NOVA SCHOOL OF
SCIENCE & TECHNOLOGY

DEPARTMENT OF MATHEMATICS

ON NOTIONS OF PROVABILITY

PAULO GUILHERME DOMINGOS CANHA MOREIRA DOS SANTOS
Master in Mathematics and Applications

JOINT DOCTORATE IN
MATHEMATICS AT NOVA University Lisbon
COMPUTER SCIENCE AT Universität Tübingen

January, 2023

DEPARTMENT OF MATHEMATICS

ON NOTIONS OF PROVABILITY

PAULO GUILHERME DOMINGOS CANHA MOREIRA DOS SANTOS

Master in Mathematics and Applications

Adviser: Isabel Maria Oitavem Fonseca da Rocha Kahle
Associate Professor, NOVA University Lisbon

Co-adviser: Reinhard Josef Klaus Kahle
Full Professor, Universität Tübingen

Examination Committee

Chair: João Jorge Ribeiro Soares Gonçalves de Araújo
Full Professor, NOVA University Lisbon

Rapporteurs: Matthias Franz Baaz
Full Professor, Technische Universität Wien

Gilda Maria Saraiva Dias Ferreira
Assistant Professor, Universidade Aberta

Adviser: Isabel Maria Oitavem Fonseca da Rocha Kahle
Associate Professor, NOVA University Lisbon

Members: Klaus Ostermann
Full Professor, Universität Tübingen

João Jorge Ribeiro Soares Gonçalves de Araújo
Full Professor, NOVA University Lisbon

JOINT DOCTORATE IN
MATHEMATICS AT NOVA University Lisbon
COMPUTER SCIENCE AT Universität Tübingen

January, 2023

On Notions of Provability

This document was created with the (pdf/Xe/Lua)LaTeX processor and the NOVAthesis template (v6.9.5) [66].

To Pi: the start of my mind, the galaxy of my life.
To my Parents: the anchors of my ship in any stormy day.

Acknowledgements

My journey in logic started by reading Bertrand Russell, one of my intellectual idols whose views (in logic, science, society, and love) I subscribe to an incredible great extent—my dear Pi has witnessed this journey and listened to my unending monologues on this topic (what a saint you are). I was lucky to start a deeper study of logic under the supervision of Professor Reinhard Kahle in my Bachelor, under a scholarship by *Fundação Calouste Gulbenkian*: all the meetings we had were determinant for me to have a sound and comprehensive understanding of the basic concepts of logic. I wholeheartedly thank Professor Kahle for all his help, for his support by investing in my scientific life, and for my scientific freedom. Professor Isabel Oitavem gave me a similar comprehensive background in recursion theory at that time; I am graciously thankful to Professor Oitavem for all the help, for improving my way of writing science, for guiding my investigations, for the support by investing in my scientific life, and for my scientific freedom.

I would like to profoundly thank Professor Wilfried Sieg for the truly wonderful time I had in Pittsburgh: Professor Sieg helped me to improve as a scientist, especially to express my ideas in a clear way.

I am also very thankful to Professor Albert Visser for reading some initial versions of parts of the work I present in the current thesis, and by suggesting lines of investigation.

The work I developed was funded by the *Fundação para a Ciência e a Tecnologia* scholarship SFRH/BD/143756/2019 and my stay in the USA was supported by the *Bolsa Fulbright para Investigação com o apoio da Fundação para a Ciência e a Tecnologia* (*FCT*).

I would like to thank my friends and family for the enormous support and love they gave me.

I am thoroughly thankful to my parents for their love, and profound support; they have been an essential part of the help I have had during my journey as a person and as a scientist.

My sweet Pi, I am dearly thankful for all your help; all my dreams are yours and all your dreams are mine; if I have succeeded it is only because I have you by my side.

“The meaning of a proposition is the method of its verification.” (Moritz Schlick)

Abstract

In this thesis, we study *notions of provability*, i.e. formulas $B(x,y)$ such that a formula φ is provable in T if, and only if, there is $m \in \mathbb{N}$ such that $T \vdash B(\ulcorner\varphi\urcorner, \overline{m})$ (m plays the role of a parameter); the usual notion of provability, k-step provability (also known as k-provability), s-symbols provability are examples of notions of provability.

We develop general results concerning notions of provability, but we also study in detail concrete notions. We present partial results concerning the decidability of k-provability for Peano Arithmetic (PA), and we study important problems concerning k-provability, such as *Kreisel's Conjecture* and *Montagna's Problem*:

$$(\forall n \in \mathbb{N}. T \vdash_{k \text{ steps}} \varphi(\overline{n})) \implies T \vdash \forall x. \varphi(x), \qquad \textbf{[Kreisel's Conjecture]}$$

and

$$\text{Does } \mathsf{PA} \vdash_{k \text{ steps}} \mathsf{Pr}_{\mathsf{PA}}(\ulcorner\varphi\urcorner) \to \varphi \text{ imply } \mathsf{PA} \vdash_{k \text{ steps}} \varphi? \qquad \textbf{[Montagna's Problem]}$$

Incompleteness, Undefinability of Truth, and Recursion are different entities that share important features; we study this in detail and we trace these entities to common results.

We present numeral forms of completeness and consistency, *numeral completeness* and *numeral consistency*, respectively; numeral completeness guarantees that, whenever a $\Sigma_1^b(\mathsf{S}_2^1)$-formula $\varphi(\vec{x})$ is such that $\vec{Q}\,\vec{x}.\varphi(\vec{x})$ is true (where $\vec{Q}$ is any array of quantifiers), then this very fact can be proved inside S_2^1, more precisely $\mathsf{S}_2^1 \vdash \vec{Q}\,\vec{x}.\mathsf{Pr}_\tau(\ulcorner\varphi(\dot{\vec{x}})\urcorner)$. We examine these two results from a mathematical point of view by presenting the minimal conditions to state them and by finding consequences of them, and from a philosophical point of view by relating them to Hilbert's Program.

The derivability condition "*provability implies provable provability*" is one of the main derivability conditions used to derive the Second Incompleteness Theorem and is known to be very sensitive to the underlying theory one has at hand. We create a weak theory G_2 to study this condition; this is a theory for the complexity class FLINSPACE. We also relate properties of G_2 to equality between computational classes.

Keywords: Notions of Provability, k-provability, Kreisel's Conjecture, Incompleteness, Theory of Arithmetic, Numeral Completeness, Hilbert's Program, Provability Implies Provable Provability, Linear Space

Resumo

O tema desta tese são *noções de demonstração*; estas últimas são fórmulas $B(x,y)$ tais que uma fórmula φ é demonstrável em T se, e só se, existe $m \in \mathbb{N}$ tal que $T \vdash B(\ulcorner\varphi\urcorner, \overline{m})$ (m desempenha o papel de um parâmetro). A noção usual de demonstração, demonstração em k-linhas (demonstração-k), demonstração em s-símbolos são exemplos de noções de demonstração.

Desenvolvemos resultados gerais sobre noções de demonstração, mas também estudamos exemplos concretos. Damos a conhecer resultados parciais sobre a decidibilidade da demonstração-k para a Aritmética de Peano (PA), e estudamos dois problemas conhecidos desta área, a *Conjectura de Kreisel* e o *Problema de Montagna*:

$$(\forall n \in \mathbb{N}. T \vdash_{k \text{ steps}} \varphi(\overline{n})) \implies T \vdash \forall x.\varphi(x), \qquad \textbf{[Conjectura de Kreisel]}$$

e

$$\mathsf{PA} \vdash_{k \text{ steps}} \mathsf{Pr}_{\mathsf{PA}}(\ulcorner\varphi\urcorner) \to \varphi \text{ implica } \mathsf{PA} \vdash_{k \text{ steps}} \varphi? \qquad \textbf{[Problema de Montagna]}$$

A Incompletude, a Incapacidade de Definir Verdade, e Recursão são entidades que têm em comum características relevantes; nós estudamos estas entidades em detalhe e apresentamos resultados que são simultaneamente responsáveis pelas mesmas.

Além disso, apresentamos formas numerais de completude e consistência, a *completude numeral* e a *consistência numeral*, respectivamente; a completude numeral assegura que, quando uma fórmula-$\Sigma_1^b(\mathsf{S}_2^1)$ $\varphi(\vec{x})$ é tal que $\vec{Q}\,\vec{x}.\varphi(\vec{x})$ é verdadeira, então este facto pode ser verificado dentro de S_2^1, mais precisamente $\mathsf{S}_2^1 \vdash \vec{Q}\,\vec{x}.\mathsf{Pr}_\tau(\ulcorner\varphi(\dot{\vec{x}})\urcorner)$. Este dois resultados são analisados de um ponto de vista matemático onde apresentamos as condições mínimas para os demonstrar e apresentamos consequências dos mesmos, e de um ponto de vista filosófico, onde relacionamos os mesmos com o Programa de Hilbert.

A condição de derivabilidade "*demonstração implica demonstrabilidade da demonstração*" é uma das condições usadas para derivar o Segundo Teorema da Incompletude e sabemos ser muito sensível à teoria de base escolhida. Nós criámos uma teoria fraca G_2 para estudar esta condição; esta é uma teoria para a classe de complexidade $\mathsf{FLINSPACE}$. Também relacionámos propriedades de G_2 com igualdades entre classes de complexidade computacional.

Palavras-chave: Noções de Demonstração, demonstração-k, Conjectura de Kreisel, Incompletude, Teoria de Aritmética, Completude Numeral, Programa de Hilbert, Demonstração Implica Demonstrabilidade da Demonstração, Espaço Linear

Contents

1 Introduction **1**

2 Incompleteness, Truth, Recursion **5**

2.1 Incompleteness and Rice's Theorem 9
2.2 Incompleteness via Recursion and the Non-Recursiveness of Truth . . . 14
2.3 General use of the Diagonalization Lemma 17

3 k-provability in PA **23**

3.1 Introduction 23
3.2 The theory PA′ 25
3.3 Main results 28
3.3.1 Provable schemata 28
3.3.2 Decidability of schemata 40
3.3.3 Decidability of some proof-skeletons and k-provability 46

4 Montagna's Problem and Kreisel's Conjecture **65**

4.1 Preliminaries 65
4.2 Introduction 65
4.3 General Facts on Bounded Notions of Provability 67
4.4 On the notion $\vdash_{\leq h}$ 69
4.5 Montagna's conjecture 71
4.6 Variants of Kreisel's Conjecture 75
4.7 On $\vdash_{k \text{ steps}}$ and $\vdash_{\leq h}$ 79
4.8 Conclusion 81

5 Numeral Completeness **83**

5.1 Introduction 83
5.2 Preliminaries 84
5.2.1 Notation and definitions 84
5.2.2 Derivability conditions 86
5.3 Numeral Completeness and Consistency 87
5.3.1 Numeral Completeness 87
5.3.2 Numeral Consistency 90

5.4 Gödel's Second Incompleteness Theorem and "Provability Implies Provable Provability" . . . 91
5.4.1 Numeral Completeness . . . 91
5.4.2 Gödel's Second Incompleteness Theorem for Weak Theories . . . 92
5.4.3 'Provability Implies Provable Provability' for Weak Theories . . 95
5.5 Characterization of the Provability Predicates that Satisfy Numeral Completeness and Consistency . . . 96
5.6 Finitist Consistency . . . 98
5.7 A general Negative Bound on Finitist Consistency Proofs . . . 101
5.8 Provable consistency . . . 104
5.9 Hilbert's Programs and Finitist Consistency . . . 108

6 Provability Implies Provable Provability 113
6.1 Introduction . . . 113
6.2 The Grzegorczyk class $\mathscr{E}^2$ . . . 114
6.3 A theory for FLINSPACE . . . 116
6.3.1 The theory G_2 . . . 117
6.3.2 A relation between $\mathscr{E}^2$ and $\mathscr{M}^2$ using G_2 . . . 120
6.4 Bootstrapping G_2 . . . 123
6.5 'Provability implies provable provability' in G_2 . . . 126
6.6 A form of internal completeness in G_2 . . . 134

7 Conclusions and Future Work 139

Bibliography 141

Chapter 1

Introduction

Proofs are, in their essence, justifications for (mathematical) statements. To have a proof of a given proposition is something much deeper than just knowing the fact that the proposition is true; it gives insights on why that is the case, it sometimes gives an algorithm to use the proposition, and more importantly, it gives a concrete meaning to the fact that the considered proposition is true.

Every field of enquiry has particular ways of acting and particular ways of establishing facts; but there is one aspect that is transversal to all of them: the need to give *justifications*. The scientist does not dogmatically decide how nature looks, on the contrary she collects data and creates a scientific justification for the propositions of science. Proofs are for the mathematician what justifications and verifiability are to the scientist; they constitute the scientific nature of mathematics.

The urge to study mathematical proofs gained special traction in the 20^{th} century. Since then, proofs became mathematical objects by their own right. That century became famous for the study of the limits of several areas: the limits of language, the limits of what can be precisely stated, and, of course, the limits of the use and applicability of proofs. We follow that approach here.

Proofs are the main focus of the present thesis. Here, we try to quantitatively measure them, to study their expressive power, and to find possible limitations to their use. We achieve this goal via the study of *notions of provability*; these are formulas $B(x,y)$ (for a certain theory T) with the following property:

$$T \vdash \varphi \iff \exists n \in \mathbb{N}. T \vdash B(\ulcorner \varphi \urcorner, \overline{n}). \qquad \textbf{[NProv]}$$

The role played by the n in the previous condition is the role of a parameter that measures some aspect of certain proof of φ: it can be the number of steps, the number of symbols, *et cetera*. Clearly, the usual arithmetized notion of provability, usually denoted by Pr_ξ

(for a certain numeration ξ of the axioms of T), is a notion of provability, where the parameter is trivially the code of a certain proof of the considered formula. Moreover, the arithmetized notions of k-steps provability, $T \vdash_{k \text{ steps}} \cdot$, and s-symbols provability, $T \vdash_{s \text{ symbols}} \cdot$, are notions of provability. Without danger of confusion, we identify the arithmetized notions of provability with the non-arithmetized versions; so we consider $T \vdash_{k \text{ steps}} \cdot$ and $T \vdash_{s \text{ symbols}} \cdot$ as being notions of provability without the need to arithmetize them. The defining feature of a (general) notion of provability $\mathscr{B}(\varphi, p)$ is that a formula φ is provable in T exactly when there is a parameter p such that $\mathscr{B}(\varphi, p)$ holds: this p can have very distinct natures; it can even be a function.

This general perspective of the study of notions of provability allows one to obtain a clearer landscape on how proofs work. In our thesis, we present general results concerning notions of provability—see, for example, Theorems 4.3.1 and 5.7.1—, but we also study particular notions in the spirit of quantitatively measure proofs, studying their expressive power, and finding possible limitations to their use.

k-steps provability, better known as k-provability, is one of the main focus of our work. In [73], [31], and [58] the decidability of this relation was studied for several formalizations of Peano arithmetic (PA); in Chapter 3, we study its decidability for particular values of k for the usual axiomatization of PA. Our approach is very general, since it is parametrized by unification algorithms.

Kreisel's Conjecture is an open problem about k-provability [18]:

$$(\forall n \in \mathbb{N}. T \vdash_{k \text{ steps}} \varphi(\overline{n})) \implies T \vdash \forall x. \varphi(x). \qquad \textbf{[Kreisel's Conjecture]}$$

It was studied in [73], [58], [32], [68], [74], [15], and [1]. We study this Conjecture in Chapter 4 for a new notion of provability $T \vdash_{\leq h} \cdot$ that depends on recursive functions h ($T \vdash_{\leq h} \cdot$ is a notion of provability because $T \vdash \varphi \iff \exists h. T \vdash_{\leq h} \varphi$: here the parameter is a recursive function). Moreover, we study a problem proposed by Montagna [18, p. 9]:

$$\text{Does } \mathsf{PA} \vdash_{k \text{ steps}} \mathsf{Pr}_{\mathsf{PA}}(\ulcorner \varphi \urcorner) \to \varphi \text{ imply } \mathsf{PA} \vdash_{k \text{ steps}} \varphi? \qquad \textbf{[Montagna's Problem]}$$

We give a negative answer to this problem for some axiomatizations of PA.

In Chapter 2, we dive into the limitations of what is provable: the Incompleteness phenomena. There, we relate Incompleteness to: Rice's Theorem, Kleene's Normal Form, the Undefinability of Truth, and Hilbert-Bernays Paradox. There have been several works that explored these relations, see for instance [36] or [93]—usually, such relations are either too general, in the sense that one abstracts so much that one looses track of the specificities of the original objects; or too specific, in the sense that one focuses so much in the particularities that one cannot see the "big picture". We show different relations and we try to maintain the right balance of abstractness *versus* concreteness.

The notion of provability Pr_ξ, the usual notion of provability, is studied in Chapters 5 and 6. In the former, we present a provable form of completeness, *Numeral Completeness*. This form of completeness guarantees that, whenever a $\Sigma_1^b(\mathsf{S}_2^1)$-formula $\varphi(\vec{x})$ is such that

$\vec{Q}\vec{x}.\varphi(\vec{x})$ is true (where $\vec{Q}$ is any array of quantifiers), then this very fact can be proved inside S^1_2, more precisely $\mathsf{S}^1_2 \vdash \vec{Q}\vec{x}.\mathsf{Pr}_\tau(\ulcorner \varphi(\dot{\vec{x}})\urcorner)$. This result entails a form of provable consistency, *Numeral Consistency*: if *Prf* is a $\Delta^b_1(\mathsf{S}^1_2)$-proof predicate for a consistent theory T, then, there is an $\exists\Delta^b_1(\mathsf{S}^1_2)$-numeration τ such that $\mathsf{S}^1_2 \vdash \forall x.\mathsf{Pr}_\tau(\ulcorner \neg Prf(\ulcorner\bot\urcorner, \dot{x})\urcorner)$.

