilon_0$ , however it has a different term representation. This ordinal notation system suggested an interesting analog of Hercules–Hydra game, the so-called Worm game, which was studied in

---

<sup>29</sup>For the purpose of the discussion below we ignore the additional sorting structure of this algebra.

[12]. This game provides one of the simplest examples of combinatorial principles independent from PA.

Currently the main questions in this area concern possible generalizations of the notion of graded provability algebra to systems stronger than PA. Standard examples of such systems are:  $\text{ATR}_0$ , a fragment of the second order arithmetic with the induction axiom and arithmetical transfinite recursion schema. This system was formulated by H. Friedman and its proof-theoretic ordinal is  $\Gamma_0$ . Another prominent example is a mildly impredicative theory  $\text{KP}_\omega$ , Kripke–Platek set theory with the infinity axiom. These systems were important stages in the ongoing proof-theoretic research into the ever stronger fragments of set theory and analysis currently culminating in the work of M. Rathjen and T. Arai (see [69, 65] for motivations and an overview).

**Problem 28** *Develop generalizations of the notion of graded provability algebra suitable for proof-theoretic ordinal analysis of  $\text{ATR}_0$  and  $\text{KP}_\omega$ .*

The recent paper [11] makes the first step towards a suitable treatment of  $\text{ATR}_0$  by developing a provability-algebraic ordinal notation system up to  $\Gamma_0$ . Thus, we believe that the treatment of  $\text{ATR}_0$  and the other predicative systems by these methods is within reach.

The treatment of  $\text{KP}_\omega$  seems to be a bit more problematic. The success of graded provability algebra approach in the study of formal arithmetic was based on a well-known correspondence between induction and reflection schemata in fragments of PA. In the case of fragments of  $\text{KP}_\omega$  some such correspondences hold as well, however the currently known picture is far from complete. In fact, the analogy between set-theoretic and arithmetical reflection principles needs to be clarified. Therefore, as a first stage in solving Problem 28 one is confronted with the following question.

**Problem 29** *Develop versions of reflection principles suitable for axiomatizing fragments of  $\text{KP}_\omega$  over some weak basic set theory. Clarify the analogy between set-theoretic and arithmetical reflection principles.*

The latter statement deserves a comment. The idea of the analogy can be very simply explained as follows.<sup>30</sup> The arithmetical reflection principle asserts that if a sentence  $\varphi$  is provable, then  $\varphi$  is true. This statement

---

<sup>30</sup>Interestingly, Kreisel and Lévy [61] wrote that they could not agree on whether using the same term ‘reflection’ both for the arithmetical and the set-theoretical reflection principle was a mere figure of speech and the principles had anything to do with each other.

could be read dually: if  $\neg\varphi$  holds, then  $\neg\varphi$  is consistent, i.e.,  $\neg\varphi$  has a model. This is the familiar form of the set-theoretic reflection principle, except for the more specialized notion of model in set theory. In fact, different kinds of models (omega-models, beta-models, etc.) yield a much richer variety of set-theoretic reflection principles. It seems very plausible that the sought generalizations of the arithmetical graded provability algebras will be obtained by considering these higher reflection principles as new modal operators.

We note in passing that analogs of Problem 29 also make sense in two different contexts: intuitionistic arithmetic and bounded arithmetic. It seems interesting to find versions of reflection principle suitable for an axiomatization of, e.g., fragments  $S_2^i$  and  $T_2^i$  of bounded arithmetic  $S_2$ . Similarly, very little is known about fragments of Heyting arithmetic **HA**. Wolfgang Burr [22] suggested a hierarchy of formula classes and fragments which match classical fragments  $III_n$  of **PA**. Can these fragments be axiomatized by suitable reflection principles over intuitionistic version of **EA**?

Next we mention a simple modal logic question concerning the standard notion of (arithmetical) graded provability algebra.

