Bibliographical Remarks and Further Reading

Preliminaries

We try to indicate for each important result or notion its author and the corresponding publication, and possibly also reference to another work where the result is presented. Our aim is to be as precise as possible; on the other hand, these remarks are not intended to be a complete historical source and they serve only for orientation. This concerns mainly remarks on old results. The reader interested in deeper investigation of origins of the metamathematics of arithmetic is referred to source books: Gödel's Collected Works I [Gödel 86], the books From Frege to Gödel [van Heijenort 87] and The Undecidable [Davis 65]. We find also [Meschkowski 81] very informative. [Smoryński 91 – Logical] contains rather detailed historical information.

The material covered in the Preliminaries belongs to classics and can be found in standard monographs on mathematical logic, notably [Shoenfield 67], [Mendelson 64], [Bell-Machover 77] and others; note that the syntax of first order predicate logic is systematically developed in [Hilbert-Ackermann 28], largely in the style which is still in use. They formulate the problem of completness; Gödel presents his solution in his dissertation (1930) published as [Gödel 31 – Monatsh.]. [Skolem 20] contains a proof of what we now call the downward Löwenheim-Skolem theorem, using what we call Skolemization. The above works seem to deal with validity and satisfiability as intuitively clear notions without trying to formalize them. Tarski's paper ([Tarski 33] in Polish; a German version is [Tarski 36], for an English version see [Tarski 56]) presents conditions defining the satisfaction of a compound formula from the satisfaction of its components – these are our "Tarski's truths conditions" or "Tarski's satisfaction conditions" (for a detailed analysis of Tarski's approach see [Hodges 85]).

Herbrand's theorem is contained in his thesis published as [Herbrand 30]; as commented in [van Heijenort 67] p. 526, Herbrand considered his theorem to be a more precise statement of the well-known Löwenheim-Skolem theorem. From the above mentioned monographs, only [Shoenfield 67] elaborates on Herbrand's theorem.

Origins of first order arithmetic were described in the Introduction; let us add some details. Arithmetical hierarchy was introduced in [Kleene 43] and [Mostowski 47]; the notation Σ_n, Π_n, Δ_n is due to [Addison 58] and [Mostowski 59] (cf. Kleene's introductions in [Gödel 86]). Definition of functions from other functions by primitive recursion were known to Dedekind, Skolem, Hilbert and Ackermann. Gödel introduces in [Gödel 31 – Monatsh.] a class of functions that he calls recursive; in our present terminology, these are just primitive recursive functions. For origins of general recursive functions see [Péter 34], [Kleene 36]; monographs on recursive functions and recursion theory include [Rogers 67], [Soare 87]. Gödel introduces and uses coding of finite sequences of natural numbers by natural numbers [Gödel 31] and arithmetization of syntax. We shall comment more on arithmetization in remarks on Chaps. I and III.

Matiyasevič's theorem is a final step of a long research into Hilbert's tenth problem in which J. Robinson, M. Davis and H. Putnam played prominent roles; besides Matiyasevič's papers, see [Davis 73 Am. Math. Monthly] for history and a detailed proof. Finally let us note that the decidability of the set of all sentences of the [language of arithmetic without multiplication] true in N mentioned in 0.49 is proved in [Presburger 30]. Similarly, [Skolem 30] proved that the theory of N in the language with multiplication and equality is decidable; investigation of the theory of multiplication has been continued, see [Nadel 80], [Cegielsky 81].

Chapter I

(1) The theory Q was introduced by R.M. Robinson and is developed in [Tarski, Mostowski, Robinson 53]. First-order arithmetic is mentioned in Gödel's postcript [Gödel 32 – Ergebnisse] to his main paper on incompleteness [Gödel 31]; [Skolem 23] develops a first order quantifier free system known presently as primitive recursive arithmetic. His paper is devoted to "positive results": no metamathematical results are obtained. Coming back to the question why first order arithmetic with induction for all formulas is commonly called Peano arithmetic, let us mention that the first occurence of this term known to us is in [Tarski-Mostowski-Robinson 53] where the authors say that the theory in question "is known as Peano arithmetic". [Kleene 52] studies the same system (or an equivalent system); in an introduction in [Gödel 86] Kleene says that his system "can be described as Peano arithmetic PA". Kleene stresses the necessity of distinguishing PA from the original Peano's axioms that are second-order.

In developing I_{open} we freely follow [Mendelson 64]; but we leave I_{open} rather soon. It should be mentioned that there are several papers devoted to the metamathematics of I_{open}, notably [Shepherdson 65], [Wilkie 77], [van der Dries 80], [Adamowicz 86, 87] and others.

$I\Sigma_0$ and related systems of bounded arithmetics are studied extensively in Chap. V (see bibliographical remarks to that chapter). In developing $I\Sigma_1$ we partially anticipate what will be done in Sect. 2 for arbitrary $I\Sigma_n$ $(n > 1)$; the main task – arithmetization – is done in a rather standard way, usual in presenting Gödel's results on PA, and it is seen that $I\Sigma_1$ suffices to prove all necessary things. (The auxiliary coding of o-sequences follows [Shoenfield 67].) Partial truth definitions are of great importance; they were studied by [Montague 59] ([Schütte 60] has a similar notion). [Kreisel-Lévy 68] is an important paper dealing with partial truth definitions; we shall mention it in remarks to Chap. III.