We characterize the provability predicates for which Numeral Completeness and Numeral Consistency hold, and we study the Second Incompleteness Theorem. In addition, we present a general negative bound on finitist proofs of consistency for notions of provability. In the second part of this Chapter, the philosophical part, we relate our results to some versions of Hilbert's Program.

Chapter 6 is devoted to the study of the derivability condition 'provability implies provable provability', i.e. $\mathsf{Pr}_\xi(x) \to \mathsf{Pr}_\xi(\ulcorner \mathsf{Pr}_\xi(\dot{x})\urcorner)$. This condition is very sensitive to the underlying theory, for example it is an open problem if it holds for $\mathsf{I}\Delta_0$. We create a weak theory G_2 to study this condition; this is a theory for the complexity class FLINSPACE. This is an interesting approach since we study the desired condition for a weak theory and since it is uncommon to study metamathematics in theories that are obtained from computational classes limited on space, and not on time. We also relate properties of G_2 to equality between computational classes.

Each chapter of our thesis has an autonomous nature and notation: the reader can independently read each one of them, the interdependence between any given two chapters is minimal (as a consequence of this, some definitions are very slightly different from chapter to chapter, since we want to focus different aspects in each chapter).

Chapter 2

Incompleteness, Undefinability of Truth, Recursion

Introduction

The Incompleteness Theorems and the Undefinability of Truth Theorem are among the most relevant results in logic; from model theory to proof theory, they changed the way this field of knowledge is conceived. Similarly, the Second Recursion Theorem is one of the pinnacles of recursion theory, as well as Rice's Theorem. Despite their major significance, the proofs of these results are deceptively simple when one is given the suitable formal background; furthermore, when one compares them, one gets the idea that there is a lot in common between them. Moreover, even the results have a similar nature, e.g. the First Incompleteness Theorem (G1) and Rice's Theorem claim very similar limitations, the former on provability, the latter on recursive properties of functions. There have been several works that explored some of these relations, see for instance [36] or [93]—usually, such relations are either too general, in the sense that one abstracts so much that one looses track of the specificities of the original objects (theories of arithmetic); or too specific, in the sense that one focuses so much in the particularities that one cannot see the "big picture", *id est* one does not, in fact, relate them. In this chapter, we show different relations and we try to maintain the right balance of abstractness *versus* concreteness, that is to say, we present general results relating the mentioned Theorems always inside theories of arithmetic.

There are several papers in the direction we aim at, but with different considerations. For instance, [54] studies the interplay between Kolmogorov complexity and the Second Incompleteness Theorem (G2); [67] studies incompleteness and jump hierarchies; Visser in [100] derives G2 from the Undefinability of Truth; and [4] presents a general setting to study incompleteness.

The current chapter is divided in three sections. In the first section, we explore how G1 is related to Rice's Theorem by developing a version of Kleene's Normal Form with formulas and provability, and by presenting an encompassing result of these facts.

The second section is devoted to the interplay between G1 and the non-recursiveness of truth via recursion theory—we exhibit a general notion of incompleteness for recursively enumerable sets.

In the third and last section, we present a general arithmetical form of the Diagonalization Lemma—one of the responsible for the feeling initially described about the shared characteristics of the proofs of the incompleteness results and the Undefinability of Truth—that encompasses: G1, the Undefinability of Truth, and Hilbert-Bernays Paradox.

Preliminaries

Throughout this chapter, T denotes a (first-order) r.e. theory of arithmetic that is a consistent extension of $\mathsf{EA} := \mathsf{I}\Delta_0 + \mathtt{exp}$, where $\mathtt{exp}$ denotes the totality of the exponentiation function (we implicitly assume the soundness of T; see [39, pp. 294–315] for details on EA). Unless otherwise stated, we assume that T has the same language as Q (see [90, pp. 55, 56] for details on this theory). Given a class of formulas Γ, we say that a formula φ is a $\Gamma(T)$-formula if there is a Γ-formula φ_0 T-equivalent to T. The classes Σ_n and all standard notation are taken from [39, pp. 13–18, 62]. Furthermore, we use $\#\varphi$ to denote the Gödel-number of φ, and $\ulcorner\varphi\urcorner$ to denote the numeral of the Gödel-number of φ, i.e., $\overline{\#\varphi}$ (here we assume the efficient numerals [39, p. 304]).

As usual, we denote the *standard proof predicate for T* [90, p. 170] by $\mathtt{Proof}_T(x,y)$—with the meaning that "y is the code of a T-proof of the formula coded by x" (this corresponds to fixing a particular $\Delta_0(T)$-enumeration of the axioms of T, see [33] and [39]). We define $\mathtt{Pr}_T(x) := \exists y.\mathtt{Proof}_T(x,y)$, the *standard provability predicate for T*. We assume that the reader is aware of the derivability conditions for $\mathtt{Pr}_T$, see [39, p. 163] for further details.

Definition 2.0.1. We say that a formula $Prf(x,y)$ is a *proof predicate for T* if it satisfies the following conditions:

Prf 1: $Prf(x,y)$ is $\Delta_1(T)$;

Prf 2: For all $n \in \mathbb{N}$ and formulas φ, $\mathbb{N} \models \mathtt{Proof}(\ulcorner\varphi\urcorner,\overline{n}) \leftrightarrow Prf(\ulcorner\varphi\urcorner,\overline{n})$;

Prf 3: $T \vdash \forall x.\forall x'.\forall y.(Prf(x,y) \wedge Prf(x',y) \rightarrow x = x')$.

A *provability predicate for T* is a predicate $P(x)$ that satisfies

$$T \vdash \forall x.(P(x) \leftrightarrow \exists y.Prf(x,y)),$$

for a certain proof predicate $Prf(x,y)$ for T. For a fixed provability predicate P, we also define $\mathtt{Con}_P := \neg P(\ulcorner\bot\urcorner)$ and $\mathtt{Con}^1_P := \neg P(\ulcorner P(\ulcorner\bot\urcorner)\urcorner)$.

We assume $\dot{\neg}$ such that $T \vdash \dot{\neg}\ulcorner\varphi\urcorner = \ulcorner\neg\varphi\urcorner$ and $\dot{\rightarrow}$ such that $T \vdash \ulcorner\varphi\urcorner \dot{\rightarrow} \ulcorner\psi\urcorner = \ulcorner\varphi \rightarrow \psi\urcorner$, for all formulas φ and ψ (formally, these can be either actual function-symbols of T or represented by $\Delta_1(T)$-formulas inside T); we assume similar notations for the other connectives. We assume $\mathsf{S}(x,y)$ such that for all formulas $\varphi(x)$ and terms t, $T \vdash \mathsf{sub}(\ulcorner\varphi(x)\urcorner, \ulcorner t\urcorner) = \ulcorner\varphi(t)\urcorner$ and $\mathtt{numeral}$ such that $T \vdash \mathtt{numeral}(\overline{n}) = \ulcorner\overline{n}\urcorner$, we define Feferman's dot notation by $\ulcorner\varphi(\dot{x})\urcorner := \mathsf{sub}(\ulcorner\varphi(y)\urcorner, \mathtt{numeral}(x))$.

We remember that a theory T is *complete* if for every sentence φ, either $T \vdash \varphi$ or $T \vdash \neg\varphi$; it is said to be *incomplete* if it is not complete. Considering T, the completeness is equivalent to[1]: for all sentences φ, $\mathbb{N} \models \varphi \implies T \vdash \varphi$. We say that T is *ω-consistent* if there is *no* formula $\varphi(x)$ such that $T \vdash \exists x.\varphi(x)$ and, for all $n \in \mathbb{N}$, $T \vdash \neg\varphi(\overline{n})$. Now we present a version of G1 (see the original in [37]).

Theorem 2.0.1 (G1)**.** *If T is ω-consistent, then T is incomplete.*

For more details on the previous result and on technical details such as ω-consistency we recommend [91]. Using Rosser's argument, we can substitute ω-consistent for the usual consistency [39, p. 161].

The Second Incompleteness Theorem (G2) is:

Theorem 2.0.2 (G2)**.** *If T is consistent, then $T \nvdash \mathsf{Con}_{\mathsf{Pr}_T}$.*

The Diagonalization Lemma was extracted by Carnap from Gödel's initial proof of G1:

Theorem 2.0.3 (Diagonalization Lemma)**.** [78, p. 250] *For every formula $\varphi(x)$ there is a sentence ψ such that $T \vdash \psi \leftrightarrow \varphi(\ulcorner\psi\urcorner)$.*

There are several generalizations of the previous result, for example:

Theorem 2.0.4 (Two-Variable Diagonalization Lemma)**.** [69, p. 208] *For every $\alpha(x,y)$ there is a formula $\varphi(x)$ such that $T \vdash \forall z.\varphi(z) \leftrightarrow \alpha(\ulcorner\varphi(x)\urcorner, z)$.*

For the rest of this section, we assume that T has function-symbols for each primitive recursive function. There is an analogue result for terms that was originally presented in [51]:

Theorem 2.0.5 (Strong-Diagonalization Lemma)**.** [90, p. 55] *For every formula $\varphi(x)$, there is a closed term t such that $T \vdash \ulcorner\varphi(t)\urcorner = t$.*

[1] Here is a proof of our claim:

Proof. Firstly, suppose that T is complete. Let φ be any sentence such that $\mathbb{N} \models \varphi$. As by assumption T is complete, then either $T \vdash \varphi$ or $T \vdash \neg\varphi$. If $T \vdash \neg\varphi$, then we would get, by the soundness of T, $\mathbb{N} \models \neg\varphi$, a contradiction; so $T \vdash \varphi$. Thus, $\mathbb{N} \models \varphi \implies T \vdash \varphi$.

Conversely, assume that, for all sentences φ, $\mathbb{N} \models \varphi \implies T \vdash \varphi$. Let φ be any sentence. We know that either $\mathbb{N} \models \varphi$ or $\mathbb{N} \models \neg\varphi$; on the first case, using the implication, we get $T \vdash \varphi$, and on the second one we get $T \vdash \neg\varphi$. In all, $T \vdash \varphi$ or $T \vdash \neg\varphi$. ⊣

The following result is similar to the Strong-Diagonalization Lemma.

Theorem 2.0.6 (Term-Diagonalization Lemma)**.** [75] *For every term $h(x)$ there is a closed term t such that $T \vdash h(\ulcorner t \urcorner) = t$.*

It is not hard to see that one can prove the Diagonalization Lemma using the Strong-Diagonalization Lemma, but the converse does not hold in the sense of the following Theorem that claims that we cannot use the instances of the Diagonalization Lemma to prove the instances of the Strong-Diagonalization Lemma.

Theorem 2.0.7. [83, §2] *There is* no *formula $\alpha(x,y)$ such that: given a formula $\varphi(x)$, if ψ is the sentence obtained from the Diagonalization Lemma applied to $\alpha(\ulcorner \varphi(x) \urcorner, x)$, then there is a term t such that ψ is $\varphi(t)$ and $T \vdash \ulcorner \varphi(t) \urcorner = t$.*

Despite this intrinsic distinction between (all) the Diagonalization Lemmas, in the end of this chapter we will present a result that generalizes their use, it unifies diagonalization of formulas and terms. Finally, we present two important facts that are very similar in their layout (we will latter confirm that they share a common arithmetical structure).

Theorem 2.0.8 (Tarski's Undefinability of Truth)**.** [39, p. 159] *There is* no *formula $Tr(x)$ such that, for all formulas φ, $T \vdash Tr(\ulcorner \varphi \urcorner) \leftrightarrow \varphi$.*

The next result was original mentioned in the famous *Grundlagen der Mathematik*, by Hilbert and Bernays; the reader can find further details in [75]. This result is relevant because, just like the Undefinability of Truth, it shows the non-existence of a total form of representability; the latter states the non-representability of truth, and this result affirms the non-representability of values when T has function-symbols for each primitive recursive function.

Theorem 2.0.9 (Hilbert-Bernays Paradox)**.** *There is* no *term $h(x)$ such that for all closed terms t, $T \vdash h(\ulcorner t \urcorner) = t$.*

We present a proof of the previous result, since it is not a very well-known fact.

Proof. Let $\mathtt{diag}$ denote the primitive recursive function such that, for each term $t(x)$, $T \vdash \mathtt{diag}(\ulcorner t(x) \urcorner) = \ulcorner t(\ulcorner t(x) \urcorner) \urcorner$. Consider $r(x) := h(\mathtt{diag}(x)) + \overline{1}$. Then, $T \vdash \mathtt{diag}(\ulcorner r \urcorner) = \ulcorner r(\ulcorner r \urcorner) \urcorner = \ulcorner h(\mathtt{diag}(\ulcorner r \urcorner)) + \overline{1} \urcorner$. Set $s := \mathtt{diag}(\ulcorner r \urcorner)$. As $T \vdash h(\ulcorner r(\ulcorner r \urcorner) \urcorner) = r(\ulcorner r \urcorner)$, we get

$$T \vdash h(s) = h(\mathtt{diag}(\ulcorner r \urcorner)) = h(\ulcorner r(\ulcorner r \urcorner) \urcorner) = r(\ulcorner r \urcorner) = h(\mathtt{diag}(\ulcorner r \urcorner)) + \overline{1} = h(s) + \overline{1},$$

a contradiction. ⊣

It is important to observe that the previous result does not contradict Lemma 1.66 of [39, p. 55] that guarantees the existence of a function evaluating terms in the language $\{0, \mathsf{S}, +, \times\}$; here we have a terms for each primitive recursive function (the essential feature is to have a term $\mathtt{diag}$), something that does not occur in that context. Another way to read the previous result is that we cannot extend the construction of Lemma 1.66 from [39, p. 55] to the general context of having a term for each primitive recursive function.

2.1 Incompleteness and Rice's Theorem

In this section, we trace the Second Recursion Theorem to the Diagonalization Lemma. Moreover, we prove a generalization of Rice's Theorem that is also responsible for G1.

Definition 2.1.1. A partial function $f : \mathbb{N}^k \rightharpoonup \mathbb{N}$ is *strongly representable in T* (*as a partial function*) if there is a formula $\varphi(x_0,\dots,x_{k-1},y)$ (not necessarily Σ_1, this notion is more general than the usual one) such that

R1: For all $m_0,\dots,m_{k-1},n \in \mathbb{N}$, $f(m_0,\dots,m_{k-1}) \simeq n \iff T \vdash \varphi(\overline{m_0},\dots,\overline{m_{k-1}},\overline{n})$;

R2: $T \vdash \forall x_0 \cdots \forall x_{k-1}.\exists! y.\varphi(x_0,\dots,x_{k-1},y)$.

The previous notion is clearly distinct from the usual one using Σ_1-formulas. In [80], the reader can find a very simple proof of the following fact.

Theorem 2.1.1. *All partial recursive functions are strongly representable in any r.e. extension of* Q *as partial functions.*

This means that strongly representability, although allowing formulas that are not necessarily Σ_1, is a notion that can be captured in the very weak Q. Kleene's Normal Form gives a simple way to characterise all (partial) recursive functions [94, p. 15]. The next result presents an alternative way to characterize them using provability instead of Kleene's Predicate.

Theorem 2.1.2 (Kleene's Normal Form with Provability). *The (partial) recursive functions are exactly the functions f that can be defined by*[2]

$$f(\vec{m}) := \text{the first } n \text{ that satisfies: } \mathsf{Q} \vdash \varphi(\overline{\vec{m}},\overline{n}) \text{ and}$$
$$\text{for all } k < n \text{ it holds } \mathsf{Q} \vdash \neg\varphi(\overline{\vec{m}},\overline{k}),$$

for a certain formula $\varphi(\vec{x})$.

Proof. Suppose that $g(m_0,\dots,m_{k-1})$ is a (partial) recursive function. Then, by Theorem 2.1.1, there is a formula $\varphi(x_0,\dots,x_k)$ such that, for all $m_0,\dots,m_{k-1},n \in \mathbb{N}$, one has $g(m_0,\dots,m_{k-1}) \simeq n \iff \mathsf{Q} \vdash \varphi(\overline{m_0},\dots,\overline{m_{k-1}},\overline{n})$. Considering such a formula $\varphi(x_0,\dots,x_k)$, since $\mathsf{Q} \vdash \forall x_0 \cdots \forall x_{k-1}.\exists! y.\varphi(x_0,\dots,x_{k-1},y)$, it is clear that

$$g(m_0,\dots,m_{k-1}) \simeq \text{the first } n \text{ that satisfies: } \mathsf{Q} \vdash \varphi(\overline{\vec{m}},\overline{n}) \text{ and}$$
$$\text{for all } k < n \text{ it holds } \mathsf{Q} \vdash \neg\varphi(\overline{\vec{m}},\overline{k}),$$

as wanted.

[2]Here $\varphi(x_0,\dots,x_n)$ is used with the meaning that $\mathsf{var}(\varphi) \subseteq \{x_0,\dots,x_n\}$, where $\mathsf{var}(\varphi)$ denotes the set of the free-variables of φ.

Conversely, it is clear that if $\varphi(\vec{x})$ is a formula, then

$$f(\vec{m}) := \text{the first } n \text{ that satisfies: } \mathsf{Q} \vdash \varphi(\overline{\vec{m}}, \overline{n}) \text{ and}$$
$$\text{for all } k < n \text{ it holds } \mathsf{Q} \vdash \neg\varphi(\overline{\vec{m}}, \overline{k})$$

is a (partial) recursive function, since the following algorithm computes the desired function, where $Proof(x,y) \subseteq \mathbb{N}^2$ is the recursive relation defined by $\mathtt{Proof}_{\mathsf{Q}}(x,y)$:

```
input  : m⃗
output : n

c ← 0, j ← 0, a ← 0;
while a = 0 do
    if Proof(#¬φ(m̄⃗, c̄), j) then
        c ← c + 1, j ← 0;
    end
    if Not Proof(#φ(m̄⃗, c̄), j) and Not Proof(#¬φ(m̄⃗, c̄), j) then
        j ← j + 1;
    end
    if Proof(#φ(m̄⃗, c̄), j) then
        a ← 1;
    end
end
return c
```

We are making use of the known fact that, given a formula, we can recursively specify its arity, and the order of the variables. Clearly, for sentences the algorithm always outputs 0 whenever the sentence is Q-provable. ⊣

By the previous result, the set of Gödel-numbers of the formulas constitute a set of indices for the partial recursive functions. More precisely, given a formula φ, one can consider, in a recursive manner, the (partial) function defined by φ using the previous proof. One can develop the theory of recursive functions using the described indices and obtain all the standard results. In particular, one can consider the notation $\{\#\varphi\}$ to denote the (partial) function defined by the formula φ in the previous way.