**Problem 30** *Axiomatize the equational theory of the reduct of  $\mathcal{M}_T^\infty$  in the language with only the following operations:  $\top$ ,  $\wedge$ ,  $\langle n \rangle$ , for all  $n$ .*

Before presenting a motivation for this question, let us remark that the equational theory in question deals with formulas of the form  $\varphi(\vec{x}) = \psi(\vec{x})$ , where  $\varphi$  and  $\psi$  are terms in the above language. Hence, the equational theory is equivalent to a fragment of Japaridze logic **GLP** consisting of formulas of the form  $\varphi(\vec{x}) \leftrightarrow \psi(\vec{x})$ . We already know a lot about **GLP**, in particular that it is decidable, thus, we believe the problem to be easy. One valid principle of the equational theory is

$$\langle n \rangle(\varphi \wedge \langle m \rangle \psi) = \langle n \rangle \varphi \wedge \langle m \rangle \psi, \quad \text{for } n > m,$$

and we conjecture that it will be its principal modal axiom.

The reason to be interested in that question is twofold. Firstly, only the formulas of this restricted language seem to play a role in the provability-algebraic analysis of Peano arithmetic. Thus, using the language from the outset may further simplify some proofs.

Secondly, this language admits a wider class of arithmetical interpretations. Propositional variables can now be understood as possibly infinitely

axiomatized (but elementary presented) arithmetical theories.<sup>31</sup> The treatment of infinitely axiomatized theories in the context of graded provability algebras seems to be useful for possible generalizations to stronger theories. For example, PA is naturally represented as a filter generated by the elements  $\{\langle 0 \rangle \top, \langle 1 \rangle \top, \langle 2 \rangle \top, \dots\}$ . However, we cannot naturally define  $\text{Con}(\text{PA})$  in the standard graded provability algebra of EA. To do so, one wants to be able to legally use expressions such as  $\langle 0 \rangle \{\langle n \rangle \top : n < \omega\}$  or  $\forall n \langle 0 \rangle \langle n \rangle \top$ .

A difficulty with generalizing this basic example is that the meaning of the  $\langle n \rangle$  operator applied to an infinite theory depends on a representation (numeration) of the theory rather than the theory itself. This is precisely the same problem that lead to serious difficulties with the program of classifying arithmetical sentences by transfinite progressions of iterated reflection principles. However, the advantage of graded provability algebra is that it provides numerations to sets of its elements and it has its own very weak language with particularly simple notion of definability. Thus, the problem might be resolved by allowing the use of expressions of the form  $\langle n \rangle X$ , where  $X$  is a definable set of elements of the graded provability algebra. The meaning of the term ‘definable’ has to be made more precise here, but a possible candidate could be the notion of definability by purely existential formulas (in the language of graded provability algebra). Thus, we come to the following challenging question.

**Problem 31** *Develop a definability theory for graded provability algebras that would allow the modalities to be applicable to (definably) infinite sets of elements.*

A number of other, more technical, questions concerning graded provability algebras are known. One such question concerns the topological semantics for Japaridze logic **GLP**. A nice topological semantics for **GL** was given independently by Merab Abashidze [1] and Andreas Blass [17]. The diamond operator can be interpreted as a derived set operator in scattered topological spaces. In fact, a completeness theorem for **GL** was proved w.r.t. the order topology of the ordinal  $\omega^\omega$ . In the case of **GLP** the situation is more difficult, as this system is not per se Kripke complete. However, there is a closely related subsystem **GLP**<sup>−</sup> which is (see [11]).

**Problem 32** *Is there a natural topological semantics for **GLP** and **GLP**<sup>−</sup>?*

---

<sup>31</sup>In the presence of negation there is a question how to interpret negation of an infinitely axiomatized theory. However, in the restricted language all formulas are positive.

Another interesting question concerns the decidability of the elementary theory of the 0-generated subalgebra of the graded provability algebra of **PA**, which is isomorphic to the Lindenbaum algebra of the letterless fragment of **GLP**. For the language with one modality the corresponding theory is mutually interpretable with the weak monadic second order theory of order  $(\omega, <)$ , commonly denoted **WS1S**, which is decidable by a well-known theorem of Büchi [5]. In the case of several modal operators the question becomes more difficult, but also more interesting.

**Problem 33** *Is the elementary theory of the 0-generated subalgebra of the graded provability algebra of **PA** decidable?*

The following list summarizes the problems mentioned in the paper. Their numeration is slightly different from the one used in the main text.