(2) The main starting point for the study of theories $I\Sigma_n$ and $B\Sigma_{n+1}$ is the paper [Paris-Kirby 78]; this was preceded by [Parsons 70, 72] where some important partial results were obtained. The paper by Paris-Kirby contains among other things the equivalence of $I\Sigma_n$, $I\Pi_n$, $L\Sigma_n$, $L\Pi_n$ (Th. 2.4) and the implications $I\Sigma_{n+1} \Rightarrow B\Sigma_{n+1} \Rightarrow I\Sigma_n$, $B\Sigma_{n+1} \Leftrightarrow B\Pi_n$ (Th. 2.5). The fact (2.4) that $I\Sigma_n \Rightarrow I\Sigma_0(\Sigma_n)$ appears in [Clote 85 – Caracas] (with a remark that Paris and Kossak independently proved the same result); the paper also contains a proof of the implication $B\Sigma_{n+1} \Rightarrow L\Delta_{n+1} \Rightarrow I\Delta_{n+1}$ (2.5). The implication $L\Delta_{n+1} \Rightarrow B\Sigma_{n+1}$ was proved by R.O. Gandy (unpublished). Note that H. Friedman circulated a preprint (on fragments of Peano arithmetic, before 1985) in which he claimed $B\Sigma_{n+1}$ and $I\Delta_{n+1}$ to be equivalent; but it still seems to be open whether $I\Delta_{n+1}$ proves $B\Sigma_{n+1}$ (or, equivalently, $L\Delta_{n+1}$). The fact that in $B\Sigma_{n+1}$, Σ_n and Π_n formulas are closed under bounded quantification is due to [Parsons 70].

The notion "piecewise coded" is studied in [Clote 85 – Caracas] and was introduced by Kossak. Theorem 2.7 appears in the same paper. Concerning theorem 2.23: The equivalence $S\Sigma_{n+1} \Leftrightarrow S\Sigma_n \Leftrightarrow I\Sigma_n$ is in [Clote 85 – Caracas], properties of PHP in [Dimitracopoulos-Paris 86], properties of the axioms of regularity in [Mills-Paris 84]. 2.24 (1) seems to be a folklore, properties of local and partial approximations (T and P) appear in [Hájek-Paris 86]. Theorem 2.25 is due to V. Švejdar (unpublished).

Finite axiomatizability of $I\Sigma_n$, $B\Sigma_{n+1}$ ($n \geq 1$) is an easy consequence of the existence of partial satisfactions; it is difficult to say who first observed it.

Relativized satisfaction (for $\Sigma_n^{\bullet}(X)$-formulas, X being a set) is introduced and investigated in [Hájek-Kučera 89]. [Clote 85 – Oberwolfach] works with $\Sigma_n(X)$-definable subsets of a model M of $I\Sigma_1$, X being a subset of M; but he does not formalize relativized satisfaction inside $I\Sigma_1$. The development of a theory of low Δ_{n+1} sets and sets of the order type of the universe (2.63–2.71) follows Clote's work but is made *inside* the theories $B\Sigma_{n+1}$, not from outside using models. The strengthening to $I\Sigma_n$ and replacement of (low) Δ_{n+1} by (low) $\Sigma_0^*(\Sigma_n)$ is due to [Hájek-Kučera 89]. We shall comment more on low sets in remarks to Sect. (3).

Finally, arithmetic with a top is studied in [Dimitracopoulos-Paris 82] and [Cegielsky-McAloon-Wilmers 82]. In I.2 we have presented only rather elementary facts on this arithmetic.

(3) The limit theorem is a classical result of recursive theory due to [Putnam 65]. [Clote 85 – Caracas] proves the limit theorem for Δ_2 sets in $I\Sigma_1$ and for Δ_2 functions in $B\Sigma_2$. The Limit theorem for Δ_2 functions in $B\Sigma_1$ was proved by Švejdar (3.2 – unpublished).

The low basis theorem is a recursion-theoretic version of the celebrated König's lemma [König 27] and is due to [Jockusch-Soare 72 – TAMS]. In the recursion-theoretic language, a degree a is low if $a' = 0'$ (see e.g. [Soare 87]). It is easy to show that the Turing degree of a set X is $0'$ iff $X \in \Delta_2$ and that $dg(X)$ is low iff X is low Δ_2 in our meaning. The low basis theorem says that if T is a recursive dyadic infinite tree then T has an infininite low branch. [Clote 85] shows that this is true in each $M \models B\Sigma_2$; we follow his proof, working inside $B\Sigma_2$. The improvement to $I\Sigma_1$ and low $\Sigma_0^*(\Sigma_1)$ is in [Hájek-Kučera 88]. 3.24 was proved by Paris (unpublished). [Gaifman 70] was the first to realise that Matiyasevič's theorem is provable in *PA*; Dimitracopoulos proved Matiyasevič's theorem in $I\Sigma_0(\exp)$; cf. [Gaifman-Dimitracopoulos 82].

(4) As we already said in comments to the Preliminaries, the arithmetization of logic goes back to Gödel. The first systematic elaboration of logic inside first order arithmetic (arithmetization of syntax as well as semantics including the Arithmetized Completness Theorem) is contained in [Hilbert-Bernays 34, 39]. (Clearly, they do not have the Low Arithmetized Completness Theorem.) Checking that (some) usual proofs of logical facts formalize in $I\Sigma_1$ is more or less immediate (disregarding some difficulties as shown in the proof of 4.10). The Arithmetized Completness Theorem can be stated as follows. *PA* proves that each consistent Δ_1 theory has a full Δ_2 model. Partial improvements were obtained by [Paris 81] and [McAloon 78 – TAMS]; our Low Arithmetized Completness Theorem is more general.

Theorem 4.33 as well as Corollary 4.34 is rather important; it is closely related to the results of [Leivant 81]. Note that 4.34 (1) ($I\Sigma_{k+1} \vdash Con^{\bullet}(I\Sigma_k)$) can be obtained by proof-theoretical methods, see [Takeuti 75]. (We shall comment on proof theory in remarks to Chap. III).