One interesting fact about the previous result is that G1 is related to the partiality of functions (these ideas and constructions will be used in the last result of this section, Corollary 2.1.2). It is a known fact that G1 can be obtained from Kleene's Normal Form [90, p. 312]. The following result establishes a relation between two seemingly unrelated forms of diagonalization—the Diagonalization Lemma and the Second Recursion Theorem. Versions of it can be found in Theorems 0.6.9 and 0.6.12 of [92, pp. 50, 52], Section 5.2 of [95], and [6]. Here we state it more explicitly and for the very weak theory Q.

Theorem 2.1.3 (Second Recursion Theorem for Formulas). *For each formula $\varphi(x,y,z)$, there is a formula $\psi(x,y)$ such that, for each $n \in \mathbb{N}$,*

$$\{\#\psi\}(n) \simeq \{\#\varphi\}(\#\psi, n).$$

Proof. From a version of the two-variables Diagonalization Lemma (this result works for $T = \mathsf{Q}$) for more variables, we know that there is a formula $\psi(x,y)$ such that

$$\mathsf{Q} \vdash \psi(x,y) \leftrightarrow \varphi(\ulcorner \psi \urcorner, x, y).$$

Thus, for $n, m \in \mathbb{N}$ we have that:

1.) $\mathsf{Q} \vdash \psi(\overline{n}, \overline{m}) \iff \mathsf{Q} \vdash \varphi(\ulcorner \psi \urcorner, \overline{n}, \overline{m})$;

2.) $\mathsf{Q} \vdash \neg\psi(\overline{n}, \overline{m}) \iff \mathsf{Q} \vdash \neg\varphi(\ulcorner \psi \urcorner, \overline{n}, \overline{m})$.

Hence, from Theorem 2.1.2, we conclude that $\{\#\psi\}(n) \simeq \{\#\varphi\}(\#\psi, n)$. ⊣

We now present another proof of G1 using the very existence of partial functions.

Proof of Theorem 2.0.1. Suppose, aiming a contradiction, that T is complete. Consider $f : \mathbb{N} \rightharpoonup \mathbb{N}$ a partial recursive function arbitrarily fixed such that there is $n_0 \in \mathbb{N}$ where f is not defined. By Theorem 2.1.1, we conclude that there is a formula $\varphi(x,y)$ that strongly represents f in T. In particular, by **R2**, $T \vdash \forall x.\exists! y.\varphi(x,y)$. So, $T \vdash \exists y.\varphi(\overline{n_0}, y)$. Thus, $\mathbb{N} \models \exists y.\varphi(\overline{n_0}, y)$. Therefore, by definition of a model, there is an $m_0 \in \mathbb{N}$ such that $\mathbb{N} \models \varphi(\overline{n_0}, \overline{m_0})$. As, by hypothesis, T is complete, it follows that[3] $T \vdash \varphi(\overline{n_0}, \overline{m_0})$. By **R1** it follows that $f(n_0) \simeq m_0$, which is a contradiction because we assumed that f was not defined for n_0. ⊣

The next result generalizes two important facts: G1 and Rice's Theorem. It captures the main reasoning behind those two theorems that state forms of a negative result: the former incompleteness, the latter undecidability. As the reader will confirm, the main feature that is being used is a version of the Diagonalization Lemma.

Theorem 2.1.4 (General Incompleteness). *There are* no *formulas $R(x,y,z,t)$, $\Pi(x)$, $\Phi(x)$, φ, ψ, and χ satisfying the following conditions:*

C1: *For every formula α, $T \vdash \Pi(\ulcorner \alpha \urcorner) \implies T \vdash R(\ulcorner \varphi \urcorner, \ulcorner \alpha \urcorner, \ulcorner \psi \urcorner, \overline{n})$, with $n \in \mathbb{N}$ arbitrary*[4]*;*

C2: *For every formula α, $T \vdash \neg\Pi(\ulcorner \alpha \urcorner) \implies T \vdash R(\ulcorner \varphi \urcorner, \ulcorner \alpha \urcorner, \ulcorner \chi \urcorner, \overline{n})$, with $n \in \mathbb{N}$ arbitrary;*

[3] Let us prove that, under the assumption that T is complete, then $\mathbb{N} \models \varphi \implies T \vdash \varphi$.

Proof. Let φ be a true sentence, *videlicet* $\mathbb{N} \models \varphi$. As T is complete, either $T \vdash \varphi$ or $T \vdash \neg\varphi$. Assume, by way of contradiction, that $T \nvdash \varphi$. Then, $T \vdash \neg\varphi$; as T is sound, it follows $\mathbb{N} \models \neg\varphi$. But then $\mathbb{N} \models \varphi \land \neg\varphi$, a contradiction. This yields $T \vdash \varphi$. ⊣

[4] That is to say, universally quantified.

C3: *There is a formula* φ_0 *such that* $T \vdash R(\ulcorner\varphi\urcorner, \ulcorner\varphi_0\urcorner, \ulcorner\varphi_0\urcorner, \overline{n})$*, with* $n \in \mathbb{N}$ *arbitrary, and* $T \vdash \Phi(\ulcorner\varphi_0\urcorner)$*;*

C4: *For formulas* α *and* β *satisfying* $T \vdash \Phi(\ulcorner\alpha\urcorner) \wedge \Phi(\ulcorner\beta\urcorner)$*, if* $T \vdash R(\ulcorner\varphi\urcorner, \ulcorner\alpha\urcorner, \ulcorner\psi\urcorner, \overline{n}) \wedge R(\ulcorner\varphi\urcorner, \ulcorner\alpha\urcorner, \ulcorner\beta\urcorner, \overline{n})$*, for all* $n \in \mathbb{N}$*, then* $T \vdash \neg\Pi(\ulcorner\beta\urcorner)$*;*

C5: *For formulas* α *and* β *satisfying* $T \vdash \Phi(\ulcorner\alpha\urcorner) \wedge \Phi(\ulcorner\beta\urcorner)$*, if* $T \vdash R(\ulcorner\varphi\urcorner, \ulcorner\alpha\urcorner, \ulcorner\chi\urcorner, \overline{n}) \wedge R(\ulcorner\varphi\urcorner, \ulcorner\alpha\urcorner, \ulcorner\beta\urcorner, \overline{n})$*, for all* $n \in \mathbb{N}$*, then* $T \vdash \Pi(\ulcorner\beta\urcorner)$*;*

C6: *For all formulas* α *such that* $T \vdash \Phi(\ulcorner\alpha\urcorner)$*,* $T \nvdash \Pi(\ulcorner\alpha\urcorner) \implies T \vdash \neg\Pi(\ulcorner\alpha\urcorner)$*.*

In the previous result: Π stands for an intuitive notion of provability/truth; R stands for a relation that has some shared properties with a general form of equivalence; Φ is identifying the considered domain (formulas, sentence, and so on); ψ represents a general form of $\top$; χ represents a form of $\bot$; and φ a formula to which a form of negative diagonalization is going to be applied.

Proof of Theorem 2.1.4. Suppose, aiming a contradiction, that R, Π, Φ, φ, ψ, and χ do satisfy C1–C6. By **C3**, there is a formula φ_0 such that $T \vdash R(\ulcorner\varphi\urcorner, \ulcorner\varphi_0\urcorner, \ulcorner\varphi_0\urcorner, \overline{n})$, for each $n \in \mathbb{N}$, and $T \vdash \Phi(\ulcorner\varphi_0\urcorner)$. Suppose, aiming a contradiction, that $T \vdash \Pi(\ulcorner\varphi_0\urcorner)$. Then, by **C1**, $T \vdash R(\ulcorner\varphi\urcorner, \ulcorner\varphi_0\urcorner, \ulcorner\psi\urcorner, \overline{n})$, and so $T \vdash R(\ulcorner\varphi\urcorner, \ulcorner\varphi_0\urcorner, \ulcorner\psi\urcorner, \overline{n}) \wedge R(\ulcorner\varphi\urcorner, \ulcorner\varphi_0\urcorner, \ulcorner\varphi_0\urcorner, \overline{n})$. By **C4** follows that $T \vdash \neg\Pi(\ulcorner\varphi_0\urcorner)$, which is a contradiction. Hence, $T \nvdash \Pi(\ulcorner\varphi_0\urcorner)$, and so by **C6**, $T \vdash \neg\Pi(\ulcorner\varphi_0\urcorner)$. By **C2** and by what was previously concluded, it follows that $T \vdash R(\ulcorner\varphi\urcorner, \ulcorner\varphi_0\urcorner, \ulcorner\chi\urcorner, \overline{n}) \wedge R(\ulcorner\varphi\urcorner, \ulcorner\varphi_0\urcorner, \ulcorner\varphi_0\urcorner, \overline{n})$. Therefore, by **C5**, $T \vdash \Pi(\ulcorner\varphi_0\urcorner)$, which is an absurdity. ⊣

It is important to observe that **C3** plays a major role in the proof; furthermore, it has clear similarities with the Diagonalization Lemma (in a sense, the φ is being diagonalised by φ_0). In fact, for the two next corollaries, **C3** will follow immediately from that result. We now prove that G1 is a particular instance of the previous Theorem, namely an instance that does not use all the variables.

Corollary 2.1.1 (Adapted G1)**.** *Suppose that* $T \nvdash \mathsf{Pr}_T(\ulcorner\bot\urcorner)$*. It is* not *the case that for all sentences* α*,* $T \vdash \mathsf{Pr}_T(\ulcorner\alpha\urcorner)$ *or* $T \vdash \mathsf{Pr}_T(\ulcorner\neg\alpha\urcorner)$*.*

Proof. Suppose, aiming a contradiction, that for all sentences α, $T \vdash \mathsf{Pr}_T(\ulcorner\alpha\urcorner)$ or $T \vdash \mathsf{Pr}_T(\ulcorner\neg\alpha\urcorner)$. Consider $\Pi(x) := \mathsf{Pr}_T(x)$, $\Phi(x) := \mathsf{Sent}(x)$ (represents the sentences in T), $\varphi := \neg\mathsf{Pr}_T(x)$, $\psi := \bot$, $\chi := \top$, and R such that $T \vdash R(\ulcorner\varphi_0\urcorner, \ulcorner\varphi_1\urcorner, \ulcorner\varphi_2\urcorner, \overline{n}) \leftrightarrow \mathsf{Pr}_T(\ulcorner\varphi_0(\ulcorner\varphi_1\urcorner) \leftrightarrow \varphi_2(\overline{n})\urcorner)$. One should keep in mind that if α is a sentence, we assume that $T \vdash \alpha(\overline{n}) \leftrightarrow \alpha$. Let us see that the conditions of the previous Theorem are satisfied. Clearly, $T \vdash \mathsf{Pr}_T(\ulcorner\alpha\urcorner) \implies T \vdash \mathsf{Pr}_T(\ulcorner\neg\mathsf{Pr}_T(\ulcorner\alpha\urcorner) \leftrightarrow \bot\urcorner)$; this means that **C1** holds. Similarly, we have $T \vdash \neg\mathsf{Pr}_T(\ulcorner\alpha\urcorner) \implies T \vdash \mathsf{Pr}_T(\ulcorner\neg\mathsf{Pr}_T(\ulcorner\alpha\urcorner) \leftrightarrow \top\urcorner)$, which means that **C2** is satisfied. Condition **C3** corresponds to the Diagonalization Lemma (with φ_0 the Gödel-sentence). Suppose that $T \nvdash \mathsf{Pr}_T(\ulcorner\alpha\urcorner)$, with α a sentence. Then, by hypothesis, $T \vdash$

$\mathsf{Pr}_T(\ulcorner\neg\alpha\urcorner)$, hence $T\vdash\mathsf{Pr}_T(\ulcorner\alpha\urcorner)\to\mathsf{Pr}_T(\ulcorner\bot\urcorner)$. By hypothesis again, either $T\vdash\mathsf{Pr}_T(\ulcorner\mathsf{Pr}_T(\ulcorner\bot\urcorner)\urcorner)$ or $T\vdash\mathsf{Pr}_T(\ulcorner\neg\mathsf{Pr}_T(\ulcorner\bot\urcorner)\urcorner)$. As $T\nvdash\mathsf{Pr}_T(\ulcorner\bot\urcorner)$, then $T\nvdash\mathsf{Pr}_T(\ulcorner\mathsf{Pr}_T(\ulcorner\bot\urcorner)\urcorner)$, and so $T\vdash\mathsf{Pr}_T(\ulcorner\neg\mathsf{Pr}_T(\ulcorner\bot\urcorner)\urcorner)$, consequently $T\vdash\neg\mathsf{Pr}_T(\ulcorner\bot\urcorner)$. Hence, $T\vdash\neg\mathsf{Pr}_T(\ulcorner\alpha\urcorner)$, which confirms **C6**. Let us see that the remaining conditions are satisfied. Suppose that we have $T\vdash\mathsf{Pr}_T(\ulcorner\neg\mathsf{Pr}_T(\ulcorner\alpha\urcorner)\leftrightarrow\bot\urcorner)\wedge\mathsf{Pr}_T(\ulcorner\neg\mathsf{Pr}_T(\ulcorner\alpha\urcorner)\leftrightarrow\beta\urcorner)$, with α and β sentences. Hence, $T\vdash\mathsf{Pr}_T(\ulcorner\neg\beta\urcorner)$. Suppose, aiming a contradiction, that $T\vdash\mathsf{Pr}_T(\ulcorner\beta\urcorner)$. Then, $T\vdash\mathsf{Pr}_T(\ulcorner\bot\urcorner)$, which is a contradiction. So, $T\nvdash\mathsf{Pr}_T(\ulcorner\beta\urcorner)$. From a previously made reasoning, it follows that $T\vdash\neg\mathsf{Pr}_T(\ulcorner\beta\urcorner)$. This means that **C4** holds. The confirmation that **C5** holds is similar. The result follows from the previous theorem. ⊣

Now we prove that Rice's Theorem [71, p. 150] is a particular instance of Theorem 2.1.4. The proof that we will present will be long since we will show that it can be carried inside T, the increment is due to the technical apparatus and not to some possible distance from Theorem 2.1.4. As the reader will see, the mentioned Theorem and its proof perfectly fit as a generalization of the next fact.

Corollary 2.1.2 (Rice's Theorem). *Let P be a property of functions and $S:=\{\#\varphi|\{\#\varphi\}$ has the property $P\}$ such that there are $\#\psi\notin S$ and $\#\chi\in S$, with $\{\#\psi\}$ and $\{\#\chi\}$ total functions. Then, S is not decidable. All this can be proved in T.*

Proof. Suppose that S is decidable. Consider f the recursive function such that

$$f(n):=\begin{cases}1, & n\in S\\ 0, & n\notin S.\end{cases}$$

As f is recursive, by Theorem 2.1.1, there is $\Psi(x,y)$ that strongly represents it. Consider Eq such that $T\vdash\mathsf{Eq}(\ulcorner\alpha\urcorner,\langle\overline{n_0},\dots,\overline{n_k}\rangle,x)\leftrightarrow\mathsf{Pr}_T(\ulcorner\alpha(\overline{n_0},\dots,\overline{n_k},\dot{x})\urcorner)\wedge\forall y<x.\mathsf{Pr}_T(\ulcorner\neg\alpha(\overline{n_0},\dots,\overline{n_k},\dot{y})\urcorner)$, for every formula α. Take φ such that

$$\{\#\varphi\}(n,m):=\begin{cases}\{\#\psi\}(m), & n\in S\\ \{\#\chi\}(m), & n\notin S.\end{cases}$$

Consider R such that $T\vdash R(\ulcorner\alpha\urcorner,\ulcorner\beta\urcorner,\ulcorner\gamma\urcorner,\overline{n})\leftrightarrow\exists x.\mathsf{Eq}(\ulcorner\alpha\urcorner,\langle\ulcorner\beta\urcorner,\overline{n}\rangle,x)\wedge$ $\mathsf{Eq}(\ulcorner\gamma\urcorner,\langle\overline{n}\rangle,x)$. Take $\Phi(x)$ defining all formulas (we can omit it), and $\Pi(x):=\Psi(x,\overline{1})$. Let us see that the conditions of the Theorem are satisfied. Observe that, by hypothesis, $\{\#\varphi\}$ is total.