## 10 List of problems

### Informal concepts of proof

- P1. (“Hilbert’s 24-th”) What is the simplest proof of a given theorem?
- P2. (Kreisel, equivalence of proofs) What proofs are essentially the same, i.e., represent the same informal proof?
- P3. Find a proof system, or a class of proof systems, in which every (informal) mathematical proof can be faithfully represented on its own level of abstraction and complexity.
- P4. (Coordinate-free proof theory) Develop the theory of proofs on a sufficiently abstract axiomatic basis.
- P5. Develop alternative, non-deductive models of provability.

### Intuitionistic arithmetic

- P6. The provability logic of Heyting arithmetic **HA**: decidability, axiomatization.
- P7. Characterize the propositional logic of **HA** + **MP** + **ECT**<sub>0</sub>.
- P8. (Markov) Characterize the propositional logics of Kleene realizability.

- P9. (Plisko) Characterize the propositional logics of Gödel's *Dialectica* interpretation.
- P10. Suppose  $\text{HA} + A$  is consistent. Is it always the case that the propositional logic of  $\text{HA} + A$  is IPC?
- P11. What are the propositional admissible rules of  $\text{HA} + \text{MP}$  and of  $\text{HA} + \text{ECT}_0$ ?
- P12. Extend the language of propositional logic with a second sort of propositional variables  $s_1, s_2, \dots$ . Realizations send ordinary variables to arithmetical sentences and the new variables to  $\Sigma_1$ -sentences. Characterize the rules for this language admissible in  $\text{HA}$ .
- P13. What is the intersection of all provability logics for recursively enumerable extensions of  $\text{HA}$  in the language of  $\text{HA}$ ?
- P14. What are the closed fragments of the provability logics of  $\text{HA} + \text{MP}$  and of  $\text{HA} + \text{MP} + \text{ECT}_0$ ?

#### **Bounded arithmetic**

- P15. The provability logic of bounded arithmetics  $\text{S}_2^1$  and  $\text{S}_2$ : decidability, axiomatization.
- P16. Does the Friedman–Goldfarb–Harrington principle hold in  $\text{S}_2^1$ ?
- P17. Does  $\exists \Sigma_1^b$ -disjunction property for a  $\Sigma_1^b$ -presented theory  $T$  imply its  $\exists \Sigma_1^b$ -reflection principle in  $\text{S}_2^1 + \text{Con}(T)$ ?

#### **Bimodal and polymodal logics**

- P18. Classification of bimodal provability logics for pairs of r.e. theories containing a sufficiently strong fragment of  $\text{PA}$ .
- P19. Characterize the provability logics of any natural pair of theories  $(T, U)$  such that  $U$  is a  $\Pi_1$ -conservative extension of  $T$ , but  $U$  is not conservative over  $T$  w.r.t. boolean combinations of  $\Sigma_1$ -sentences.
- P20. What are the bimodal provability logics of  $(I\Sigma_1, I\Pi_2^-)$  and other pairs of incomparable fragments of  $\text{PA}$ ?
- P21. Give an effective description of all possible provability logics with a constant for a single sentence over  $\text{PA}$ .

### **Magari algebras and Lindenbaum Heyting algebras**

- P22. Is the  $\forall^*\exists^*$ -fragment of the first order theory of the provability algebra of PA decidable? The same question for the  $\forall^*\exists^*\forall^*$ -fragment.
- P23. Give sharp necessary and sufficient conditions for the isomorphism of Magari algebras of reasonable theories. In particular, are the algebras of PA and  $\text{PA} + \text{Con}(\text{PA})$  isomorphic?
- P24. Characterize the first-order definable elements in the Magari algebra of PA.
- P25. Characterize (r.e.) subalgebras of the Lindenbaum Heyting algebra of HA.
- P26. Is the elementary theory of the Lindenbaum Heyting algebra of HA decidable? Which fragments of it are?
- P27. Are the Lindenbaum Heyting algebras of HA and  $\text{HA} + \text{RFN}(\text{HA})$  isomorphic?

### **Interpretability logic and its kin**

- P28. Characterize the interpretability logic of all reasonable theories.
- P29. (Ignatiev) Characterize the logics of  $\Sigma_1$ - and  $\Sigma_2$ -conservativity over PA.
- P30. Characterize the interpretability logic and the  $\Pi_1$ -conservativity logic for PRA.
- P31. Characterize the logic of the principles for local interpretability valid in all reasonable theories.