Finally, Theorem 4.37 is extracted from [Paris 80] and will be used in Chap. II where results of Paris's paper will be presented in detail.

Chapter II

As mentioned in the introduction to the present chapter, the interest in developing finite combinatorics in *PA* and its fragments arose from metamathematical investigations by Paris and his school. Paris's pioneering contribution is twofold: he produced a mathematically well-understood combinatorial Π_2-theorem which is true but unprovable in *PA*, and he showed the unprovability by model-theoretic means, not by self-reference (as Gödel did when considering his unprovable true sentences). The original formulation of the principle now known as the Paris-Harrington principle was different from that presented here and did not use the notion of a large set. It circulated in preprints and may be found in

[Paris 78], [McAloon 79]. Harrington contributed by reformulating the principle into its elegant form using the notion of a (relatively) large finite set and showing the principle to be equivalent to $Con^{\bullet}(PA + Tr(\Pi_1))$ (cf. [Paris-Harringtron 77]); the last equivalence was independently proved by [McAloon 80 – rapports]. Investigation of instances of the Paris-Harrington principle and their relation to fragments of *PA* (as well as of instances of the principle of ordinal-large intervals, see below) is the main content of Paris's papers [Paris 80], [Paris 81]. Positive results on fragments contained in these two papers form the main topic of the present chapter.

(1) Theorem 1.9 in particular is due to Paris. But we do not present it following Paris (since he obtained the provabilities in question by indirect, model-theoretical means). We follow and elaborate proofs of [Clote 85 – Oberwolfach] based on the Low Basis Theorem. In particular, Theorems 1.5, 1.6, 1.7 are due to Clote. Clote also reasons model-theoretically but his arguments easily yield direct proofs in fragments. Furthermore, we improve Clote's results by using our stronger version of Low Basic Theorem. 1.10 is folklore.

(2) Here we elaborate the proof of [Paris 81]; the subsection (b) on combinatorics relies on [Harrington-Paris 77].

(3) The history of what we call the Schwichtenberg-Wainer hierarchy or fast growing hierarchy is as follows (see [Buchholz-Wainer 87]): [Kreisel 52 – non fin.] showed that the functions provably recursive in *PA* can all be defined by recursions on certain well orderings of type $< \varepsilon_0$. Later [Schwichtenberg 71] and [Wainer 70, 72] independently generalized earlier results of [Grzegorczyk 53] and [Robbin 65] to show that Kreisel's functions can be characterized by means of the present fast growing hierarchy below ε_0. As further reference we mention [Löb-Wainer 70], [Schwichtenberg 77], [Rose 84], [Buchholz 84].

[Ketonen and Solovay 81] related this hierarchy to the Paris-Harrington principle and established, using purely combinatorial means, sharp upper and lower bounds to the function

$$\sigma(n) = \min\{a \mid [0,a] \underset{*}{\rightarrow} (n+1)^n_n\}\,;$$

from this, they reproved Paris and Harrington's result. In their paper, Ketonen and Solovay introduced and studied the notion of an α-large finite set (due originally to Ketonen), α being an ordinal. It follows from their investigations that the principle

$$\text{(W)} \qquad (\forall \alpha \leq \varepsilon_o)(\forall x)(\exists y)([x,y] \text{ is } \alpha\text{-large})$$

is (meaningful and) unprovable in *PA*. [Paris 80, 81] introduced and investigated instances of (W) and related them to fragments of *PA*. We freely follow Paris's papers using [Ketonen-Solovay 81]. Theorem 3.18 does not occur explicitely in [Paris 81]; it can be found e.g. in [Takeuti 75], see also [Kurata 86].

There are various other important combinatorial principles that can be analyzed with respect to fragments, notably: a principle proposed independently by Pudlák (original paper unpublished, see [Hájek-Paris 86]), Kanamori-McAloon's principle [Kanamori-NcAloon 87], principles due to [Clote-McAloon 83] and possibly others.

Chapter III

(1) The informal notion of an interpretation of a theory in another one appears to be rather old; but we did not attempt to identify particular references. The first work dealing with interpretations in connection with systems of first order arithmetic is [Tarski-Mostowski-Robinson 53]; an old paper is [Montague 57]. Feferman's fundamental paper [Feferman 60] deals also with interpretability; [Montague 63] contains a model-theoretic characterization of interpretability (cf. also [Hájek 66]).

Similarly, we present no information about the origins of the notion of partial conservativity; but the first result concerning partial conservativity is [Kreisel 62] showing that

$\neg Con^{\bullet}(PA)$ is Π_1-conservative over PA. We shall give detailed references to later works below (point (4)).

Sequential theories were introduced by [Pudlák 85]; a similar notion was investigated by [Vaught 67]. Friedman also had a similar notion but did not present an exact definition, cf. [Smoryński 85 – Friedman's research].

ACA_0 is one of various interesting subsystems of second order arithmetic; we refer to Simpson's forthcoming book on this topic [Simpson 86–90]. The model-theoretic proof of conservativity of ACA_0 over PA (1.16) is a folklore; the proof of finite axiomatizability is a variant of Gödel's proof of the fact that GB proves the schema of comprehension [Gödel 40]. See also below.

The notions of a binumeration and numeration go back to [Gödel 31] and are studied in detail in [Feferman 60]. Gödel introduced also the notion of of an ω-consistent theory; this is what we called a sound theory (1.21). 1-consistent theories were introduced in [Feferman-Kreisel-Orey 62]. Rosser invented witness comparison [Rosser 36]; theorem 1.24 appears in [Tarski, Mostowski, Robinson 53].