Suppose that $T\vdash\Pi(\ulcorner\alpha\urcorner)$. Then, $\#\alpha\in S$, and so $\{\#\varphi\}(\#\alpha,n)\simeq\{\#\psi\}(n)$, with $n\in\mathbb{N}$. Fix $n\in\mathbb{N}$. Take $n_0:=\{\#\psi\}(n)$ (it is defined because $\{\#\psi\}$ is total by construction). Then, $T\vdash\psi(\overline{n},\overline{n_0})$ and, for all $k<n_0$, $T\vdash\neg\psi(\overline{n},\overline{k})$. Hence, $T\vdash\mathsf{Pr}_T(\ulcorner\psi(\overline{n},\overline{n_0})\urcorner)$ and, for all $k<n_0$, $T\vdash\mathsf{Pr}_T(\ulcorner\neg\psi(\overline{n},\overline{k})\urcorner)$. As $T\vdash k<\overline{n_0}\leftrightarrow k=0\vee k=\overline{1}\vee\dots\vee k=\overline{n_0-1}$, it follows that $T\vdash\mathsf{Eq}(\ulcorner\psi\urcorner,\langle\overline{n}\rangle,\overline{n_0})$. As $\{\#\varphi\}(\#\alpha,n)\simeq n_0$, we conclude that $T\vdash\varphi(\ulcorner\alpha\urcorner,\overline{n},\overline{n_0})$ and, for all $k<n_0$, $T\vdash\neg\varphi(\ulcorner\alpha\urcorner,\overline{n},\overline{k})$. Thus, $T\vdash\mathsf{Pr}_T(\ulcorner\varphi(\ulcorner\alpha\urcorner,\overline{n},\overline{n_0})\urcorner)$ and, for all $k<n_0$, $T\vdash\mathsf{Pr}_T(\ulcorner\neg\varphi(\ulcorner\alpha\urcorner,\overline{n},\overline{k})\urcorner)$. This means that $T\vdash\mathsf{Eq}(\ulcorner\varphi\urcorner,\langle\ulcorner\alpha\urcorner,\overline{n}\rangle,\overline{n_0})$. In all, $T\vdash R(\ulcorner\varphi\urcorner,\ulcorner\alpha\urcorner,\ulcorner\psi\urcorner,\overline{n})$. All together means that **C1** holds.

Suppose that $T \vdash \neg\Pi(\ulcorner\alpha\urcorner)$. By the consistency of T, $T \nvdash F(\ulcorner\alpha\urcorner)$. As $T \vdash \Psi(\ulcorner\alpha\urcorner, 0)$ or $T \vdash \Psi(\ulcorner\alpha\urcorner, \overline{1})$, we conclude that $T \vdash \Psi(\ulcorner\alpha\urcorner, 0)$, and so $\#\alpha \notin S$. Following the reasoning from before, it is easy to see that $T \vdash R(\ulcorner\varphi\urcorner, \ulcorner\alpha\urcorner, \ulcorner\chi\urcorner, \overline{n})$, so **C2** is satisfied. Take φ_0 the formula given by Theorem 2.1.3. Then, $\{\#\varphi\}(\#\varphi_0, n) \simeq \{\#\varphi_0\}(n)$. Following the reasoning of **C1**, it is easy to see that $T \vdash R(\ulcorner\varphi\urcorner, \ulcorner\varphi_0\urcorner, \ulcorner\varphi_0\urcorner, \overline{n})$, and so **C3** holds.

Consider α and β such that, for all $n \in \mathbb{N}$, $T \vdash R(\ulcorner\varphi\urcorner, \ulcorner\alpha\urcorner, \ulcorner\psi\urcorner, \overline{n}) \wedge R(\ulcorner\varphi\urcorner, \ulcorner\alpha\urcorner, \ulcorner\beta\urcorner, \overline{n})$. Then, $T \vdash \exists x.\mathsf{Eq}(\ulcorner\varphi\urcorner, \langle\ulcorner\alpha\urcorner, \overline{n}\rangle, x) \wedge \mathsf{Eq}(\ulcorner\psi\urcorner, \langle\overline{n}\rangle, x)$ and $T \vdash \exists x.\mathsf{Eq}(\ulcorner\varphi\urcorner, \langle\ulcorner\alpha\urcorner, \overline{n}\rangle, x) \wedge \mathsf{Eq}(\ulcorner\beta\urcorner, \langle\overline{n}\rangle, x)$. So, there is $m \in \mathbb{N}$ such that $\mathbb{N} \models \mathsf{Eq}(\ulcorner\varphi\urcorner, \langle\ulcorner\alpha\urcorner, \overline{n}\rangle, \overline{m}) \wedge \mathsf{Eq}(\ulcorner\psi\urcorner, \langle\overline{n}\rangle, \overline{m})$. Hence, $\mathbb{N} \models \mathsf{Pr}_T(\ulcorner\varphi(\ulcorner\alpha\urcorner, \overline{n}, \overline{m})\urcorner) \wedge \forall k < \overline{m}.\mathsf{Pr}_T(\ulcorner\neg\varphi(\ulcorner\alpha\urcorner, \overline{n}, \dot{k})\urcorner)$, so $\{\#\varphi\}(\#\alpha, n) = m$. Similarly, $\{\#\psi\}(n) = m$, and so $\{\#\varphi\}(\#\alpha, n) = \{\#\psi\}(n)$. As this holds for each $n \in \mathbb{N}$, it follows that $\lambda x.\{\#\varphi\}(\#\alpha, x) = \{\#\psi\}$. In a similar, way one has that $\lambda x.\{\#\varphi\}(\#\alpha, x) = \{\#\beta\}$. Hence, $\{\#\psi\} = \{\#\beta\}$, and so $\#\beta \notin S$. This means that $T \vdash \Psi(\ulcorner\beta\urcorner, 0)$. As $\Psi(x, y)$ strongly represents f, in particular $T \vdash \forall x.\exists! y.\Psi(x, y)$, consequently $T \vdash \neg\Psi(\ulcorner\beta\urcorner, \overline{1})$, i.e., $T \vdash \neg\Pi(\ulcorner\beta\urcorner)$. This confirms **C4**. The confirmation of **C5** is very similar. Finally, suppose that $T \nvdash \Pi(\ulcorner\alpha\urcorner)$. As $T \vdash \Psi(\ulcorner\alpha\urcorner, 0)$ or $T \vdash \Psi(\ulcorner\alpha\urcorner, \overline{1})$, it follows that $T \vdash \Psi(\ulcorner\alpha\urcorner, 0)$. As before, it follows that $T \vdash \neg\Pi(\ulcorner\alpha\urcorner)$, which confirms **C6**. ⊣

The main idea of the previous proof was the construction of the relation R, that is intended to capture functional equality, via the relation Eq. A more general result could be established, namely one where we would not require the totality of $\{\#\psi\}$ and $\{\#\chi\}$. That would correspond to a change in the relation R (intuitively, instead of claiming existence it would claim unicity)—we decided to present the previous version since the increment in generality would not correspond to the greater increment in complexity of the relation.

We can conclude that the Diagonalization Lemma was the main feature of Theorem 2.1.4, making it responsible for G1 and, more unexpectedly, for Rice's Theorem.

2.2 Incompleteness via Recursion and the Non-Recursiveness of Truth

In this section, we present a version of G1 using the "language" of recursion theory, more concretely, we state G1 in terms of recursive functions. We start with several well-known facts. Our approach is similar to the one considered in the study of Weihrauch Complexity [8].

Folklore 2.2.1. *The following statements are equivalent:*

1 *A is a recursively enumerable set;*

2 *There is a (partial) recursive function f such that, for all $n \in \mathbb{N}$,*

$$f(n) = 0 \iff n \in A;$$

3 *$A = \emptyset$ or there is a total recursive function f such that $f(\mathbb{N}) = A$.*

Folklore 2.2.2. *The set* $\mathtt{Prov}_T := \{\#\varphi | T \vdash \varphi\}$ *is recursively enumerable.*

The next result uses the Diagonalization Lemma, Gödels' original paper [37] is the first place where we have a clear indication of it.

Lemma 2.2.1. *The set* $\mathtt{True}_{\mathbb{N}} := \{\#\varphi | \mathbb{N} \models \varphi\}$ *is not recursively enumerable.*

Proof. (This result can be established using Post's Theorem, but we decided to show the reader yet another use of the Diagonalization Lemma.) Suppose, aiming a contradiction, that $\mathtt{True}_{\mathbb{N}}$ is recursively enumerable. Then, by Folklore 2.2.1, there is a recursive function f such that

$$f(n) = 0 \iff n \in \mathtt{True}_{\mathbb{N}}.$$

Let e be the index of f. Consider $T(x,y,z)$ as being the Kleene T-Predicate and $U(x)$ as being the primitive recursive function created by Kleene to output the result of computations. It is a known fact that $T(x,y,z)$ is a primitive recursive relation. Consider $\mathsf{T}(x,y,z)$ a Σ_1-formula that represents $T(x,y,z)$, and consider $\mathsf{U}(x,y)$ a Σ_1-formula that represents the function U. Now, take the $\Sigma_1(T)$-formula

$$\Phi(x,y) := \exists z.(\mathsf{T}(\overline{e},x,z) \wedge \mathsf{U}(z,y)).$$

By the Diagonalization Lemma, there is a sentence φ such that

$$T \vdash \varphi \leftrightarrow \neg\Phi(\ulcorner\varphi\urcorner, 0).$$

It is clear that $\mathbb{N} \models \varphi$ or $\mathbb{N} \models \neg\varphi$. Suppose that $\mathbb{N} \models \varphi$. Then, $f(\#\varphi) = 0$, and so there is $w \in \mathbb{N}$ such that $T(e, \#\varphi, w)$ and $U(w) = 0$. This means that $T \vdash \mathsf{T}(\overline{e}, \ulcorner\varphi\urcorner, \overline{w}) \wedge \mathsf{U}(\overline{w}, 0)$. Thus, $T \vdash \exists z.(\mathsf{T}(\overline{e}, \ulcorner\varphi\urcorner, z) \wedge \mathsf{U}(z, 0))$, i.e., $T \vdash \Phi(\ulcorner\varphi\urcorner, 0)$. Therefore, $\mathbb{N} \models \neg\varphi$, which is a contradiction. Hence, $\mathbb{N} \models \neg\varphi$. So, $\mathbb{N} \models \Phi(\ulcorner\varphi\urcorner, 0)$. This means that there is $w \in \mathbb{N}$ such that $\mathbb{N} \models \mathsf{T}(\overline{e}, \ulcorner\varphi\urcorner, \overline{w}) \wedge \mathsf{U}(\overline{w}, 0)$. As $\mathsf{T}(\overline{e}, \ulcorner\varphi\urcorner, \overline{w}) \wedge \mathsf{U}(\overline{w}, 0)$ is a $\Sigma_1(T)$-sentence and as T is Σ_1-complete, we can conclude that $T \vdash \mathsf{T}(\overline{e}, \ulcorner\varphi\urcorner, \overline{w}) \wedge \mathsf{U}(\overline{w}, 0)$. So, $T(e, \#\varphi, w)$ and $U(w) = 0$. Hence, $f(\#\varphi) = 0$, and so $\mathbb{N} \models \varphi$, which is a contradiction. ⊣

The previous result is a generalisation of Tarski's Theorem on the Undefinability of Truth, as the following consequence confirms. Besides being a generalization of Tarski's result, its interest comes from the fact that it is expressed in terms of recursive functions.

Corollary 2.2.1. [7, p. 222] *Truth is not definable in* T.

Proof. Suppose, aiming a contradiction, that truth is definable by a predicate $\mathtt{Truth}(x)$ in T. Then, for all sentences φ,

$$T \vdash \mathtt{Truth}(\ulcorner\varphi\urcorner) \iff \mathbb{N} \models \varphi.$$

Take $Proof(x,y) \subseteq \mathbb{N}^2$ as the recursive relation that expresses in $\mathbb{N}$ that "y is a T-proof of x". It is well-known that $Proof(x,y)$ is primitive recursive [78, p. 233]. Take $Not\ Proof(x,y)$

as being the relation that expresses that y is not a proof of x. Consider the following algorithm

```
input  : n
output : r
a ← 0, c ← 0, r ← 0;
while a = 0 do
    if Proof(#Truth(n̄), c) then
        a ← 1;
    end
    if Not Proof(#Truth(n̄), c) then
        c ← c + 1;
    end
end
return r
```

Let f be the function computed by the previous algorithm. Clearly,

$$n \in \mathtt{True}_{\mathbb{N}} \iff f(n) = 0,$$

from where we conclude that $\mathtt{True}_{\mathbb{N}}$ is recursively enumerable, which goes against Lemma 2.2.1. ⊣

We say that a set A *is complete with respect to a set* B if there is a total recursive function f such that $f(A) = B$. We say that A *is incomplete with respect to* B if A is not complete with respect to B.

The next fact establishes a relation between the notion of completeness in terms of recursive functions and the notion of completeness in T.

Theorem 2.2.1. *If T is complete, then $\mathtt{Prov}_T$ is complete with respect to $\mathtt{True}_{\mathbb{N}}$.*

Proof. Suppose that T is complete. Then for all sentences φ,

$$\mathbb{N} \models \varphi \iff T \vdash \varphi.$$

Consider $\mathtt{id}$ as being the identity function. Clearly $\mathtt{id}$ is a total recursive function. By the previous equivalence we conclude that $\mathtt{Prov}_T = \mathtt{True}_{\mathbb{N}}$. Hence, $\mathtt{id}(\mathtt{Prov}_T) = \mathtt{True}_{\mathbb{N}}$. From this we conclude that $\mathtt{Prov}_T$ is complete with respect to $\mathtt{True}_{\mathbb{N}}$. ⊣

The concept of completeness with respect to a given set is related to the concepts used in the study of Weihrauch Complexity (see [8] for details on this topic).

The next result corresponds to Lemma 3.2.1. from [50, p. 51]. It is a very naïve fact and seemingly innocent, but we are going to use it to frame the incompleteness of T using the concepts we have just introduced; hence, we decided to call the result "Recursive Transfer" because we are *transfering* recursive enumerability.

Theorem 2.2.2 (Recursive Transfer)**.** *If A is a recursively enumerable set and B is not recursively enumerable, then A is incomplete with respect to B.*

Proof. Suppose that A is a recursively enumerable set and B is not recursively enumerable. Consider a total recursive function $f : \mathbb{N} \to A$ in the conditions of Folklore 2.2.1, i.e. such that $f(\mathbb{N}) = A$. Suppose, aiming a contradiction, that A is complete with respect to B. Then, there is a total recursive function g such that $g(A) = B$. Take $h := g \circ f$. Clearly, $h : \mathbb{N} \to B$ is a total recursive function such that $h(\mathbb{N}) = B$, hence, by Folklore 2.2.1, B is recursively enumerable, which is a contradiction. ⊣

Finally, we can conclude that G1 follows from the Recursive Transfer.

Proof of Theorem 2.0.1. By Theorem 2.2.1, it suffices to prove that Prov_T is incomplete with respect to $\mathsf{True}_\mathbb{N}$. By Folklore 2.2.2, we know that Prov_T is recursively enumerable, and by Lemma 2.2.1 we have that $\mathsf{True}_\mathbb{N}$ is not recursively enumerable. Thus, by the Recursive Transfer, we can conclude that Prov_T is incomplete with respect to $\mathsf{True}_\mathbb{N}$. ⊣

2.3 General use of the Diagonalization Lemma

In this last section, we present a result that generalizes the use of the Diagonalization Lemma and the Term-Diagonalization Lemma. We are aware that there are general results that trace back several important diagonalization reasonings to a common theorem (for instance [83, §5], [83, §6], and [104]). Although such approaches are very relevant, since they generalize the use of diagonalization, we believe that they lose information when one wants to see how far one can get just using the Diagonalization Lemmas (there, the Diagonalization Lemmas correspond just to a function feature, they are not relevant facts by their own right). Our objective with this section is not to present the use of diagonalization in arithmetic in its most general form, rather we are interested in studying just the use of the Diagonalization Lemmas and concluding that they are enough to get the majority of the relevant meta-theorems of arithmetic. Our goal is to unify diagonalization of formulas and diagonalization of terms in arithmetic. As we will see, the next result delivers our goal, it can capture both types of reasoning.

Theorem 2.3.1 (General Diagonalization)**.** *Suppose that T has function-symbols* sb *and* ng, *and formulas $\Phi(x)$, $\Phi'(x)$, and $R(x,y)$ satisfying:*

GD1: *$T \vdash$ "R is an equivalence relation" $\land \forall x.(\Phi(x) \to \Phi'(x))$;*

GD2: *For all $n \in \mathbb{N}$ satisfying $T \vdash \Phi'(\overline{n})$, there is $m \in \mathbb{N}$ satisfying $T \vdash \Phi(\overline{m})$ such that $T \vdash R(\mathsf{sb}(\overline{n},\overline{m}),\overline{m})$;*

GD3: *For all $n \in \mathbb{N}$ such that $T \vdash \Phi(\overline{n})$, $T \vdash \neg R(\mathsf{ng}(\overline{n}),\overline{n})$;*

GD4: *If $n, m \in \mathbb{N}$ satisfy $T \vdash \Phi'(\overline{n}) \land \Phi(\overline{m})$, then $T \vdash R(\mathsf{ng}(\mathsf{sb}(\overline{n},\overline{m})),\mathsf{sb}(\mathsf{ng}(\overline{n}),\overline{m}))$;*

GD5: *For all $n \in \mathbb{N}$ satisfying $T \vdash \Phi'(\overline{n})$, there is $m \in \mathbb{N}$ such that $T \vdash \Phi'(\overline{m}) \wedge \overline{m} = \mathsf{ng}(\overline{n})$. We assume the same property substituting Φ' for Φ;*

GD6: *If $n, m \in \mathbb{N}$ satisfy $T \vdash \Phi'(\overline{n}) \wedge \Phi(\overline{m})$, then there is $s \in \mathbb{N}$ such that $T \vdash \Phi(\overline{s}) \wedge \overline{s} = \mathsf{sb}(\overline{n}, \overline{m})$.*

Then, there is no *$n \in \mathbb{N}$ satisfying $T \vdash \Phi'(\overline{n})$ such that: for all $m \in \mathbb{N}$ satisfying $T \vdash \Phi(\overline{m})$, we have $T \vdash R(\mathsf{sb}(\overline{n}, \overline{m}), \overline{m})$.*

Proof. Suppose, aiming a contradiction, that $n \in \mathbb{N}$ satisfies the desired conditions. From conditions **GD5** and **GD2**, there is $m \in \mathbb{N}$ satisfying $T \vdash \Phi(\overline{m})$ and $T \vdash R(\mathsf{sb}(\mathsf{ng}(\overline{n}), \overline{m}), \overline{m})$. From **GD4** we get $T \vdash R(\mathsf{ng}(\mathsf{sb}(\overline{n}, \overline{m})), \mathsf{sb}(\mathsf{ng}(\overline{n}), \overline{m}))$, from **GD1**, $T \vdash R(\mathsf{ng}(\mathsf{sb}(\overline{n}, \overline{m})), \overline{m})$. By the hypothesis on n, $T \vdash R(\mathsf{sb}(\overline{n}, \overline{m}), \overline{m})$. Together with what was concluded and using **GD1**, one has that $T \vdash R(\mathsf{ng}(\mathsf{sb}(\overline{n}, \overline{m})), \mathsf{sb}(\overline{n}, \overline{m}))$, and so, from **GD6** and **GD3**, $T \vdash \bot$, which by its turn goes against the consistency of T. ⊣

In the previous Theorem, the relation R can be interpreted as being equality or logical equivalence. In that context, **GD2** is then a generalization of the Diagonalization Lemma and of the Term-Diagonalization Lemma. Furthermore, ng is intended to represent a function that always gives something different—"through the eyes of R"—from the initial object, this constitutes a weak form of negation. sb is intended to represent a substitution function compatible with ng according to R. Condition **GD4** expresses how substitution acts according to R. Conditions **GD5** and **GD6** are technical ones.