### **Graded Provability Algebras**

- P32. Develop generalizations of the notion of graded provability algebra suitable for the proof-theoretic analysis of  $\text{ATR}_0$  and  $\text{KP}_\omega$ .
- P33. Develop versions of reflection principles suitable for axiomatizing fragments of  $\text{KP}_\omega$  over some weak basic set theory. Clarify the analogy between set-theoretic and arithmetical reflection principles.
- P34. The same question for bounded arithmetic theories  $S_2^i$  and  $T_2^i$  over PV and for the fragments of Heyting arithmetic HA.



- P35. Axiomatize the equational theory of the reduct of graded provability algebra in the language with only the following operations:  $\top$ ,  $\wedge$ ,  $\langle n \rangle$ , for all  $n$ .
- P36. Develop a definability theory for graded provability algebras that would allow the modalities to be applicable to (definably) infinite sets of elements.
- P37. Find a new combinatorial independent principle that would be motivated by Kripke models for **GLP**.
- P38. Is there a natural topological semantics for **GLP**?
- P39. Is the elementary theory of the 0-generated subalgebra of the graded provability algebra of **PA** decidable?

## References

- [1] M.A. Abashidze. Ordinal completeness of the Gödel–Löb modal system. In *Intensional Logics and Logical Structure of Theories: Material from the fourth Soviet–Finnish Symposium on Logic, Telavi, May 20–24, 1985*, pages 49–73. Metsniereba, Tbilisi, 1988. In Russian.
- [2] S. Abramsky. Algorithmic game semantics. A tutorial introduction. In H. et al. Schwichtenberg, editor, *Proof and system-reliability. Proceedings of the NATO Advanced Study Institute, Marktoberdorf, Germany, July 24 - August 5, 2001*, pages 21–47. Kluwer Academic Publishers. NATO Sci. Ser. II, Math. Phys. Chem. 62, Dordrecht, 2002.
- [3] S. Artemov. Explicit provability and constructive semantics. *Bulletin of Symbolic Logic*, 7(1):1–36, 2001.
- [4] S.N. Artemov. Applications of modal logic in proof theory. In *Questions of Cybernetics: Nonclassical logics and their applications*, pages 3–20. Nauka, Moscow, 1982. In Russian.
- [5] S.N. Artëmov and L.D. Beklemishev. On propositional quantifiers in provability logic. *Notre Dame Journal of Formal Logic*, 34:401–419, 1993.
- [6] S.N. Artemov and L.D. Beklemishev. Provability logic. In D. Gabbay and F. Guenther, editors, *Handbook of Philosophical Logic, 2nd ed.*, volume 13, pages 229–403. Kluwer, Dordrecht, 2004.

- [7] J. Barwise and J. Etchemendy. Visual information and valid reasoning. In W. Zimmerman and S. Cunningham, editors, *Visualization in teaching and learning mathematics*, pages 9–24. Mathematical Association of America, Washington, D.C., 1990.
- [8] 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.
- [9] L.D. Beklemishev. On bimodal logics of provability. *Annals of Pure and Applied Logic*, 68(2):115–160, 1994.
- [10] L.D. Beklemishev. Bimodal logics for extensions of arithmetical theories. *Journal of Symbolic Logic*, 61(1):91–124, 1996.
- [11] L.D. Beklemishev. Provability algebras and proof-theoretic ordinals, I. *Annals of Pure and Applied Logic*, 128:103–123, 2004.
- [12] L.D. Beklemishev. The Worm principle. Logic Group Preprint Series 219, University of Utrecht, March 2003. <http://preprints.phil.uu.nl/lgps/>.
- [13] L.D. Beklemishev, M. Pentus, and N. Vereshchagin. Provability, complexity, grammars. *American Mathematical Society Translations, Series 2*, 192, 1999.
- [14] L.D. Beklemishev and A. Visser. On the limit existence principles in elementary arithmetic and related topics. Technical Report LGPS 224, Department of Philosophy, Utrecht University, 2004.
- [15] A. Berarducci. The interpretability logic of Peano arithmetic. *The Journal of Symbolic Logic*, 55:1059–1089, 1990.
- [16] A. Berarducci and R. Verbrugge. On the provability logic of bounded arithmetic. *Annals of Pure and Applied Logic*, 61:75–93, 1993.
- [17] A. Blass. Infinitary combinatorics and modal logic. *Journal of Symbolic Logic*, 55:761–778, 1990.
- [18] A. Blass and Y. Gurevich. Algorithms vs. Machines. *Bulletin of the European Association for Theoretical Computer Science*, 77:96–118, June 2002.