(2) Gödel was first to construct a particular self-referential formula [Gödel 31 – Monatsh.]. The non-parametric version of the fixed-point theorems 2.1 occurs in [Carnap 34]; Gödel acknowledges Carnap's priority in [Gödel 34 – Princeton]. A parametric version occurs first in [Montague 62], where also existence of a self-referential pairs is proved (2.6). For a historical survey containing self-reference see [Smoryński 81 – fifty]. Corollary 2.3 is the celebrated result on the undefinability of truth, due to Tarski [Tarski 33, 35]. Tarski refers to [Gödel 31 – Monatsh.] for the method of diagonalization. Results of subsection (e) are classical: Gödel's main paper [Gödel 31 – Monatsh.] contains his first incompleteness theorem (as well as the statement of his second incompleteness theorem) for a system related to Whitehead-Russell's Principia Mathematica (1913); this paper was preceded by a short announcement [Gödel 30 – Anzeiger] and complemented by a short paper [Gödel 32 – Ergebnisse]; the last paper formulates Gödel's theorems for first order systems like our PA. Rosser's improvement (replacing ω-consistency by mere consistency) is in [Rosser 36]. Flexible formulas were invented by [Mostowski 61]; our proof follows [Kripke 62]. Gödel formulated his second incompletness theorem for a particular system and a particular proof predicate and did not need to formulate exciplit provability conditions. Provability conditions are formulated in [Hilbert-Bernays 39]. The fact that PA is not finitely axiomatizable (2.4) is due to [Ryll-Nardzewski 52]. For Löb's theorem and its motivation see [Löb 54], [Henkin 52]. A paper showing that PA cannot be axiomatized by formulas of limited arithmetical complexity is [Rabin 61]; this is a predecessor of 2.27. For Craig's theorem see [Craig 72].

Essential reflexivity of PA (2.35) was proved in [Feferman 60]. Our 2.37 is a slight improvement of a theorem of [Feferman 60]. Theorems characterizing interpretability in pure extensions of PA (2.34–2.40) are based on [Orey 59, 61], the equivalence 2.39 (i) $\Leftrightarrow$ (ii) being often called Orey's compactness theorem (cf. [Hájek 72], [Guaspari 76]).

The method of shortening definable cuts was discovered by Solovay; but he has never published the result. See [Hájek 81 – Int. II] and [Paris-Dimitracopoulos 83]; the latter paper proves that there are definable cuts having no shortening to a cut closed under exponentiation (see also [Dimitracopoulos 80]).

The rest of Sect. 3 (subsections (b)–(d)) is an elaboration of (parts of) [Pudlák 85]; only Theorem 3.20 (2) is due to Hájek. The idea of simultaneous use of two provability predicates (3.6), essential for 3.9, is due to Mycielski (unpublished).

(4) Section 4 presents results on interpretability and partial conservativity for theories containing $I\Sigma_1$; many of them were first obtained for PA. We shall refer to the original results even if they are less general than the theorem presented. An extended abstract of this section appeared as [Hájek 87 – CMUC].

4.5 (1) [Feferman 60] for $T \supseteq PA$; [Švejdar 78]. (2) [Kreisel 68] for $T \supseteq PA$; [Smoryński 80]. (3) [Kreisel 68] for $T \supseteq PA$; Švejdar (unpublished).

4.25 is due to Guaspari and Solovay (for $T \supseteq PA$): see [Guaspari 67] (their examples are more complicated than ours).

4.26 (1) is classical (cf. [Rogers 67]); (2) [Hájek 87]; (3) for $T = ZF$, $\Gamma = \Pi_1$ [Hájek 71], generalized for $T \supseteq PA$ by [Lindström 84]. (4) implicit in [Lindström 84] for $T \supseteq PA$.

4.27 For $T = ZF$, $\Gamma = \Pi_1$ Solovay (unpublished) for PA, $\Gamma = \Pi_n$ and *Consv* [Hájek 79]; for PA, $\Gamma = \Sigma_n$ and *Consv* [Quinsey 81]; in full generality for $T \supseteq PA$ [Lindström 84].

Fixed point theorem 4.29 [Shepherdson 60], [Smoryński 81 – Fifty]. Theorem 4.35 (Σ_1-numerability) was first obtained by [Shepherdson 60].

4.55 The first example of a Σ_2 formula φ such that $(ZF + \varphi)$ is interpretable in ZF but $(GB + \varphi)$ is not in GB is in [Hájek 71] under the assumption of soundness; this assumption was removed in [Hájek, Hájková 72]. A Σ_1 formula of the desired properties was constructed by Solovay.

4.57 first appeared in [Hájek 87]; 4.50 is from [Lindström 84].

*

We shall now comment on some important topics related to the contents of Chapter III but not covered. We shall indicate only basic references, not full literature on the subject.

(a) *Reflection principles.* Roughly, a reflection principle says "if a formula is provable then it is true" (or "if all instances of a formula are provable then the formula is true"). We met an example of a provable reflection principle in I.4.34; an example of an unprovable principle is the schema

$$Pr^{\bullet}_{PA}(\overline{\varphi}) \to \varphi$$

for all φ($Pr^{\bullet}_{PA}$ being given by a Σ_1-definition of (PA)) since its particular case (for φ being $0 = 1$) clearly implies $Con^{\bullet}_{PA}$. On the other hand, the last schema is clearly true in N and may be used as a natural strengthening of PA. This has been studied systematically; the reader many consult [Kreisel-Lévy 68], [Schmerl 79, 80, 82]; early results on reflection principles are also discussed in detail in Smoryński's survey paper [Smoryński 77] in the Handbook of Mathematical logic.