Definition 2.3.1. We say that T is a *uniform theory of arithmetic* if:

U1: T is a consistent extension of EA;

U2: There are a formula $\mathtt{Fml}(x)$ in T that identifies all formulas, i.e., $T \vdash \mathtt{Fml}(\ulcorner \varphi \urcorner)$ if, and only if, φ is a formula; a formula $\mathtt{Sent}(x)$ that identifies all sentences; and a formula $\mathtt{oneVarFml}(x)$ that identifies 1-variable formulas (with possibly several occurrences) and also sentences[5];

U3: There is a function-symbol $\dot{\neg}$ that represents negation, i.e., such that for all formulas φ, $T \vdash \dot{\neg}(\ulcorner \varphi \urcorner) = \ulcorner \neg \varphi \urcorner$. We will write $\dot{\neg} x := \dot{\neg}(x)$;

U4: There is a function-symbol $\dot{\leftrightarrow}$ that represents equivalence, i.e., such that for all formulas φ and ψ, $T \vdash \dot{\leftrightarrow} (\ulcorner \varphi \urcorner, \ulcorner \psi \urcorner) = \ulcorner \varphi \leftrightarrow \psi \urcorner$. We will write $x \dot{\leftrightarrow} y := \dot{\leftrightarrow} (x, y)$;

U5: There are a function-symbol $\mathtt{sub}(x, y)$ and $\mathtt{numeral}$ such that, for all formulas $\varphi(x)$ and ψ, $T \vdash \mathtt{sub}(\ulcorner \varphi(x) \urcorner, \ulcorner \psi \urcorner) = \ulcorner \varphi(\ulcorner \psi \urcorner) \urcorner$ and $T \vdash \mathtt{numeral}(\overline{n}) = \ulcorner \overline{n} \urcorner$. We assume that if φ is a sentence, then $T \vdash \mathtt{sub}(\ulcorner \varphi \urcorner, \ulcorner \psi \urcorner) = \ulcorner \varphi \urcorner$. We use Feferman's dot notation: $\ulcorner \varphi(\dot{x}) \urcorner := \mathtt{sub}(\ulcorner \varphi(y) \urcorner, \mathtt{numeral}(x))$;

U6: T has a provability predicate P that satisfies the following properties:

[5]This name might confuse the reader, but one should keep in mind that oneVarFml includes sentences.

P1: $T \vdash \forall \mathtt{Fml}(x).\forall \mathtt{Fml}(y).\forall \mathtt{Fml}(z).P(x \dot{\leftrightarrow} x) \wedge (P(x \dot{\leftrightarrow} y) \wedge (P(y \dot{\leftrightarrow} z)) \to P(x \dot{\leftrightarrow} z)) \wedge (P(x \dot{\leftrightarrow} y) \to P(y \dot{\leftrightarrow} x))$;

P2: For all formulas φ, $T \vdash P(\ulcorner \neg\varphi \leftrightarrow \varphi \urcorner) \to P(\ulcorner \bot \urcorner)$;

P3: For all formulas φ and ψ, $T \vdash P(\dot{\neg}\mathtt{sub}(\ulcorner\varphi\urcorner, \ulcorner\psi\urcorner) \dot{\leftrightarrow} \mathtt{sub}(\dot{\neg}\ulcorner\varphi\urcorner, \ulcorner\psi\urcorner))$.

The next result is a generalization of Theorem 2.0.8.

Corollary 2.3.1 (Undefinability of Truth)**.** *Given T a uniform theory of arithmetic (P the considered provability predicate for T), let S be a consistent theory that contains $T + \mathtt{Con}_P$. Then, there is* no *formula $Tr(x)$ such that for all sentences ψ,*

$$S \vdash P(\ulcorner Tr(\ulcorner\psi\urcorner) \leftrightarrow \psi \urcorner).$$

Proof. Consider $\Phi(x) := \mathtt{Sent}(x)$, $\Phi'(x) := \mathtt{oneVarFml}(x)$, $\mathtt{sb}(x,y) := \mathtt{sub}(x,y)$, $\mathtt{ng}(x) := \dot{\neg}x$, and $R(x,y) := P(x \dot{\leftrightarrow} y)$. Clearly, one can obtain condition **GD1** of Theorem 2.3.1 from condition **P1**. Condition **GD2** follows from the Diagonalization Lemma. Furthermore, using condition **P2** and the fact that S extends $T + \mathtt{Con}_P$, one can obtain **GD3**. Condition **GD4** is obtainable from **P3**. Finally, condition **GD5** follows from **U3** and condition **GD6** from **U5**. The result is a direct application of Theorem 2.3.1. ⊣

Definition 2.3.2. We say that T is a *Gödel-like theory* if:

G1: T is a uniform theory of arithmetic;

G2: The provability predicate P also satisfies:

P4: For all formulas φ and ψ, $T \vdash P(\ulcorner \varphi \wedge \psi \urcorner) \leftrightarrow (P(\ulcorner\varphi\urcorner) \wedge P(\ulcorner\psi\urcorner))$;

P5: For all formulas φ, $T \vdash P(\ulcorner\varphi\urcorner) \to P(\ulcorner P(\ulcorner\varphi\urcorner)\urcorner)$;

P6: For all formulas φ and ψ, $T \vdash P(\ulcorner \varphi \to \psi \urcorner) \to (P(\ulcorner\varphi\urcorner) \to P(\ulcorner\psi\urcorner))$.

The next fact generalizes G1.

Corollary 2.3.2 (Formalized G1)**.** *Let T be a Gödel-like theory (P the considered provability predicate for T). Then, it does* not *hold that, for all sentences φ,*

$$T + \mathtt{Con}_P + \mathtt{Con}^1_P \vdash P(\ulcorner\varphi\urcorner) \vee P(\ulcorner\neg\varphi\urcorner).$$

Proof. Suppose, aiming a contradiction, that the result holds. Consider φ a sentence. From **P4**, we have that $T + P(\ulcorner \varphi \wedge \neg P(\ulcorner\varphi\urcorner)\urcorner) \vdash P(\ulcorner\varphi\urcorner) \wedge P(\ulcorner \neg P(\ulcorner\varphi\urcorner)\urcorner)$. From **P5** follows that $T + P(\ulcorner \varphi \wedge \neg P(\ulcorner\varphi\urcorner)\urcorner) \vdash P(\ulcorner P(\ulcorner\varphi\urcorner)\urcorner) \wedge P(\ulcorner \neg P(\ulcorner\varphi\urcorner)\urcorner)$. From **P6** we conclude that $T + P(\ulcorner \varphi \wedge \neg P(\ulcorner\varphi\urcorner)\urcorner) \vdash P(\ulcorner\bot\urcorner)$, hence $T + \mathtt{Con}_T \vdash \neg P(\ulcorner \varphi \wedge \neg P(\ulcorner\varphi\urcorner)\urcorner)$. By the assumption, this means that $T + \mathtt{Con}_P + \mathtt{Con}^1_P \vdash P(\ulcorner \varphi \to P(\ulcorner\varphi\urcorner)\urcorner)$.

Using **P4**, $T + P(\ulcorner P(\ulcorner\varphi\urcorner) \wedge \neg\varphi\urcorner) \vdash P(\ulcorner P(\ulcorner\varphi\urcorner)\urcorner) \wedge P(\ulcorner\neg\varphi\urcorner)$. Thus, from **P5**, $T + P(\ulcorner P(\ulcorner\varphi\urcorner) \wedge \neg\varphi\urcorner) \vdash P(\ulcorner P(\ulcorner\varphi\urcorner)\urcorner) \wedge P(\ulcorner P(\ulcorner\neg\varphi\urcorner)\urcorner)$. From **P6** and **P4**, we conclude that $T + P(\ulcorner P(\ulcorner\varphi\urcorner) \wedge$

$\neg\varphi^\urcorner) \vdash P(\ulcorner P(\ulcorner \bot^\urcorner)^\urcorner)$, and so $T + \mathsf{Con}_P + \mathsf{Con}^1_P \vdash \neg P(\ulcorner P(\ulcorner\varphi^\urcorner) \wedge \neg\varphi^\urcorner)$. By the assumption one has that $T + \mathsf{Con}^1_P \vdash P(\ulcorner P(\ulcorner\varphi^\urcorner) \to \varphi^\urcorner)$.

In all, for all sentences φ,

$$T + \mathsf{Con}_P + \mathsf{Con}^1_P \vdash P(\ulcorner P(\ulcorner\varphi^\urcorner) \leftrightarrow \varphi^\urcorner),$$

which goes against Corollary 2.3.1 for $S := T + \mathsf{Con}_P + \mathsf{Con}^1_P$. ⊣

In the previous proof, the key-idea was that if $T + \mathsf{Con}_P + \mathsf{Con}^1_P \vdash P(\ulcorner\varphi^\urcorner) \vee P(\ulcorner\neg\varphi^\urcorner)$, then $P(x)$ would be a truth-predicate in the sense of the Undefinability of Truth. This means that we proved G1 using the Undefinability of Truth. This idea has already appeared in the literature [100], but with a different use and a different purpose (it was used to prove G1 avoiding diagonalization). It is important to observe that, in the previous section, G1 was obtained in a similar way, namely by means of the non-recursiveness of truth.

Definition 2.3.3. We say that T is a *strongly-uniform theory of arithmetic* if:

SU1: T is a consistent extension of EA;

SU2: There is a formula $\mathtt{Term}(x)$ in T that identifies all terms, i.e., $T \vdash \mathtt{Term}(\ulcorner t^\urcorner)$ if, and only if, t is a term;

SU3: There is a function-symbol that represent successor, i.e., such that for all terms t, $T \vdash (\ulcorner t^\urcorner) = \ulcorner t+1^\urcorner$;

SU4: There is a function-symbol $\dot{=}$ that represent equality, i.e., such that for all terms t_0 and t_1, $T \vdash \dot{=}(\ulcorner t_0^\urcorner, \ulcorner t_1^\urcorner) = \ulcorner t_0 = t_1^\urcorner$. We will write $x \dot{=} y := \dot{=}(x,y)$;

SU5: There is a function-symbol $\mathsf{sub}(x,y)$ such that, $T \vdash \mathsf{sub}(\ulcorner t(x)^\urcorner, \ulcorner t_0^\urcorner) = \ulcorner t(\ulcorner t_0^\urcorner)^\urcorner$, for all term $t(x)$ and t_0. We assume that if $t(x)$ has no free-variables, then $T \vdash \mathsf{sub}(\ulcorner t^\urcorner, \ulcorner t_0^\urcorner) = \ulcorner t^\urcorner$;

SU6: T has a provability predicate P that satisfies the following properties:

PU1: $T \vdash \forall\mathtt{Term}(x).\forall\mathtt{Term}(y).\forall\mathtt{Term}(z).P(x \dot{=} x) \wedge (P(x \dot{=} y) \wedge (P(y \dot{=} z)) \to P(x \dot{=} z)) \wedge (P(x \dot{=} y) \to P(y \dot{=} x))$;

PU2: For all terms t, $T \vdash P(\ulcorner t = t+1^\urcorner) \to P(\ulcorner\bot^\urcorner)$;

PU3: For all terms t_0 and t_1, $T \vdash P((\mathsf{sub}(\ulcorner t_0^\urcorner, \ulcorner t_1^\urcorner)) \dot{=} \mathsf{sub}((\ulcorner t_0^\urcorner), \ulcorner t_1^\urcorner))$.

Finally, we present a generalization of Theorem 2.0.9.

Corollary 2.3.3 (Hilbert-Bernays Paradox)**.** *Let T be a strongly-uniform theory of arithmetic (P the considered provability predicate for T). Then, there is* no *term $h(x)$ such that for all closed terms t,*

$$T + \mathsf{Con}_P \vdash P(\ulcorner h(\ulcorner t^\urcorner) = t^\urcorner).$$

Proof. Take $\Phi(x), \Phi'(x) := \mathtt{Term}(x)$, $R(x,y) := P(x \doteq y)$, $\mathtt{ng}(x) := (x)$, and $\mathtt{sb}(x,y) := \mathtt{sub}(x,y)$. Consider $T + \mathtt{Con}_P$. Condition **PU1** corresponds to condition **GD1** from Theorem 2.3.1, the Term-Diagonalization Lemma corresponds to condition **GD2**, condition **PU2** corresponds to conditions **GD3**, and condition **PU3** corresponds to condition **GD4**. Furthermore, **GD5** follows from **SU3** and **GD6** follows from **SU5**. By Theorem 2.3.1, the result follows. ⊣

Chapter 3

k-provability in PA

3.1 Introduction

k-provability is the notion of provability '$\vdash_{k \text{ steps}}$', i.e. the notion of *being provable, in a certain theory, with at most k steps*. This notion has been studied for different theories and with different purposes. In [73], [31], and [58] the decidability of this relation was studied for several formalizations of Peano Arithmetic (PA). Kreisel's conjecture—an open problem in k-provability [18]—was studied in [73], [58], [32], [68], [74], [15], and [1]. We recommend [72] for a detailed account of this and other notions of provability.

In [13], it was proved that k-provability is undecidable for the sequent calculus of arithmetic with an infinite number of relation-symbols. Furthermore, in [31], this relation was proved to be decidable for several formulations of PA where the universal instantiation schema is replaced by other schemata. The usual universal instantiation schema is:

Uni. Inst $(\forall x.\varphi) \to \varphi^x_t$, where t is substitutable for x in φ.

It is an open problem whether k-provability for PA with the usual instantiation schema is decidable [31]. From [58], [31], and [72, p. 103] we know that the proof-skeleton problem is undecidable for PA with the usual instantiation schema; by the *proof-skeleton problem* we mean the problem of deciding if a given formula has a proof whose skeleton (the list of axioms and rules that were used) is the considered one.

The work we present in this chapter was awarded the Amílcar Sernadas Logic Prize[1] and was published in [85]. Here, we address the proof-skeleton problem and the decidability of k-provability; we:

[1] https://math.tecnico.ulisboa.pt/pacs/.

1. Characterize some proof-skeletons for which it is decidable whether a given formula has a proof with the considered skeleton;

2. Characterize some values of k for which it is decidable whether a formula can be proved in k steps.

These characterizations are natural ones—in the sense that they emerge from simple generalization of concepts—and parameterized by unification algorithms (for a type of systems that we are going to develop). Our approach is valid for several theories that extend PA. We will consider theories of arithmetic formulated in a Hilbert-style systems having the following logical axioms (see [28, p. 112] for further details):

L1. $(\varphi \to (\psi \to \mu)) \to ((\varphi \to \psi) \to (\varphi \to \mu))$;

L2. $\varphi \to (\psi \to \varphi)$;

L3. $(\neg\varphi \to \neg\psi) \to (\psi \to \varphi)$;

L4. $(\forall x.\varphi) \to \varphi^x_t$, where t is substitutable for x in φ;

L5. $\forall x.(\varphi \to \psi) \to (\forall x.\varphi \to \forall x.\psi)$;

L6. $\varphi \to \forall x.\varphi$, where x does not occur free in φ;

L7. $\forall x.x = x$;

L8. $\forall x.\forall y.\forall z.(x = y \wedge y = z \to x = z)$;

L9. $\forall x.\forall y.x = y \to y = x$;

L10. $\forall x_0.\forall x_1.\forall x_2.\forall x_3.(x_0 = x_1 \wedge x_2 = x_3 \to x_0 + x_2 = x_1 + x_3)$;

L11. $\forall x.\forall y.(x = y \to S(x) = S(y))$.

We do not allow the occurrence of any other predicates besides ‘=’ (for instance, we assume that one is not given a predicate ‘<’ for the usual relation $<$ in $\mathbb{N}$). Furthermore, we consider the following two rules (the rule Gen can be removed from this axiomatization, but we decided to keep it because it is useful in practise):

$$\frac{\varphi \quad \varphi \to \psi}{\psi}\text{MP} \qquad\qquad \frac{\varphi}{\forall x.\varphi}\text{Gen}$$

It is important to observe that these axioms are *schemata* in the sense that they can be substituted for any formula and any variable which satisfy certain conditions. The non-logical axioms of Robinson Arithmetic (Q) are:

Q1. $\forall x.\forall y.(S(x) = S(y) \to x = y)$;

Q2. $\forall x.\neg 0 = S(x)$;

Q3. $\forall x.x + 0 = x$;

Q4. $\forall x.\forall y.x + S(y) = S(x + y)$;

Q5. $\forall x.x \times 0 = 0$;

Q6. $\forall x.\forall y.x \times S(y) = (x \times y) + x$;

Q7. $\forall x.(\neg x = 0 \to \exists y.x = S(y))$.