- [19] A. Blass and Y. Gurevich. Algorithms: A Quest for Absolute Definitions. *Bulletin of the European Association for Theoretical Computer Science*, 81:195–225, Oct. 2003.
- [20] G. Boolos. *The Logic of Provability*. Cambridge University Press, Cambridge, 1993.
- [21] G. Boolos and G. Sambin. Provability: the emergence of a mathematical modality. *Studia Logica*, 50(1):1–23, 1991.
- [22] W. Burr. Fragments of Heyting arithmetic. *Journal of Symbolic Logic*, 65(3):1223–1240, 2000.
- [23] S. Buss. *Bounded arithmetic*. Bibliopolis, Napoli, 1986.
- [24] T. Carlson. Modal logics with several operators and provability interpretations. *Israel Journal of Mathematics*, 54:14–24, 1986.
- [25] 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.
- [26] D.H.J. de Jongh. The maximality of the intuitionistic predicate calculus with respect to Heyting’s Arithmetic. *The Journal of Symbolic Logic*, 36:606, 1970.
- [27] D.H.J. de Jongh and F. Veltman. Provability logics for relative interpretability. In [64], pages 31–42, 1990.
- [28] D.H.J. de Jongh and A. Visser. Embeddings of Heyting algebras. In [48], pages 187–213, 1996.
- [29] K. Došen. Identity of proofs based on normalization and generality. *Bulletin of Symbolic Logic*, 9:477–503, 2003.
- [30] S. Feferman. Arithmetization of metamathematics in a general setting. *Fundamenta Mathematicae*, 49:35–92, 1960.
- [31] S. Feferman. Transfinite recursive progressions of axiomatic theories. *Journal of Symbolic Logic*, 27:259–316, 1962.
- [32] 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, Heidelberg, New York, 1973.
- [33] H. Friedman. The disjunction property implies the numerical existence property. *Proc. Nat. Acad. USA*, 72:2877–2878, 1975.
  - [34] Yu. V. Gavrilenko. Recursive realizability from the intuitionistic point of view. *Soviet Mathematical Dokl.*, 23:9–14, 1981.
  - [35] G. Gentzen. Die Widerspruchsfreiheit der reinen Zahlentheorie. *Math. Ann.*, 112(4):493–565, 1936.
  - [36] G. Gentzen. Neue Fassung des Widerspruchsfreiheitsbeweises für die reine Zahlentheorie. *Forschungen zur Logik und Grundlegung der exakten Wissenschaften*, 4:19–44, 1938.
  - [37] S. Ghilardi. Unification in intuitionistic logic. *Journal of Symbolic Logic*, 64:859–880, 1999.
  - [38] J.-Y. Girard. Linear logic. *Theoretical Computer Science*, 50:1–102, 1987.
  - [39] J.-Y. Girard. *Proof theory and logical complexity, Vol. 1*. Bibliopolis, Napoli, 1987.
  - [40] J.-Y. Girard. Locus solum: From the rules of logic to the logic of rules. *Math. Struct. Comput. Sci.*, 11:301–506, 2001.
  - [41] J.-Y. Girard, Y. Lafont, and P. Taylor. *Proofs and Types*. Cambridge University Press, 1989.
  - [42] K. Gödel. Eine Interpretation des intuitionistischen Aussagenkalküls. *Ergebnisse Math. Kolloq.*, pages 39–40, 1933.
  - [43] O. Goldreich. *Zero-knowledge twenty years after its invention*. Tutorial, URL: <http://www.wisdom.weizmann.ac.il/~oded/zk-tut02.html>, 2004.
  - [44] S. Goldwasser, S. Micali, and C. Rackoff. The knowledge complexity of interactive proof systems. *SIAM Journal on Computing*, 18:186–208, 1989.
  - [45] S.V. Goryachev. On interpretability of some extensions of arithmetic. *Mat. Zametki*, 40:561–572, 1986. In Russian.