(b) *Transfinite progressions of theories.* One can iteratively add unprovable true sentences to a sound theory, e.g. investigate the sequence PA, $PA + Con^{\bullet}(PA^{\bullet})$, $PA + Con^{\bullet}(PA^{\bullet}) + Con^{\bullet}(PA^{\bullet} + \overline{Con(PA^{\bullet})}), \ldots$ or iteratively add some reflection principles. Using ordinal notations [Kleene 44] one can iterate transfinitely and try to characterize which part of the truth ($Th(N)$) can be grasped. See [Feferman 62], [Feferman-Spector 62], [Fenstad 68], [McAloon 80, 82].

Arithmetic and lattices: Lindström and Švejdar independently introduced and studied the lattice of intenpretability types over theories T containing some arithmetic ([Lindström 79 – Aalborg, 84 – Notre Dame], [Švejdar 78]). Put $\varphi \preccurlyeq \psi$ if $(T + \varphi)$ is interpretable in $(T + \psi)$; this is a quasiorder and the corresponding factors are interpretability types; they form a distributive lattice. Many deep results about this lattice are known, in particular for $T = PA$.

[Hájková 71] and [Palúch 79] investigated lattices of consistency statements: e.g. take PA and investigate all Σ_1 binumerations of PA in PA. For each such binumeration α let $Con^{\bullet}_{\alpha}$ be the corresponding consistency statement. Put $\alpha \leq \beta$ if $PA \vdash Con^{\bullet}_{\alpha} \to Con^{\bullet}_{\beta}$. This again gives an extremely rich lattice.

Investigation of both lattices uses extremely tricky self-referential constructions. Note that [Mycielski, Pudlák and Stern 90] investigate a lattice of interpretability types of all first order theories (not necessarily Δ_1 axiomatizable and not necessarily containing arithmetic); their lattice differs drastically from Lindström-Švejdar's lattice.

(d) *Faithful interpretations.* An interpretation $*$ of T in S is Γ-faithful if for each $\varphi \in \Gamma$, $T \vdash \varphi$ is equivalent to $S \vdash \varphi^*$. See [Feferman-Kreisel-Orey 62], [Guaspari 79] (model theoretic characterization), [Lindström 84].

(e) *Arithmetic and modal logic.* This is an extremely rich and fruitful domain: one considers formulas of the modal propositional calculus (i.e. for each formula A, we have a formula $\Box A$-necessarily A) and investigates their arithmetical interpretations. Such an interpretation $*$ assigns a sentence of arithmetic to each propositional variable, commutes with connectives and interprets $\Box$ as provability: $(\Box A)^*$ is $Pr^\bullet(\overline{A^*})$ (for a fixed provability predicate for a fixed arithmetic T). There is a natural axiomatic system L (or G) satisfying (for reasonable T and $Pr^\bullet$) arithmetical completeness: $G \vdash A$ iff for each arithmetical interpretation $*$, $T \vdash A^*$. A pioneering paper is [Solovay 76]; an extensive monograph is [Smoryński 85 – Self].

Further important names (see bibliography): Boolos, Sambin, de Jongh, Magari, Montagna, Artemov and others. One can introduce a further modality $\triangleright$ of interpretability and investigate modal interpretability logics and logics of partial conservativity (Švejdar, Visser, de Jongh, Veltman, Berarducci, Šavrukov, Hájek – Montagna).

(f) In this book we pay very little attention to advanced methods of *proof theory*; in particular, cut elimination is used only in Chap. V, Sect. 5. Proof theory is an extremely large domain and we shall not try to sketch its aims in a few words; instead, we refer to basic monographs: [Schütte 60], [Takeuti 71], [Pohlers 89]. See also [Schwichtenberg 77]. An application of proof *theory to fragments of* arithmetic is [Sieg 85]; we obtain some of the results presented by him using model theoretical rather than proof theoretical methods (in Chap. IV.)

Chapter IV

(1) The construction of a non-standard model of *PA* using definable ultrapower goes back to [Skolem 33] (Th. 1.7). For Loś's lemma 1.9 see [Loś 55]. Another early paper on models of arithmetic is [Müller 61].

Lemma 1.20 (which is the only place in this chapter explicitly using Matiyasevič's theorem) is due to [Gaifman 72] (for *PA*). Lemma 1.22 is due to Kirby and Theorem 1.24 elaborates a theorem of [Gaifman 72], cf. 1.25 (2). Theorem 1.29 (1) is due to [Paris-Kirby 77], (2)–(3) to Hájek, see [Hájek-Paris 86].

The method of proof of Theorem 1.33 goes back to [Paris-Kirby 78] (cf. also [Lessan 78]), but they did not deal with $P\Sigma_k$; results in 1.33 not concerning $P\Sigma_k$ are due to [Paris and Kirby 78], for the rest see [Hájek and Paris 86]. As we mentioned elsewhere, the fact that *PA* is not finitely axiomatizable (Corol. 1.34 (3)) was first proved by [Ryll-Nardzewski 52].

We are unable to credit anauthor for 1.51 and 1.52; but both theorems follow rather easily from the preceding.

Theorem 1.53 is due to [Paris-Kirby 78]. Theorem 1.59 is the famous Paris-Friedman conservation theorem; Friedman has never published his proof, Paris's proof is in [Paris 80] and in [Paris 81]. We present Kaye's proof (1.60, not published as a paper, see [Kaye 91 – Models]) and a modification of a Paris's proof (1.61) suitable for formalization in Sect. 4.