PA is obtained from Q by adding the induction schema:

PA1. $\varphi^y_0 \wedge \forall x.(\varphi^y_x \to \varphi^y_{S(x)}) \to \forall x.\varphi^y_x$, where y is free in φ and x is substitutable for y in φ.

Observe that we are considering the signature of the logic as only having the universal quantifier, implication sign, and negation sign: whenever another connective appears, it should be written using only implication and negation signs; for instance $\varphi \wedge \psi := \neg(\varphi \to \neg\psi)$. Other options could have been made here.

3.2 The theory PA′

In this section, we develop a version of PA, namely PA′. To achieve that goal, we present some useful results.

Theorem 3.2.1. *The schema*

Inst. 1: $(\forall x.\varphi) \to \varphi^x_t$*, where t is substitutable for x in φ*

has the same instances as the two following schemata when considered together:

Inst. 2: $(\forall x.\varphi) \to \varphi^x_t$*, where t is substitutable for x in φ and x does not occur in t;*

Inst. 3: $(\forall x.\varphi^y_x) \to \varphi^y_t$*, where t is substitutable for y in φ, x does not occur free in φ, the variable y does not occur free under the scope of a $\forall x$ quantifier in φ, the variable y is not the variable x, and y does not occur in t.*

Proof. Let us analyze the following cases:

Inst. 2 $\Longrightarrow$ **Inst. 1:** This is immediate, since all instances of **Inst.** 2 are directly instances of **Inst. 1**.

Inst. 3 $\Longrightarrow$ **Inst. 1:** Suppose that one is given $\mu := (\forall x.\varphi^y_x) \to \varphi^y_t$, where t is substitutable for y in φ, x does not occur free in φ, the variable y does not occur free under the scope of a $\forall x$ quantifier in φ, the variable y is not the variable x, and y does not occur in t. Take $\xi := \varphi^y_x$. Consider the two following situations:

y **does not occur free in** φ**:** For this case, $\xi = \varphi^y_x = \varphi = \varphi^y_t$. As x does not occur free in φ, we conclude that $\xi^x_t = \xi$ and t is substitutable for x in φ, thus t is substitutable for x in ξ. Consequently,

$$\mu = ((\forall x.\varphi^y_x) \to \varphi^y_t) = ((\forall x.\xi) \to \xi) = ((\forall x.\xi) \to \xi^x_t),$$

so μ is an instance of **Inst. 1**.

y **occurs free in** φ**:** Suppose, aiming a contradiction, that t is *not* substitutable for x in $\xi = \varphi^y_x$. Then, x occurs free in ξ and there is a variable z in t which is captured by a quantifier $\forall z$ in ξ^x_t. As x do not occur free in φ by hypothesis, this means that there is a variable z in t which is captured by a quantifier $\forall z$ in φ^y_t; which contradicts the fact that t is substitutable for y in φ. So, t is substitutable for x in ξ. As the variable y does not occur free under the scope of a $\forall x$ quantifier and x does not occur free in φ, we conclude that $\xi^x_t = \left(\varphi^y_x\right)^x_t = \varphi^y_t$. Hence,

$$\mu = ((\forall x.\varphi^y_x) \to \varphi^y_t) = ((\forall x.\xi) \to \xi^x_t),$$

and so μ is an instance of **Inst. 1**.

Inst. 1 $\Longrightarrow$ **Inst. 2, Inst. 3:** Consider $\mu' := (\forall x.\varphi) \to \varphi^x_t$, where t is substitutable for x in φ. Consider the following cases:

x **does not occur in** t**:** In this case, μ' is an immediate instance of **Inst. 2**.

x **occurs in** t**:** Take $\chi := \varphi^x_y$, where y is a fresh variable not occurring in φ (not even in the quantifiers of φ) and in t. Clearly, the variable y does not occur free under the scope of a $\forall x$ quantifier in χ. As t is substitutable for x in φ and y does not occur in φ, it follows that t is substitutable for y in χ. Furthermore, x does not occur free in χ. It is clear that $\chi^y_x = \varphi$. As y does not appear in φ (not even in quantifiers of φ) and all free occurrences of x in φ are being replaced by y in χ, we have that $\chi^y_t = \left(\varphi^x_y\right)^y_t = \varphi^x_t$. Thus,

$$((\forall x.\chi^y_x) \to \chi^y_t) = ((\forall x.\varphi) \to \varphi^x_t) = \mu',$$

and so μ' is an instance of **Inst. 3**.

The result follows by the previous case analysis. ⊣

Theorem 3.2.2. *The following schemata have the same instances:*

Ind. 1: $\varphi^y_0 \wedge \forall x.(\varphi^y_x \to \varphi^y_{\mathsf{S}(x)}) \to \forall x.\varphi^y_x$*, where y is free in φ and x is substitutable for y in φ;*

Ind. 2: $\varphi^y_0 \wedge \forall x.(\varphi^y_x \to \varphi^y_{\mathsf{S}(x)}) \to \forall x.\varphi^y_x$*, where y is free in φ, the variable x is not the variable y, and x is substitutable for y in φ.*

Proof. We have the following cases to study:

Ind. 2 $\implies$ **Ind. 1:** It is immediate, since the instances of **Ind.** 2 are directly instances of **Ind.** 1.

Ind. 1 $\implies$ **Ind. 2:** Suppose that $\mu := (\varphi^y_0 \wedge \forall x.(\varphi^y_x \to \varphi^y_{\mathsf{S}(x)}) \to \forall x.\varphi^y_x)$ is an instance of **Ind. 1**. If x is not y, then it is immediately an instance of **Ind.** 2. So, suppose that x is y in μ. Thus, $\mu = (\varphi^x_0 \wedge \forall x.(\varphi \to \varphi^x_{\mathsf{S}(x)}) \to \forall x.\varphi)$. Take y a fresh variable not occurring in φ (not even in quantifiers) and $\psi := \varphi^x_y$. As x is free in φ and y does not appear at all in φ, it follows that y is free in ψ. As x is free in φ, we conclude that x is substitutable for y in ψ. Furthermore, as y does not occur at all in φ, $\psi^y_x = \left(\varphi^x_y\right)^y_x = \varphi$, $\psi^y_0 = \left(\varphi^x_y\right)^y_0 = \varphi^x_0$, and $\psi^y_{\mathsf{S}(x)} = \left(\varphi^x_y\right)^y_{\mathsf{S}(x)} = \varphi^x_{\mathsf{S}(x)}$. Therefore, we have

$$\mu = (\varphi^x_0 \wedge \forall x.(\varphi \to \varphi^x_{\mathsf{S}(x)}) \to \forall x.\varphi) = (\psi^y_0 \wedge \forall x.(\psi^y_x \to \psi^y_{\mathsf{S}(x)}) \to \forall x.\psi^y_x)$$

All the cases were considered. ⊣

Now we are ready to define PA′.

Definition 3.2.1. Let PA′ be PA from before where the universal instantiation axiom is replaced by the schemata **Inst.** 2 and **Inst. 3**, and where the induction axiom is replaced by **Ind.** 2.

Theorem 3.2.3. *The two following statements are equivalent:*

1. $\mathsf{PA} \vdash_{k\ steps} \varphi$;

2. $\mathsf{PA}' \vdash_{k\ steps} \varphi$.

Proof. From Theorems 3.2.1 and 3.2.2, we know that the axioms in PA′ that are a replacement of the axioms of PA have exactly the same instances. So, in a proof of PA an occurrence of **Inst. 1** can be replaced by an occurrence of **Inst.** 2 or **Inst. 3** to obtain a proof in PA′ exactly with the same formulas, in particular with the same length. The same idea applies to substitutions of **Ind. 1** in proofs of PA by **Ind.** 2 to obtain proofs in PA′. Furthermore, **Inst.** 2 and **Inst. 3** can be replaced by **Inst. 1** in the same fashion, and **Ind.** 2 by **Ind. 1**. ⊣

By the previous result, we know that the decidability of $\mathsf{PA} \vdash_{k\ \mathrm{steps}}$ reduces to the decidability of $\mathsf{PA}' \vdash_{k\ \mathrm{steps}}$. We will consider the axioms **Inst.** 2 and **Inst.** 3, and **Ind.** 2—they have the nice syntactical feature that in the replacements one cannot have a variable being substituted by a term where that very variable occurs. We are considering the number of steps as being the number of rules that are being applied—here one could also consider the number of proof lines, the results that we are going to present can be adapted for that situation.

3.3 Main results

We were inspired by Parikh systems (see, for instance, [31]) for the systems that we have developed, but we use very similar terminology to the one used in [31] with very different meanings (the reader should always have this in mind). The biggest difference between our approach and the approach followed in [31] is that we have developed a general way to obtain the provable formulas via schemata and in the latter the authors' focus in schemata occurs mainly in the axioms (they do not extend that notion to the provable formulas as we do). In that paper, it was developed a technique to study the decidability of k-provability for some theories using unification. We develop a new technique that depends on a different way to unify—we create a technique to unify some of the schemata that generate the provable formulas.

3.3.1 Provable schemata

The general idea of our approach is to attach a meaning to the combinatorial nature of general proof structures, namely to the different ways to combine, in a given number of steps, the axioms of the considered theory. Let us see, as an example, the general structure that corresponds to the following arrangement of the axioms: MP([L2],MP([L2],[L1])). This means that one firstly applies MP to an L1 implication using an L2 axiom, and to the result of that, which must be an implication, one applies MP using an axiom of the form L2. Starting from the first application, to apply to the left side of $(\varphi \to (\psi \to \mu)) \to ((\varphi \to \psi) \to (\varphi \to \mu))$ something of the form of $\varphi \to (\psi \to \varphi)$, one must have $\mu = \varphi$. Hence, the application of the first MP yields something of the form $(\varphi \to \psi) \to (\varphi \to \varphi)$. Now, to apply L2 to $(\varphi \to \psi) \to (\varphi \to \varphi)$, one needs $\psi = \xi \to \varphi$. In this conditions, the general shape/structure of MP([L2],MP([L2],[L1])) is $\varphi \to \varphi$.

Clearly, any other way to arrange the axioms in the considered shape is a particular case of $\varphi \to \varphi$. Moreover, $\varphi \to \varphi$ codifies, in a unique schema, all the ways to combine the axioms in MP([L2],MP([L2],[L1])). A similar analysis could be carried out for (some of) the other combinations of axioms, resulting in a finite list of schemata that generate, via substitutions, all instances of the other schemata that are obtainable in a given number of steps, k (this only works for some values of k due to undecidability issues). Hence, for some values of k, there are finitely many provable schemata that give rise to the formulas that are provable in k steps. It is important to observe that the previous idea does not work for all proof-skeletons (see Theorem 5.1 from [58], and Theorem 14.1 from [72, p.103]).

Now we move to formalize the previous ideas. Having in mind what was previously observed, the general shape of a schema is nothing but $F[\varphi_0,\ldots,\varphi_{n_0},t_0,\ldots,t_{n_1},v_0,\ldots,v_{n_2}]\&C$, where $\varphi_0,\ldots,\varphi_{n_0}$ stand for formula-variables, where $t_0,\ldots,t_{n_1}$ stand for term-variables, and where $v_0,\ldots,v_{n_2}$ stand for variable-variables; F stands for the arrangement of the logical symbols; and C stands for a condition on the variables, on the formulas, and on the terms. All the variables in the previous schema are exactly that, variables, they do not stand for actual entities; for instance, the formula-variables do not stand for actual formulas. Let us see two examples:

- '$\varphi \to (\psi \to \varphi)$' is a schema, where $F[\varphi_0, \varphi_1] := \varphi_0 \to (\varphi_1 \to \varphi_0)$ only has formula-variables, and where there are no further conditions;
- '$\varphi^x_0 \land \forall y.(\varphi^x_y \to \varphi^x_{\mathsf{S}(y)}) \to \forall y.\varphi^x_y$, where y is free in φ, the variable x is not the variable y, and x is substitutable for y in φ' is a schema where $F[\varphi, x] := \varphi^x_0 \land \forall y.(\varphi^x_y \to \varphi^x_{\mathsf{S}(y)}) \to \forall y.\varphi^x_y$, and $C :=$'y is free in φ, the variable x is not the variable y, and x is substitutable for y in φ'.

Let us fix throughout the rest of the chapter: $C_0(\varphi, x) :=$'x is free in φ', $C_1(\varphi, x) :=$'x is not free in φ', $C_2(\varphi, t, x) :=$'t is substitutable for x in φ', $C_3(t, x) :=$'x does not occur in t', and $C_4(x, y) :=$'the variable x is different from the variable y', $C_5(\varphi, x, y) :=$ 'the variable y does not occur free under the scope of a $\forall x$ quantifier in φ', $C_6(\varphi, x) :=$ 'x occurs in φ', $C_7(\varphi, x) :=$'x does not occur free in φ', and $C_8(\varphi, x, y) :=$'there is a free occurrence of the variable y in φ that does not occur under the scope of a $\forall x$ quantifier'. Observe that all the previous conditions are decidable. By C_0 we mean that all the occurrences of x in φ are free. We will assume that T is a (fixed) theory of arithmetic that extends the presented version of PA' by adding schemata that depend on formulas, terms, and variables, without having any conditions on the schemata. We now move to define formally what a proof-skeleton is.

Definition 3.3.1. We define inductively[2] the notion of *proof-skeleton*:

Basis case: A_i is a proof-skeleton if A_i is the number[3] of an axiom of T;

Induction step: If $\mathcal{S}_0$ and $\mathcal{S}_1$ are proof-skeletons, then $\mathrm{MP}(\mathcal{S}_0, \mathcal{S}_1)$ and $\mathrm{Gen}(\mathcal{S}_0)$ are proof-skeletons.

We say that a proof-skeleton $\mathcal{S}$ *has k steps* if it has k applications of rules (k might be 0).

Now we define the notion of a schema.

Definition 3.3.2. We define inductively the notion of *term-structure*:

Basis case: Every variable-variable, every term-variable, and 0 are a term-structure.

Induction step: If r and s are term-structures, then $\mathsf{S}(r)$, $r + s$, $r \times s$, and $s^{\vec{x}}_{\vec{t}}$, where $\vec{x}$ are variable-variables and $\vec{t}$ are term-structures, are also term-structures.

We define inductively the notion of *formula-structure*:

[2] Throughout our Thesis, we use the term 'induction' for the proof technique and to define objects like sets; we reserve the term 'recursion' for the definition of functions.

[3] Here by 'number' we mean the syntactical entity.

Basis case: Every formula-variable is a formula-structure. Furthermore, $r = s$ is a formula-structure, with r and s term-structures.

Induction step: If F and G are formula-structure, then the following are formula-structures:

- $F \to G$,
- $\neg F$,
- $\forall v.F$, where v is a variable-variable,
- $(F)^{\vec{v}}_{\vec{t}}$, where $\vec{v}$ are variable-variables, and $\vec{t}$ are term-structures.

We say that F is a *sub-formula-structure* of G if F is a formula-structure that occurs in G. We say that A is an *atom* if A is a formula-variable, or if $A = \varphi^{x_0 \dots x_\ell}_{t_0 \dots t_\ell}$, where φ is either a formula-variable or a formula-structure of the form $r = s$, with r and s term-structures. We say that an expression of the form

$$F[\varphi_0, \dots, \varphi_{n_0}, t_0, \dots, t_{n_1}, v_0, \dots, v_{n_2}] \& \bigvee_{i \in I} \mathop{\&}_{j \in J_i} \sim^{k^0_j} C_{k^1_j}(A_i, t_{k^2_j}, v_{k^3_j})$$

or of the form

$$F[\varphi_0, \dots, \varphi_{n_0}, t_0, \dots, t_{n_1}, v_0, \dots, v_{n_2}] \& \perp$$

is a *schema*, where I and J_i are sets of indices (possibly empty, in which case we omit the conditions), $F[\varphi_0, \dots, \varphi_{n_0}, t_0, \dots, t_{n_1}, v_0, \dots, v_{n_2}]$ is a formula-structure, and A_i are atoms. Here $C_{k^1_j}$ stand for a (syntactical) representation of the condition in the theory T previously mentioned (we also allow formula-structures to occur inside the conditions, but this will be avoided using several conventions). We allow term-structures to occur inside the conditions.

Every axiom of PA′ is a schema. Furthermore, every axiom of T is a schema.

Convention 3.3.1. In every occurrence of $\varphi^{\vec{x}}_{\vec{s}}$ or $t^{\vec{x}}_{\vec{s}}$ in schemata, we do not allow the variables that are being changed to occur in the replacing term. Furthermore, we do not allow a variable to occur in a replacement being mapped to different terms, and we do not allow repeated occurrences of the same change in the replacement (for instance $t^{x\,x}_{s\,s}$).

It is important to stress that schemata are syntactical objects, even the conditions in them are syntactical (that have a semantical interpretation). The symbols $\sim$, $\bigvee$, and $\&$ are syntactical representations of the connectives in the meta-language (negation, disjunction, and conjunction, respectively).

Convention 3.3.2. As $(\varphi \to \psi)^x_t = \varphi^x_t \to \psi^x_t$, $(\neg\varphi)^x_t = \neg\varphi^x_t$, and

$$(\forall y.\varphi)^x_t = \begin{cases} \forall y.\varphi, & x = y \\ \forall y.\varphi^x_t, & x \neq y, \end{cases}$$

hold for all formulas, we assume these identities for schemata. This means that in a schema one can move all occurrences of $(\cdot)^x_t$ inside the formula-structure.

With the conventions that we are going to present, we will extend the syntactical equality (and we will continue to denote it simply by '='). Sometimes to emphasize that $\mathscr{E}_0 = \mathscr{E}_1$ syntactically we will say that '$\mathscr{E}_0$ is $\mathscr{E}_1$' (we will also use it to express that $\mathscr{E}_0$ and $\mathscr{E}_1$ are syntactically the same after some suitable substitution).