- [46] P. Hájek and F. Montagna. The logic of  $\Pi_1$ -conservativity. *Archiv für Mathematische Logik und Grundlagenforschung*, 30:113–123, 1990.
- [47] P. Hájek and P. Pudlák. *Metamathematics of First Order Arithmetic*. Springer-Verlag, Berlin, Heidelberg, New York, 1993.
- [48] W. Hodges, M. Hyland, C. Steinhorn, and J. Truss, editors. *Logic: from foundations to applications*. Clarendon Press, Oxford, 1996.
- [49] R. Iemhoff. On the admissible rules of intuitionistic propositional logic. *Journal of Symbolic Logic*, 66(1):281–294, 2001.
- [50] R. Iemhoff. *Provability logic and admissible rules*. PhD thesis, University of Amsterdam, Amsterdam, 2001.
- [51] K.N. Ignatiev. Partial conservativity and modal logics. ITLI Prepublication Series X-91-04, University of Amsterdam, 1991.
- [52] K.N. Ignatiev. On strong provability predicates and the associated modal logics. *Journal of Symbolic Logic*, 58:249–290, 1993.
- [53] G. Japaridze and D. de Jongh. The logic of provability. In S. Buss, editor, *Handbook of proof theory*, pages 475–546. North-Holland Publishing Co., amsterdam edition, 1998.
- [54] G.K. Japaridze. The modal logical means of investigation of provability. Thesis in Philosophy, in Russian, Moscow, 1986.
- [55] J.J. Joosten. *Interpretability formalized*. PhD thesis, University of Utrecht, Utrecht, the Netherlands, 2004.
- [56] J.J. Joosten and A. Visser. The interpretability logic of *all* reasonable arithmetical theories. *Erkenntnis*, 53(1–2):3–26, 2000.
- [57] J.J. Joosten and A. Visser. How to derive principles of interpretability logic. A toolkit. In J. van Benthem, A. Troelstra, F. Veltman, and A. Visser, editors, *Liber Amicorum for Dick de Jongh*. ILLC, (<http://www.illc.uva.nl/D65/>), 2004.
- [58] A. Kolmogorov and V. Uspensky. On the definition of algorithm. *Uspekhi Mat. Nauk*, 13(4):3–28, 1958. In Russian. English translation in *AMS Translations* 29 (1963), 217–245.
- [59] J. Krajíček. *Bounded arithmetic, Propositional logic, and Complexity theory*. Cambridge University Press, Cambridge, 1995.

- [60] G. Kreisel. A survey of proof theory. II. In *Proc. 2nd Scandinav. Logic Sympos. 1970, Studies Logic Foundations Math. 63*, pages 109–170. 1971.
- [61] 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.
- [62] 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.
- [63] Y. Moschovakis. What is an algorithm? In B. Engquist and W. Schmid, editors, *Mathematics Unlimited*, pages 919–936. Springer Verlag, Berlin, 2001.
- [64] P.P. Petkov, editor. *Mathematical logic, Proceedings of the Heyting 1988 summer school in Varna, Bulgaria*. Plenum Press, Boston, 1990.
- [65] 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.
- [66] H. Prakken. *Logical Tools for Modelling Legal Argument. A Study of Defeasible Reasoning in Law*. Dordrecht, 1997.
- [67] D. Prawitz. Ideas and results in proof theory. In *Proc. 2nd Scandinav. Logic Sympos. 1970, Studies Logic Foundations Math. 63*, pages 235–307. 1971.
- [68] N. Preining. Sketch-as-proof. In G. Gottlob, A. Leitsch, and D. Mundici, editors, *Lecture Notes in Computer Science 1289. Computational Logic and Proof Theory, 5th K. Gödel Colloquium KGC’97, Proceedings*, pages 264–277. Springer-Verlag, Berlin, 1997.
- [69] M. Rathjen. Recent advances in ordinal analysis:  $\Pi_2^1$ -CA and related systems. *Bulletin of Symbolic Logic*, 1(4):468–485, 1995.
- [70] G.F. Rose. Propositional calculus and realizability. *Transactions of the American Mathematical Society*, 61:1–19, 1953.
- [71] V.V. Rybakov. A criterion for admissibility of rules in the modal system **S4** and intuitionistic logic. *Algebra and Logic*, 23:369–384, 1984.