(2) We commented on the arithmetic with a top above (remark to Chap. II Sect. 2 (e)), Theorem 2.2 appears in [Cegielsky-McAloon-Wilmers 82] and is attributed to Paris. Subsections (b) and (c) present and elaborate part of [Paris 80, 81]. (See more on cuts and standard systems below.) Theorem 2.40 is due to [Ressayre 83, 86]; note that an analogous result for models of *PA* was first proved by [Friedman 73].

(3) Section 3 is central to Chap. IV: we present results of [Paris 80, 81]. In particular, Paris invented the notion of an indicator of a family of cuts. The term "envelope" (3.3) is borrowed from [McAloon 86]. Paris's main theorems are 3.5 and 3.7; the corollary 3.6 (characterizing $I\Sigma_1$ provably recursive functions as primitive recursive) is due independently to [Minc 76], [Takeuti 75] and Parsons (unpublished). A related paper is [Ratajczyk 88].

Section 4 presents results on model theory formalized in fragments; the result 4.8 (provability of the Paris-Friedman conservation theorem in $I\Sigma_1$) is due to Clote and

Hájek independently, cf. [Clote, Hájek, Paris 90]. The last paper contains a stronger result obtained by proof-theoretical methods.

*

Chapter IV presents only some important selected topics of model theory and its fragments. For an overview of the development of model theory of arithmetic the reader may consult [Smoryński 84 – LC82] and, with special emphasis on the work of Friedman, [Smoryński 85 – Friedman]. [Kaye 91 – Models] is a monograph devoted to models of arithmetic. We mentioned our main omission in the introduction to Chap. IV; now let us offer some references.

(a) *Standard systems.* See 2.13; for $I = N$, $SS_N(M)$ (or just $SS(M)$ is a set of subsets of N called the standard system of M (also: the set of reals of M). It is easy to show that if $M \models PA$ then $SS(M)$ is just the set of all $X \cap N$, where X is a parametrically definable subset of M. Similarly introduce $SS_0(M) = \{X \cap N \mid X$ nonparametrically definable in $M\}$. [Scott 62] was the first to investigate $SS_0(M)$ for $M \models PA$ ($SS_0(M)$ is the set of all subsets of N binumerable in the complete theory $Th(M)$) and characterized algebras of subsets of N obtainable as $SS_0(M)$ (often called Scott algebras). [Friedman 73] introduced $SS(M)$; it turns out that all $SS(M)$ coincide with all Scott algebras (for a particular M, $SS(M)$ and $SS_0(M)$ may or may not differ, cf. e.g. [Hájek 81 – alg]). Friedman obtained beautiful theorems on the embeddability of models in dependence on their standard systems. [Kaye 91 – Models] contains a detailed exposition.

(b) *Recursively saturated models.* A very fruitful domain. A ***type*** α in M consists of a set $\tau(x, y)$ of formulas $\varphi(x, y)$ and a parameter $b \in M$ interpreting y. The type is Δ_1 iff $\tau(x, y)$ is a Δ_1 set of formulas. The type α is ***finitely satisfiable*** in M if for each finite subset $\delta(x, y)$ of $\tau(x, y)$,

$$M \models (\exists a) \wedge \delta(a, b)\,.$$

The type α is ***satisfiable*** in M if there is an $a \in M$ such that $M \models \varphi(a, b)$ for each $\varphi \in \tau$. M is ***recursively saturated*** if each finitely satisfiable Δ_1 type is satisfiable. Each model M has a recursively saturated elementary extension. For an older overview of the theory of recursively saturated models of arithmetic see [Smoryński 81 – rec-sat]; see also [Kaye 91 – Models].

(c) *Cuts in models of arithmetic.* This is another very fruitful domain; we have presented only a fraction of known results. The pioneering paper is [Kirby-Paris 77] (and Kirby's dissertation [Kirby 77]). There are lots of papers on various sorts of cuts, indicators etc.; for an older survey see e.g. [Pillay 81 – cuts] further, see e.g. [McAloon 78 – (LC77)], [Kotlarski 81, 83, 84].

(d) Note also a very interesting topic – systems of PA with parameter-free induction, i.e. $I\varphi$ for φ having just one free variable. It is customary for a fragment T of arithmetic, to denote, by T^- the corresponding subtheory having the same axiom schema(s) as T but without parameters. It is easily seen that PA^- is equivalent to PA and $I\Sigma_0^-$ is equivalent to $I\Sigma_0$; but for $n \geq 1$, $I\Sigma_n^-$ is weaker than $I\Sigma_n$ and different from $I\Pi_n^-$. The basic paper is [Kaye, Paris, Dimitracopulos 88] (and Kaye's dissertation [Kaye 87]); see also [Adamowicz 88].

The above is only a selection of major topics not covered; one could continue mentioning the theory of satisfaction classes, uncountable models, indiscernibles and other developments.

Chapter V

Bounded Arithmetic was introduced by Parikh [Parikh 71]. He introduced the system which is nowadays denoted by $I\Sigma_0$ or $I\Delta_0$. Some weak fragments had been considered before. Shepherdson [Shepherdson 64, 65, 67] considered I_{open}. Goodstein [Goodstein 54] and Cleave and Rose [Cleave-Rose 67] studied so-called $\mathcal{E}^n$-arithmetics. These systems are equational systems corresponding to Grzegorczyk classes $\mathcal{E}^n$, [Grzegorczyk 53]. Another equational system was introduced by Cook [Cook 75]. His system PV is closely related to S_2^1 (this relationship is described in [Buss 86 – Bounded Arith.]). After Parikh, most of the research on $I\Sigma_0$ and $I\Sigma_0 + \Omega_1$ was done by Paris and Wilkie. The most important papers are [Paris-Wilkie 81 – Δ_0 Sets] and [Wilkie-Paris 87]. Let us note that [Wilkie-Paris 87] was published several years after the results of the paper were obtained. Later Buss entered this area with his book [Buss 86 – Bounded Arith.]. Several other mathematicians contributed to this field and the research is going on.