Convention 3.3.3. We assume the following identities:

- $\varphi^{x_0 \cdots x_n}_{t_0 \cdots t_n} = \varphi^{x_{f(0)} \cdots x_{f(n)}}_{t_{f(0)} \cdots t_{f(n)}}$, where $f : \{0, \dots, n\} \to \{0, \dots, n\}$ is a bijection (this feature is not troublesome because we are interpreting the replacements as simultaneous replacements that satisfy Convention 3.3.1).
- The usual properties of replacements, for example:

$$(x + \mathsf{S}(s))^x_t = (t + \mathsf{S}(s^x_t)).$$

- The usual identities for propositional (meta-)logic, for example:
  - $\sim^{2n} C = C$, $\sim^{2n+1} C = \sim C$,
  - $C \bigvee C = C$,
  - $\sim (C \,\&\, C') = (\sim C) \bigvee (\sim C')$ (where C and C' are conditions), and so on.
- For C_0:
  - $C_0(\neg F, x) = C_0(F, x)$,
  - $C_0(F \to G, x) = C_0(F, x) \,\&\, C_0(G, x)$,
  - $C_0(\forall y.F, x) = C_0(F, x) \,\&\, C_4(x, y)$.
- $C_1(F, x) = \sim C_0(F, x)$.
- For C_2:
  - $C_2(\neg F, t, x) = C_2(F, t, x)$,
  - $C_2(F \to G, t, x) = C_2(F, t, x) \,\&\, C_2(G, t, x)$,
  - $C_2(\forall y.F, t, x) = C_7(\forall y.F, x) \bigvee (C_3(t, y) \,\&\, C_2(F, t, x))$.
- For C_5:
  - $C_5(\neg F, x, y) = C_5(F, x, y)$,
  - $C_5(F \to G, x) = C_5(F, x, y) \,\&\, C_5(G, x, y)$,
  - $C_5(\forall z.F, x, y) = (C_4(x, z) \,\&\, C_5(F, x, y)) \bigvee (\sim C_4(x, z) \,\&\, C_7(F, y))$.
- $C_7(\forall y.F, x) = \sim C_4(x, y) \bigvee C_7(F, x)$.

- For C_8:
  - $C_8(\neg F, x, y) = C_8(F, x, y)$,
  - $C_8(F \to G, x, y) = C_8(F, x, y) \bigvee C_8(G, x, y)$,
  - $C_8(\forall z.F, x, y) = (C_4(x, z) \,\&\, C_8(F, x, y))$.
- Similarly for the other conditions.

It is important to observe that in all the conditions one can arrange the formula-structures in such a way that inside the conditions one has only atoms (this will follow from the conventions that we are going to make, when considered together).

Convention 3.3.4. Whenever we are considering, at the same time, different schemata, we implicitly assume that they do not have common variables (this is just a useful technical feature that does not have any conceptual reason).

Definition 3.3.3. A *substitution* σ is a function that assigns: formula-variables to formula-structures, term-variables to term-structures, and variable-variables to variable-variables.

It is important to observe that if one applies a substitution σ to a schema one might be increasing the number of term-variables and variable-variables.

Definition 3.3.4. We define inductively the *provable schemata* by:

Basis case: Every schema that is an axiom is a provable schema;

Induction step: If

- $F[\varphi_0^0, \ldots, \varphi_{n_0}^0, t_0^0, \ldots, t_{n_1}^0, v_0^0, \ldots, v_{n_2}^0] \& \bigvee_{i \in I^0} \&_{j \in J_i^0} \sim^{k_j^{0,0}} C_{k_j^{0,1}}(A_i^0, t_{k_j^{0,2}}, v_{k_j^{0,3}})$, and
- $G[\varphi_0^1, \ldots, \varphi_{n_3}^1, t_0^1, \ldots, t_{n_4}^1, v_0^1, \ldots, v_{n_5}^1] \to H[\varphi_0^2, \ldots, \varphi_{n_6}^2, t_0^2, \ldots, t_{n_7}^2, v_0^2, \ldots, v_{n_8}^2]$ $\& \bigvee_{i \in I^1} \&_{j \in J_i^1} \sim^{k_j^{1,0}} C_{k_j^{1,1}}(A_i^1, t_{k_j^{1,2}}, v_{k_j^{1,3}})$

are provable schemata and there is σ such that

$$F[\sigma(\varphi_0^0), \ldots, \sigma(\varphi_{n_0}^0), \sigma(t_0^0), \ldots, \sigma(t_{n_1}^0), \sigma(v_0^0), \ldots, \sigma(v_{n_2}^0)] = G[\sigma(\varphi_0^1), \ldots, \sigma(\varphi_{n_3}^1), \sigma(t_0^1), \ldots, (t_{n_4}^1), \sigma(v_0^1), \ldots, \sigma(v_{n_5}^1)],$$

one says that

$$H[\sigma(\varphi_0^2), \ldots, \sigma(\varphi_{n_6}^2), \sigma(t_0^2), \ldots, \sigma(t_{n_7}^2), \sigma(v_0^2), \ldots, \sigma(v_{n_8}^2)] \& \bigvee_{i \in I^0} \&_{j \in J_i^0} \sim^{k_j^{0,0}} C_{k_j^{0,1}}(\sigma(A_i^0), \sigma(t_{k_j^{0,2}}), \sigma(v_{k_j^{0,3}})) \& \bigvee_{i \in I^1} \&_{j \in J_i^1} \sim^{k_j^{1,0}} C_{k_j^{1,1}}(\sigma(A_i^1), \sigma(t_{k_j^{1,2}}), \sigma(v_{k_j^{1,3}})) \; (\& C)$$

is a provable schema; furthermore,

$$\forall v.F[\varphi_0^0,\ldots,\varphi_{n_0}^0,t_0^0,\ldots,t_{n_1}^0,v_0^0,\ldots,v_{n_2}^0]\&\bigvee_{i\in I^0}\mathop{\&}_{j\in J_i^0}\sim^{k_j^{0,0}}C_{k_j^{0,1}}(A_i^0,t_{k_j^{0,2}},v_{k_j^{0,3}})$$
$$(\& C)$$

is also a provable schema. C is a possibly added condition that arises from conventions, for instance from Convention 3.3.2 by adding C_4 conditions.

A provable schema S is *provable in k steps* if in the construction of S as a provable schema were used, at most, k steps (we do not count the application of conventions as steps nor the axiom case as a step). A provable schema that is an axiom has skeleton equal to the number of the axiom; if a provable schema S has skeleton $\mathcal{S}$, then the corresponding provable schema obtained using the universal rule, $\forall v.S$, has skeleton Gen($\mathcal{S}$); if S_0 and S_1 are provable schemata that have skeletons $\mathcal{S}_0$ and $\mathcal{S}_1$, respectively, and S is a schemata obtained using the MP construction from S_0 and S_1, then S has skeleton MP($\mathcal{S}_0,\mathcal{S}_1$).

Convention 3.3.5. The equality in the previous definition should be read as follows: there is a substitution σ such that, after applying the conventions to both formula-structures considered in the definition, one gets syntactical equality. For each way of applying the conventions and for a fixed substitution one might get new provable schemata (for each way one gets a new provable schema). Thus, a schema is provable if there are a substitution and several applications of the conventions that make the conditions of the definition (in particular the equality) hold. In practice, Convention 3.3.2 will be applied in the following way: in a schema, either one has the same variable occurring in a quantifier and in a replacement, and then one eliminates the replacement; or one proceeds as the convention suggests and one adds a condition C_4 to differentiate the variables. This is assumed, for instance, in the provable schemata by considering both situations after the application of σ (this information then goes to the condition C). We will make the same assumption for the other conventions that we are going to establish. Whenever we consider a schema in the previous conditions, we are, in fact, considering all the schemata that are obtained using the previous procedure. In practice, we also allow that in the previous definition more conditions are added. Furthermore, we will use the notion of the previous definition where σ represents the application of several substitutions and conventions.

We now pause to give some examples. Clearly, $(\varphi\to(\psi\to\mu))\to((\varphi\to\psi)\to(\varphi\to\mu))$ is a provable schema, since it is an axiom. We also have that $(\varphi\to\psi)\to(\varphi\to\varphi)$ is a provable schema, since it can be obtained from the schemata $(\varphi\to(\psi\to\mu))\to((\varphi\to\psi)\to(\varphi\to\mu))$ and $\varphi\to(\psi\to\varphi)$ (by considering the substitution such that $\sigma(\varphi):=\varphi$, $\sigma(\psi):=\psi$, and $\sigma(\mu):=\varphi$). Moreover, for the same reason, $(\varphi\to\varphi)\to(\varphi\to\varphi)$ (by considering the substitution such that $\sigma(\varphi):=\varphi$, $\sigma(\psi):=\varphi$, and $\sigma(\mu):=\varphi$) is a provable schema. It is a good exercise to check that $\varphi\to\varphi$ is a provable schema.

Definition 3.3.5. We say that Σ is a *concrete-substitution* for the schema

$$F[\varphi_0,\ldots,\varphi_{n_0},t_0,\ldots,t_{n_1},v_0,\ldots,v_{n_2}]\&\bigvee_{i\in I}\mathop{\&}_{j\in J_i}\sim^{k_j^0}C_{k_j^1}(A_i,t_{k_j^2},v_{k_j^3})$$

if Σ assigns formula-variables to actual formulas, term-variables to actual terms, and variable-variables to actual variables, in such a way that

$$\bigvee_{i\in I}\mathop{\&}_{j\in J_i}\sim^{k_j^0}C_{k_j^1}(\Sigma(A_i),\Sigma(t_{k_j^2}),\Sigma(v_{k_j^3}))$$

is a true condition and such that, in any occurrence of $G^{y_0\cdots y_n}_{s_0\cdots s_n}$ in F (with G a term-structure or a formula-structure), no $\Sigma(y_i)$ occurs in any $\Sigma(s_j)$, and there are no $\Sigma(y_i)$ with different attributions. It also needs to be the case that in every occurrence of $G^{y_0\cdots y_n}_{s_0\cdots s_n}$ in the schema, each $\Sigma(s_i)$ is substitutable for $\Sigma(y_i)$ in $\Sigma(G)$ and that no $\Sigma(y_i)$ is $\Sigma(y_j)$ with $i\neq j$ (there are no repetitions of the same change). For a schema S, we use the notation $\Sigma(S)$ to denote the result of performing the substitution Σ in the schema S whenever the substitution satisfies the definition.

To respect the previous definition, when one applies Convention 3.3.2, one should add the condition $C_7(\forall y.\varphi,x)\bigvee C_3(t,y)$ in the context of provable schemata.

Convention 3.3.6. We assume that $\left(t^{\vec{x}}_{\vec{s}}\right)^{y}_{r}=t^{\vec{x}\ \ y}_{\vec{s}\ \ r}$, and the analogue identity for formula-structures, if there are no other occurrences of the basis of the replacement, t, in the schema that is being considered (here we are assuming that no $\vec{x}$ is a y, the suitable changes should be applied for the other case and the respective conditions should be added in the presence of provable schemata). This offers no problem with the concrete-substitution interpretation because if one has a concrete-substitution that satisfies the left-hand-side of the equation, one can construct a concrete-substitution that satisfies the right-hand-side and vice-versa. For example, if we had $(t^x_s)^y_r$ with $\Sigma(x):=x_0$, $\Sigma(s):=\mathsf{S}(x_1)$, $\Sigma(y):=x_1$, $\Sigma(r):=x_2$, then by considering a concrete-substitution Σ' such that $\Sigma'(t):=\Sigma(t)$, $\Sigma'(x):=x_0$, $\Sigma'(y):=x_1$, $\Sigma'(s):=S(x_2)$, and $\Sigma'(r):=x_2$, we would get

$$\begin{aligned}\Sigma\left((t^x_s)^y_r\right)&=\left(\Sigma(t)^{\Sigma(x)}_{\Sigma(s)}\right)^{\Sigma(y)}_{\Sigma(r)}=\left(\Sigma(t)^{x_0}_{S(x_1)}\right)^{x_1}_{x_2}=\Sigma(t)^{x_0\ \ x_1}_{S(x_2)\ x_2}\\&=\Sigma'(t)^{x_0\ \ x_1}_{\mathsf{S}(x_2)\ x_2}=\Sigma'(t)^{\Sigma'(x)\ \Sigma'(y)}_{\Sigma'(s)\ \Sigma'(r)}=\Sigma'(t^{x\ y}_{s\ r}).\end{aligned}$$

This means that if there is a concrete-substitution of one member of the equality, then there is a concrete-substitution for the other member. This identity will hold only if none of the $\vec{x}$ occurs in r (one should add the suitable condition to express this fact).

Observe that one could also have in a schema, besides the considered term-structure, the term-structures $(t^x_z)^z_s$ and t^y_r. In that type of situations, we take x' a totally fresh variable (i.e. not occurring at all) and the first term-structure is replaced by $t^{x'\ z}_{s\ \ s}$ and the second is replaced by $t^{y\ x'}_{r\ x}$ (the same for formula-structures), assuming that z is not x (one should add the suitable conditions for that).

More generally, $\left(t^{\vec{x}}_{\vec{r}}\right)^{\vec{y}}_{\vec{s}}$ is replaced by $(t')^{\vec{x'}\ \vec{y}}_{\vec{r}^{\vec{y}}_{\vec{s}}\ \vec{s}}$, with $\vec{x'}$ all fresh and t', and in the other occurrences of t where $\vec{x}$ are not being changed (one should do a case analysis for this using conditions C_4 and one should add the fact that, for the new t, x does not occur in t), one places $(t')^{\vec{x'}}_{\vec{x}}$ and one proceeds in a similar fashion for the other cases; in the previous situation we need to assume that none $\vec{x}$ occurs in $\vec{y}$, one should also consider the case where some of the $\vec{x}$ are $\vec{y}$ and add the suitable conditions in the presence of provable schemata, which simplifies the analysis.

We assume this convention also for formula-structures. More precisely, $\left(\varphi^{\vec{x}}_{\vec{r}}\right)^{\vec{y}}_{\vec{s}}$ is replaced by $(\varphi')^{\vec{x'}\ \vec{y}}_{\vec{r}^{\vec{y}}_{\vec{s}}\ \vec{s}}$, with $\vec{x'}$ all fresh and φ' fresh, and in the other occurrences of φ where $\vec{x}$ are not being changed, one places $(\varphi')^{\vec{x'}}_{\vec{x}}$; for the previous replacement to work we need, in the context of provable schemata, to add to the conditions: $\vec{x'}$ is free in φ, $\vec{x}$ is substitutable for $\vec{x'}$ in φ, $\vec{r}$ is substitutable for $\vec{x'}$ in φ, and $\vec{x}$ is not free in φ (the justification for all this procedure is that one needs these conditions for the previous reasoning for terms to be applied for formulas, namely to be able, in the presence of a concrete-substitutions, to go back from the image of the replaced formula-structure to the image of the initial one). This means that, without loss of generality, we will assume that in the schemata all these reductions were already applied—in practice, this means that several case analysis ought to be done. As our interpretation of the schemata is obtained via concrete-substitutions, we do not have a problem, since everything fits the definition.

Convention 3.3.7. In every occurrence of x^y_t in a provable schemata, one should consider the cases where x is the same as y, that entails $x^y_t = t$, and the opposite case, that entails $x^y_t = x$. These situations will give rise to several provable schemata that are originated from a single syntactical expression. For each situation, we should add accordingly the conditions C_4 or $\sim C_4$ (c.f. Convention 3.3.5).

Lemma 3.3.1. *Consider formulas φ and ψ, and schemata*

- $S_0 := F[\varphi^0_0,\ldots,\varphi^0_{n_0},t^0_0,\ldots,t^0_{n_1},v^0_0,\ldots,v^0_{n_2}]\& \bigvee_{i\in I^0} \&_{j\in J^0_i} \sim^{k^{0,0}_j} C_{k^{0,1}_j}(A^0_i, t_{k^{0,2}_j}, v_{k^{0,3}_j})$,
- $S_1 := G[\varphi^1_0,\ldots,\varphi^1_{n_3},t^1_0,\ldots,t^1_{n_4},v^1_0,\ldots,v^1_{n_5}] \to H[\varphi^2_0,\ldots,\varphi^2_{n_6},t^2_0,\ldots,t^2_{n_7},v^2_0,\ldots,v^2_{n_8}]\& \bigvee_{i\in I^1} \&_{j\in J^1_i} \sim^{k^{1,0}_j} C_{k^{1,1}_j}(A^1_i, t_{k^{1,2}_j}, v_{k^{1,3}_j})$.