- [72] V.V. Rybakov. *Admissibility of logical inference rules*. Elsevier, Amsterdam, 1997.
- [73] A. Schönhage. Storage modification machines. *SIAM Journal on Computing*, 9:490–508, 1980.
- [74] V.Yu. Shavrukov. The logic of relative interpretability over Peano arithmetic (in Russian). Technical Report Report No.5, Stekhlov Mathematical Institute, Moscow, 1988.
- [75] V.Yu. Shavrukov. A note on the diagonalizable algebras of PA and ZF. *Annals of Pure and Applied Logic*, 61:161–173, 1993.
- [76] V.Yu. Shavrukov. Subalgebras of diagonalizable algebras of theories containing arithmetic. *Dissertationes Mathematicae*, 323, 1993.
- [77] V.Yu. Shavrukov. Isomorphisms of diagonalizable algebras. *Theoria*, 63(3):210–221, 1997.
- [78] V.Yu. Shavrukov. Undecidability in diagonalizable algebras. *Journal of Symbolic Logic*, 62(1):79–116, 1997.
- [79] T. Smiley. The logical basis of ethics. *Acta Philosophica Fennica*, 16:237–246, 1963.
- [80] C. Smoryński. Applications of Kripke Models. In A.S. Troelstra, editor, *Metamathematical Investigations of Intuitionistic Arithmetic and Analysis*, Springer Lecture Notes 344, pages 324–391. Springer, Berlin, 1973.
- [81] C. Smoryński. Fifty years of self-reference. *Notre Dame Journal of Formal Logic*, 22:357–374, 1981.
- [82] C. Smoryński. *Self-Reference and Modal Logic*. Springer-Verlag, Berlin, 1985.
- [83] R.M. Solovay. Provability interpretations of modal logic. *Israel Journal of Mathematics*, 28:33–71, 1976.
- [84] G. Sundholm. Proofs as acts versus proofs as objects: Some questions for Dag Prawitz. *Theoria*, 64:187–216, 1998.
- [85] V. Švejdar. Modal analysis of generalized Rosser sentences. *The Journal of Symbolic Logic*, 48:986–999, 1983.

- [86] A. Tarski, A. Mostowski, and R.M. Robinson. *Undecidable theories*. North-Holland, Amsterdam, 1953.
- [87] A. Troelstra. *Metamathematical investigations of intuitionistic arithmetic and analysis*. Springer Lecture Notes 344. Springer-Verlag, Berlin, 1973.
- [88] A. Troelstra and D. van Dalen. *Constructivism in Mathematics, vols 1, 2*. North-Holland, Amsterdam, 1988.
- [89] A.M. Turing. System of logics based on ordinals. *Proc. London Math. Soc.*, ser. 2, 45:161–228, 1939.
- [90] R. Verbrugge. *Efficient Metamathematics*. PhD thesis, University of Amsterdam, Amsterdam, 1993.
- [91] A. Visser. *Aspects of Diagonalization and Provability*. PhD thesis, University of Utrecht, Utrecht, The Netherlands, 1981.
- [92] A. Visser. On the completeness principle. *Annals of Mathematical Logic*, 22:263–295, 1982.
- [93] A. Visser. Evaluation, provably deductive equivalence in Heyting’s Arithmetic of substitution instances of propositional formulas. Logic Group Preprint Series 4, Department of Philosophy, Utrecht University, Heidelberglaan 8, 3584 CS Utrecht, 1985.
- [94] A. Visser. Interpretability logic. In [64], pages 175–209, 1990.
- [95] A. Visser. The formalization of interpretability. *Studia Logica*, 51:81–105, 1991.
- [96] A. Visser. An overview of interpretability logic. In M. Kracht, M. de Rijke, H. Wansing, and M. Zakhariashev, editors, *Advances in Modal Logic, v.1, CSLI Lecture Notes, No. 87*, pages 307–359. CSLI Publications, Stanford, 1998.
- [97] A. Visser. Rules and arithmetics. *Notre Dame Journal of Formal Logic*, 40(1):116–140, 1999.
- [98] A. Visser. Substitutions of  $\Sigma_1^0$ -sentences: Explorations between intuitionistic propositional logic and intuitionistic arithmetic. *Annals of Pure and Applied Logic*, 114(1–3):227–271, 2002.



- [99] A. Visser. Faith and falsity. *Annals of Pure and Applied Logic*, 131:103–131, 2005.
- [100] 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, no. 53, pages 289–326. Center for the Study of Language and Information, Stanford, 1995.