(1) The result on I_{open} is due to Shepherdson [Shepherdson 64]. For further independence results on I_{open} see [Macintyre-Marker 89], [Adamowicz 86]. Theorems 1.1 and 1.2 are due to McAloon [McAloon 82] (we present Kaye's proof of 1.1). Paris improved Theorem 1.2 to cover models of IE_1 [Paris 84]. Theorem 1.4 (only for $I\Sigma_0$) was proved in [Parikh 71]. The hierarchy of theories $I\Sigma_0 + \Omega_i$, $I\Sigma_0 + Exp$ and $I\Sigma_0 + Superexp$ appears in the papers of Paris and Wilkie. A system equivalent to $I\Sigma_0 + Exp$ is mentioned also in [Friedman 80]. The theory IE_1 and related systems were investigated in [Wilmers 85] and [Kaye 92 – Open]. $PIND$ and $LIND$ axioms were introduced in [Buss 86 – Bounded Arith.].

We mention only very briefly $\Sigma_0 PHP$, a very interesting subject with several nice results and open problems. Woods showed that $\Sigma_0 PHP$ proves Bertrand's Postulate, hence that there are infinitely many primes. (Note that by Parikh's theorem any proof of the infinitude of primes in Bounded Arithmetic must give a piece of information about the distribution of primes.) Wilkie proved a weaker version of $\Sigma_0 PHP$ in $I\Sigma_0 + \Omega_1$ which is sufficient for Woods' proof. The results are presented in [Paris-Wilkie-Woods 88]. A weak form of independence of PHP was shown by Ajtai [Ajtai 83]. He proved that if we extend $I\Sigma_0$ with a new relation symbol $R(x,y)$ (and do not add any special axioms about R except for the induction for the new formulae), then $PHP(R(x,y))$ is not provable in such a theory (which is denoted by $I\Sigma_0(R)$). He proved a similar result for the parity principle which says that an interval $[0, 2n+1)$ cannot be partitioned into two-element blocks [Ajtai 90].

(2) Computational complexity is a very broad subject. In Sect. 2 we have mentioned only a few results. There are several books about this subject; we recommend the following ones [Aho-Hopcroft-Ulman 74], [Garey-Johnson 79], [Savage 76], [Wagner-Wechsung 86], [Balcazár-Díaz-Gabarró 88, 90]. The very recent Handbook of Theoretical Computer Science [van Leeuwen 90] is also a very good source. We have quite neglected an important part of complexity theory which is the complexity of boolean function; the best reference is [Wegener 87].

Here we shall only state the authors of the theorems of Sect. 2. Theorems 2.2 (a) and (b) are due to Hartmanis and Stearns [Hartmanis-Stearns 65] and Hartmanis, Lewis and Stearns [Hartmanis-Lewis-Stearns 65], respectively. Theorem 2.4 is due to Hopcroft, Paul and Valiant [Hopcroft-Paul-Valiant 75]. Theorem 2.5 was proved by Savitch [Savitch 70]. Theorem 2.6 was proved independently by Immerman and Szelepczényi [Immerman 88] and [Szelepczényi 87]. Theorem 2.6 is due to Cook [Cook 71]; in this famous paper he formulated the $\mathsf{P} \overset{?}{=} \mathsf{NP}$ problem. Let us note in passing that already in 1956 Gödel had asked a question related to the $\mathsf{P} \overset{?}{=} \mathsf{NP}$ problem in a letter to von Neumann. He asked whether one can decide the provability of a formula by a proof of length n in time linear or quadratic in n. Now we know that this problem is NP-complete. The next influential paper was [Karp 72]; it contains also the proofs of the NP-completeness of CLIQUE and HAMILTONIAN GRAPHS. The NP-completeness of the solvable equations $ax^2 + by = c$ was proved by Manders and Adleman, see [Manders 80] for a survey. Theorem 2.8 is due to Ladner [Ladner 75]. Theorem 2.9 is due to Baker, Gill and Solovay [Baker-Gill-Solovay 75]. The Polynomial Time Hierarchy was introduced in [Stockmeyer 76]. The Linear Time Hierarchy was studied in [Wrathall 78]. The reference for Nepomnjaščij's Theorem 2.14 is [Nepomnjaščij 70]. Wrathall proved that *LinH* is equal to the class of ***rudimentary*** sets introduced in [Smullyan 61]. It is not difficult to show that rudimentary sets are just Δ_0 definable sets (using the natural coding of sequences). Thus Theorem 2.16 follows from her result. Theorems 2.17 (i) and (ii) were proved in [Hartmanis-Lewis-Stearns 65] and [Hartmanis-Stearns 65] respectively. Theorem 2.18 is due to Žák [Žák 83]. Theorem 2.19 was possibly never explicitly stated, but it is a typical application of Žák's method. Theorem 2.21 is due to Pudlák. Further results on time hierarchies can be found in [Paris-Wilkie 81].

An interesting area, which we do not cover, is *counting problems*. A typical question is the following: do Σ_0 definable sets have Σ_0 definable counting functions? ($F(x)$ is a counting function for A, if $F(x)$ is the number of elements of A smaller than x.) This is closely related to the questions about PHP in $I\Sigma_0$. If Σ_0 sets had Σ_0 counting functions and their properties were provable in $I\Sigma_0$, then $I\Sigma_0$ would prove $\Sigma_0 PHP$. The relation of counting to Bounded Arithmetic and approximations of counting functions were studies in [Paris-Wilkie 85 and 87]. Recently Toda proved a result from which it follows that the sets in *PH* do not have counting functions in *PH*, provided that *PH* does not collapse [Toda 89]. Hence, assuming that *PH* does not collapse, the answer to the above question is also negative.

(3) Bennett was the first to show a Σ_0 formula for the relation $z = x^y$ [Bennett 62]. Paris found another such formula and Dimitracopoulos [Dimitracopoulos 80] verified that the inductive clauses ((c.1) and (c.2)) are provable in $I\Sigma_0$ for Paris's formula. A different Σ_0 definition of exponentiation is in [Pudlák 83 – A definition]. In Sect. 3 we use the idea of [Nelson 86]: first to build a weak form of coding of sequences (based on binary expansions) and then to define exponentiation. However Nelson does not work in $I\Sigma_0$, he works in theories interpretable in Q. (Also he uses base four expansion of a single number instead of the binary expansions of two numbers.) A formalization of syntax in $I\Sigma_0$ is considered here for the first time, though the ideas on which it is based have been around for some time. A formalization of syntax in $I\Sigma_0 + \Omega_1$ was made in [Wilkie-Paris 87]. Theorem 3.37 is new. The corollary that context-free languages belong to Σ_0^N follows also from the result of [Wrathall 78] that context-free languages are rudimentary.

(4) The definition of the theories S_2, T_2 and their fragments S_2^i, T_2^i is due to Buss [Buss 86 – Bounded Arith.]. The theorems about the relationship between different axioms of induction and related principles are proved in that book or in [Buss 90 – Axiomatizations]; (Theorems 4.5, 4.7, 4.8, 4.10, 4.13). Our model-theoretical approach to witnessing theorems is inspired by Wilkie's proof of Buss's theorem [Wilkie 85]. Theorem 4.29 was proved by Krajíček and Takeuti, the proof here is due to Pudlák. Theorem 4.32 first appeared in [Buss 90 – Axiomatizations], it is a strengthening of Buss's theorem from [Buss 86, Bounded Arith.]. We have adapted Wilkie's proof to this strengthening. The result about models of fragments of S_2 which we have extracted from his proof (Theorem 4.31) might be of an independent interest. Theorem 4.38 is due to Krajíček, Pudlák and Takeuti [Krajíček, Pudlák and Takeuti 91]. See also [Krajíček 91 – Fragments].

A proof that S_2 is not finitely axiomatizable would strongly support the conjecture that the Polynomial Hierarchy is proper. However, proving the former statement seems to be also very hard. A possible way of proving it is proposed in [Krajíček, Pudlák 89 – Quantified]. Quantified propositional proof systems are related to the fragments S_2^i in the same way as the Extended Frege system is related to PV (hence to S_2^1) in [Cook 75] (this was done independently also by Dowd [Dowd 85]). To separate the fragments S_2^i, we need to prove a speed-up theorem about the length of proofs for the hierarchy of the quantified propositional calculi.

Versions of Buss's theorem for different fragments and different complexity function classes were proved in [Clote-Takeuti 86; 92].

The spectrum of the fragments of Bounded Arithmetic is much richer, more and more theories are being defined. A lot of research is being done on second order systems of Bounded Arithmetic. Such systems were introduced in [Buss 86 – Bounded Arith.]; for more recent results see [Takeuti 90 – S_3^i; 91 – A second], [Clote-Takeuti 92], [Krajíček 90 – Exponentiation].

(5) The truth definitions for Σ_0 were first considered in [Lesan 78] and [Paris-Dimitracopoulos 82]. The finite axiomatizability of $I\Sigma_0 + Exp$ (Theorem 5.6) was shown in [Gaifman-Dimitracopoulos 82]. Theorem 5.7 is due to Wilkie (unpublished). For related results on interpretability see [Szemielew-Tarski 52], [Nelson 86], (Nelson considers only local interpretations), and references to Chap. III. Theorem 5.12 is in [Wilkie-Paris 87]. We formalize some classical results of proof theory, the best reference is [Takeuti 80]. The important concept of restricted provability and restricted consistency is due to Paris and Wilkie [Wilkie-Paris 87]. Theorem 5.26 is due to Wilkie [Wilkie 86 – On sentences]. His original proof of this theorem is different, he uses a model-theoretical construction. Theorem 5.27, in a slightly stronger form, is in [Wilkie-Paris 87]. Theorem 5.28 is due to Pudlák [Pudlák 85]. Corollary 5.29 and Theorem 5.31 were proved in [Wilkie-Paris 87] (we give different proofs). Corollary 5.32 is from [Wilkie 86 – On sentences]. Theorem 5.33 and Corollary 5.34 are strengthenings of results of [Wilkie-Paris 87]. The last theorem is new.

There are more results about incompleteness and truth definitions in Bounded Arithmetic. Their main motivation is the problem of finite axiomatizability of Bounded Arithmetic. We cannot use ordinary consistency statements to separate fragments of Bounded Arithmetic, since even $I\Sigma_0 + Exp$ does not prove $Con_Q^\bullet$. In [Buss 86 – Bounded Arith.] Buss proposed to use various kinds of *bounded consistency*, which are consistencies with respect to proofs containing only bounded formulae. It has turned out that a similar incompleteness extends also to such consistency statements. For instance S_2 does not prove bounded consistency of S_2^1 [Pudlák 90 – A note], [Takeuti 88 – Some relations]. For further results see [Takeuti 88 – Bounded], [Krajíček-Takeuti to appear].