If there are concrete-substitutions Σ_0 and Σ_1 such that $\Sigma_0(S_0) = \varphi$ and $\Sigma_1(S_1) = \varphi \to \psi$, then there are a substitution σ such that

$$F[\sigma(\varphi^0_0),\ldots,\sigma(\varphi^0_{n_0}),\sigma(t^0_0),\ldots,\sigma(t^0_{n_1}),\sigma(v^0_0),\ldots,\sigma(v^0_{n_2})] = G[\sigma(\varphi^1_0),\ldots,\sigma(\varphi^1_{n_3}),\sigma(t^1_0),\ldots,\sigma(t^1_{n_4}),\sigma(v^1_0),\ldots,\sigma(v^1_{n_5})]$$

and a concrete-substitution Σ such that $\Sigma(S_2) = \psi$, where

$$S_2 = H[\sigma(\varphi^2_0),\dots,\sigma(\varphi^2_{n_6}),\sigma(t^2_0),\dots,\sigma(t^2_{n_7}),\sigma(v^2_0),\dots,\sigma(v^2_{n_8})]\&$$
$$\bigvee_{i\in I^0}\mathop{\&}_{j\in J^0_i}\sim^{k^{0,0}_j} C_{k^{0,1}_j}(\sigma(A^0_i),\sigma(t_{k^{0,2}_j}),\sigma(v_{k^{0,3}_j}))\&$$
$$\bigvee_{i\in I^1}\mathop{\&}_{j\in J^1_i}\sim^{k^{1,0}_j} C_{k^{1,1}_j}(\sigma(A^1_i),\sigma(t_{k^{1,2}_j}),\sigma(v_{k^{1,3}_j}))\left(\mathop{\&} C\right).$$

Proof. By hypothesis, $\Sigma_0(S_0) = \varphi$ and $\Sigma_1(S_1) = \varphi \to \psi$. This means that there are concrete (and not variable) formulas $\varphi^0_0,\dots,\varphi^0_{n_0},\varphi^1_0,\dots,\varphi^1_{n_2}$, concrete terms $t^0_0,\dots,t^0_{n_1},t^1_0,\dots,t^1_{n_3}$, and concrete variables $v^0_0,\dots,v^0_{n_2},v^1_0,\dots,v^1_{n_5}$ obeying the semantical translation of the syntactical conditions such that

$$\begin{aligned}\varphi =&F[\varphi^0_0,\dots,\varphi^0_{n_0},t^0_0,\dots,t^0_{n_1},v^0_0,\dots,v^0_{n_2}]\\ =&G[\varphi^1_0,\dots,\varphi^1_{n_2},t^1_0,\dots,t^1_{n_3},v^1_0,\dots,v^1_{n_5}].\end{aligned}$$

Hence, the outer layout of implication signs, negation signs, universal quantifier signs, and parenthesis in S_0 and in S_1 can be made the same. For instance, the formula-structures $\varphi \to (\psi \to \varphi)$ and $\varphi \to ((\varphi \to \varphi) \to \varphi)$ can have the same outer layout of signs, but $\varphi \to \varphi$ and $(\varphi \to \varphi) \to \varphi$ cannot—we use this expression to say that they have a common outer structure of parenthesis, implication signs, negation signs, and universal quantifier sings (we mention it as simply the *layout*). A layout is nothing but a sequence of symbols: parenthesis, implication signs, negation sings, and universal quantifier signs. For example, $(\forall()) \to (\neg() \to ())$ is a layout. We will say that a layout L is a *sub-layout* of L' if L is a subsequence of L'. We call *entry* to the content of a layout inside implications such that they do not contain further implications.

It is not hard to see that there are substitution σ_0 and σ_1, and concrete-substitutions Σ'_0 and Σ'_1, such that the layout of $\sigma_0(F)$ and $\sigma_1(G)$ is the same as the layout of φ and $\Sigma'_0(\sigma_0(S_0)) = \varphi$ and $\Sigma'_1(\sigma_0(S_1)) = \varphi \to \psi$. It is enough for the procedure that we are going to describe that the layout is the same with possible exception of the entries in it because the entries are going to be accounted in the procedure *per se*. For example, consider the layouts $(\neg\forall(\neg())) \to (() \to (\neg()))$ and $() \to (() \to (\neg()))$; although they are not the same, for the purposes of the procedure that we are going to develop the differences do not matter because they occur only at the level of the entries (this will be accounted for in the procedure); furthermore, when we are considering entries we always consider the biggest one, for instance, in the first layout, although $()$, $\neg()$, $\forall(\neg())$, and $\neg\forall(\neg())$ are all entries of the layout that correspond to the same position, we will consider the biggest one, i.e. $\neg\forall(\neg())$. In all, without loss of generality, we may assume that F and G have the same layout. We assume that the changes of Convention 3.3.6 were already made in such a way that one does not have compositions of replacements (from the information of the concrete-substitutions one can find the correct way to apply the convention). Using the concrete-substitutions, one can find the correct way to apply the Convention 3.3.2 and add that information to a condition C, composed of several C_4 conditions. After this, one can still find concrete-substitutions such that $\Sigma_0'(\sigma_0(S_0)) = \varphi$ and $\Sigma_1'(\sigma_0(S_1)) = \varphi \rightarrow \psi$.

We do the following to construct a substitution σ and a concrete-substitution Σ (we are implicitly considering Convention 3.3.4 and all the other conventions):

1. If x is a variable-variable that is mapped to $\Sigma_i(x)$, then x is assigned to a new fresh-variable, $\sigma(x)$, and $\Sigma(\sigma(x)) := \Sigma_i(x)$. At this stage, one should also identify the variables that are mapped to the same concrete variables and distinguish the variables that are mapped to different variables using conditions C_4 and $\sim C_4$.

2. One proceeds in a similar way with the term-variables.

3. Starting from the first entry of the (common) layout, if the content of any of the entries is a formula-variable, then one assigns that formula-variable to the content of the other entry and moves to the next entry (in the end we might need to make some adjustments to this step).

4. Suppose now that the contents of both entries are not formula-variables.

   4.1. Suppose that the contents are $X_0^0 \cdots X_{\ell_0}^0 (\varphi_i)_{t_0^0 \cdots t_n^0}^{v_0^0 \cdots v_n^0}$, in S_0, and $X_0^1 \cdots X_{\ell_1}^1 (\varphi_k)_{t_0^1 \cdots t_m^1}^{v_0^1 \cdots v_m^1}$, in S_1, where φ_i and φ_j are formula-variables (here $X_0^i \cdots X_{\ell_i}^i$ are arrays of quantifiers and negation symbols).

      4.1.1. It might be needed to make a substitution in order to $\ell_0 = \ell_1$ (for the case where one has all the layout equal with possible exception of the entries in it). If $\ell_0 < \ell_1$, this is achieved by $\sigma(\varphi_i) := X_{\ell_0+1}^1 \cdots X_{\ell_1}^1 \varphi'$, with φ' fresh (it is possible that it is necessary to make adjustments to the variables occurring in the quantifiers so that none of them is $v_0^0 \cdots v_n^0$ and to add the suitable conditions). Using σ, one unifies the variables occurring in the same place in the quantifiers (their value is already fixed by stage 1).

      4.1.2. One then assigns for each component in the entry the corresponding image through the concrete-substitution, with the difference that the occurrences of actual variables in the formulas are replaced by occurrences of variable-variables. One does the same for the variables and term-structures in the replacements.

      4.1.3. Define the concrete-substitution accordingly.

      4.1.4. Move to the next entry.

   4.2. Suppose that the contents are $X_0^0 \cdots X_{\ell_0}^0 (r_0 = s_0)_{t_0^0 \cdots t_n^0}^{v_0^0 \cdots v_n^0}$, in S_0, and $X_0^1 \cdots X_{\ell_1}^1 (r_1 = s_1)_{t_0^1 \cdots t_m^1}^{v_0^1 \cdots v_m^1}$, in S_1, where r_0, r_1, s_0, s_1 are term-structures.

      4.2.1. One proceeds as in 4.1 for the quantifier variables.

4.2.2. We can perform substitutions to both entries in such a way that they become exactly equal to the content of the actual formula, but with the difference that actual variables in the formula are represented by variable-variables (the same idea that was applied before). For example, if the contents are $x + t = y$ and $t' = t''$, and they are both mapped using the concrete-substitutions to $x+(z+\mathsf{S}(0)) = 0$, then one considers $\sigma(t) := z+\mathsf{S}(0)$, with z a variable-variable, $\sigma(t') := x + (z + \mathsf{S}(0))$, and $\sigma(t'') := 0$.

4.2.3. One does the suitable adaptions for the replacements. For instance, if the content of an entry is $t_s^x = \mathsf{S}(x) + 0$ and the concrete-substitution is defined for the entry by $\Sigma_i(t) := \Sigma(x)\times(0+x_0)$ and $\Sigma_i(s) := x_3 \times \mathsf{S}(0)$; then one defines for this entry $\sigma(t) := \sigma(x) \times (0 + y)$, and $\sigma(s) := z \times \mathsf{S}(0)$, entailing that

$$\begin{aligned}\sigma(t_s^x = \mathsf{S}(x) + 0) &= \left(\sigma(t)_{\sigma(s)}^{\sigma(x)} = \mathsf{S}(\sigma(x)) + 0\right) \\ &= ((z \times \mathsf{S}(0)) \times (0 + y) = \mathsf{S}(\sigma(x)) + 0).\end{aligned}$$

Observe that one might need to add several C_4 conditions.

4.2.4. Define accordingly the concrete-substitution for this situation by assigning the variable-variables to the corresponding concrete variables.

4.2.5. One can perform substitutions in such a way that the term-structures that appear are equal to their image through the concrete substitution where the occurrence of variables in the actual formula is replaced by an occurrence of variable-variables in the term-structures, as described before (all this can be done because we have the guarantee of the existence of the concrete-substitutions).

4.2.6. Define the concrete-substitution for this case by assigning the variable-variables to the actual variables that they represent accordingly.

4.2.7. With the previous construction, in particular we have

$$\sigma((r_0 = s_0)_{t_0^0 \cdots t_n^0}^{v_0^0 \cdots v_n^0}) = \sigma((r_1 = s_1)_{t_0^1 \cdots t_m^1}^{v_0^1 \cdots v_m^1}).$$

4.2.8. If any variable-variable or term-variable is already assigned, make the suitable adaptations (this is always possible and reduces to a case analysis).

4.2.9. Move to the next entry.

4.3. If the content of one entry is $X_0^0 \cdots X_{\ell_0}^0 (r_0 = s_0)_{t_0^0 \cdots t_n^0}^{v_0^0 \cdots v_n^0}$ and the content of the other entry is $X_0^1 \cdots X_{\ell_1}^1 (\varphi_k)_{t_0^1 \cdots t_m^1}^{v_0^1 \cdots v_m^1}$, one proceeds in a similar way by attributing, via σ, the formula-variable φ_k in such a way that in the end of the substitution one is left with a version of the actual formula where variables are replaced by variable-variables.

4.4. Move to the next entry.

5. Make the suitable changes to the variables in order to have

$$\Sigma(H[\sigma(\varphi_0^2),\ldots,\sigma(\varphi_{n_6}^2),\sigma(t_0^2),\ldots,\sigma(t_{n_7}^2),\sigma(v_0^2),\ldots,\sigma(v_{n_8}^2)]) = \psi.$$

It is not hard to see that the constructed concrete-substitution obeys the conditions of the considered schemata and the added ones (while applying the conventions in a suitable way). ⊣

We could have presented a shorter proof of the previous results, but we decided to exhibit this one because it contains important ideas that are going to be used in several contexts. We can now prove that, using concrete-substitutions, k-provability of formulas is, in a sense, the same as k-provability of schemata.

Theorem 3.3.1. *The two following statements are equivalent:*

C1: $T \vdash_{k \text{ steps}} \varphi$;

C2: *There are S a provable schema in k steps and a concrete-substitution Σ such that $\Sigma(S) = \varphi$.*

Proof. By definition of provable schema in k steps, we have that **C2** implies **C1** (this can be more formally proved by induction on k).

Let us prove that **C1** implies **C2** by induction on k, the number of steps. Clearly, if φ is an axiom, then there are an axiom schema S and a concrete-substitution Σ such that $\Sigma(S) = \varphi$. Suppose, by induction hypothesis, that the result holds for k. Furthermore, assume that $T \vdash_{k+1 \text{ steps}} \varphi$. In the last steps of a proof of φ we either apply the rule MP or the rule Gen:

MP In this case, there is a formula ψ such that $T \vdash_{k_0 \text{ steps}} \psi \to \varphi$ and $T \vdash_{k_1 \text{ steps}} \psi$, with $k = k_0 + k_1$. By induction hypothesis, there are schemata satisfying the conditions of Lemma 3.3.1. By the lemma, it follows that there is a provable schema H and a concrete-substitution Σ such that $\Sigma(H) = \varphi$. As each of the schemata used in the lemma is, by induction hypothesis, provable in k_0 and k_1 steps (respectively), it follows that H is a schema provable in $k+1$ steps.

Gen In this case, φ is of the form $\forall x.\psi$. By hypothesis, $T \vdash_{k \text{ steps}} \psi$. So, by induction hypothesis, there are a provable schema in k steps, H, and a concrete-substitution Σ such that $\Sigma(H) = \psi$. Consider the provable schema in $k+1$ steps obtained from H via the universal schema from Definition 3.3.4, let us call it H'. Clearly, one can make the suitable changes in such a way that $\Sigma(H') = \forall x.\psi = \varphi$.

The result follows by induction. ⊣

Lemma 3.3.2. *Consider a formula φ, and schemata S_0 and S_1. Suppose that there are a concrete-substitution Σ and a substitution σ (that might include the application of several substitutions and conventions) such that $\Sigma(S_0) = \varphi$ and $\sigma(S_1) = S_0$. Then, there is a concrete-substitution Σ' such that $\Sigma'(S_1) = \varphi$.*

Proof. It is not hard to see that this follows from the considered definitions and from the fact that all the conventions are compatible with the concrete-substitution interpretation. ⊣

3.3.2 Decidability of schemata

The next result has a similar content to Proposition 2.2 from [31].

Lemma 3.3.3. *Given a condition of the form* $\&_{i\in I} \sim^{k_i} C_{k^1_j}(A_i, t_i, v_i)$*, one can computationally decide if there is a concrete substitution* Σ *such that*

$$\Sigma\left(\underset{i\in I}{\&} \sim^{k_i} C_{k^1_j}(A_i, t_i, v_i)\right) = \underset{i\in I}{\&} \sim^{k_i} C_{k^1_j}(\Sigma(A_i), \Sigma(t_i), \Sigma(v_i))$$

is true; furthermore, witnesses can be found in affirmative case. The result still holds if the image of certain meta-variables are a priori *fixed under a concrete-substitution.*

Proof. The following idea is a procedure for the case where all the atoms that are not fixed are formula-variables and where the term-variables that are not fixed occur without replacements:

1. Start by considering enough variables to satisfy the occurrences of the conditions C_4 and $\sim C_4$.

2. Then, consider every formula as being equal to $0 = 0$ and every term as being equal to 0.

3. After that, focus on C_0 conditions. For every occurrence of $C_0(\varphi, x)$, make the attribution $\varphi := (\varphi \wedge x = x)$ (if x is not yet free in φ). Proceed in a similar way to $\sim C_7$ (for this condition we just need one free occurrence). For the occurrences of $C_1(\varphi, x)$ make $\varphi := (\varphi \wedge \forall x.x = x)$, in particular if $\sim C_7(\varphi, x)$ and also $C_1(\varphi, x)$, then attribute instead $\varphi := (\varphi \wedge x = x) \wedge (\forall x.x = x)$.

4. Similarly for $\sim C_3$, for every occurrence of $\sim C_3(t, x)$, consider $t := (t + x)$.

5. Each time $\sim C_2(\varphi, t, x)$ occurs, (one should always compare what one is doing here with the C_5 and C_2 conditions for the same formula-variables and variable-variables) consider y a fresh variable, and attribute $t := (t + y)$ and $\varphi := (\varphi \wedge \forall y.x = 0)$.

6. For each occurrence of $\sim C_5(\varphi, x, y)$, take $\varphi := (\varphi \wedge \forall x.y = 0)$. For $C_8(\varphi, x, y)$ we might need to consider $\varphi := (\varphi \wedge y = 0)$, in particular if $\sim C_5(\varphi, x, y)$ and $C_8(\varphi, x, y)$ occur, then attribute $\varphi := ((\varphi \wedge \forall x.y = 0) \wedge y = 0)$. If $\sim C_8(\varphi, x, y)$ occurs, one should consider two cases:

   6.1. y does not occur free in φ, i.e. $C_7(\varphi, y)$;

6.2. Or $\sim C_7(\varphi, y)$ and in all free occurrences of y in φ they occur under the scope of a $\forall x$ quantifier. For this situation one acts in a similar way to the one described for $\sim C_5(\varphi, x, y)$ in all free occurrences of y (one places $\forall x$ in all free occurrences of y in φ).

7. Now test, for the considered attributions, if the occurrences of C_0, C_1, C_2, C_3, C_5, $\sim C_6$, C_7, C_8, and $\sim C_8$ are satisfied. In negative case (one should consider all possible situations), reject.

8. If $C_6(\varphi, x)$ is in the expression, test if the condition is already satisfied for the considered attributions. If not, then take $\varphi := (\varphi \wedge \forall x.x = 0)$ and test the conditions again.

If any variable, term, or formula is already fixed, the previous analysis remain valid (some adaptations are needed, for instance in the beginning of the algorithm)—this simplifies the algorithm to a case analysis. Furthermore, it is not hard to adapt it for replacements and for more complex term-structures (see the identities bellow). Suppose now that we have atoms of the form $(r = s)^{\vec{x}}_{\vec{t}}$. Then, we do the following:

1. We start by listing all the possible ways to apply the $\vec{x}$ to $r = s$, i.e. all the ways to apply the replacements (including the way in which the variables $\vec{x}$ do not occur in $r = s$). Each possibility will give rise to a separate analysis.

2. We proceed as in step 1 until no further replacements are applicable to term-structures.

3. In the previous step, one is left with several occurrences of $t^{\vec{y}}_{\vec{t'}}$.

4. One proceeds in a way similar to the previous algorithm (if possible, i.e. if no contradiction was reached). Observe that such an analysis is simplified, since, for instance, being free in the considered formula reduces to occurring in the formula (because there are no quantifiers).

One can also adapt accordingly the idea of the initial procedure. One should also have in mind the following identities concerning t^x_s:

- For C_2, with z a totally fresh variable,

$$C_2(\varphi, t^y_s, x) = \big(C_3(t, y) \mathbin{\&} C_2(\varphi, t, x)\big) \bigvee \big(\sim C_3(t, y) \mathbin{\&} C_2(\varphi, s, x) \mathbin{\&} C_2(\varphi, t^y_z, x)\big).$$

- For $\sim C_3$,

$$\sim C_3(t^y_s, x) = \big(C_3(t, y) \mathbin{\&} \sim C_3(t, x)\big) \bigvee \big(\sim C_3(t, y) \mathbin{\&} C_4(x, y) \mathbin{\&} \big(\sim C_3(t, x) \bigvee \sim C_3(s, x)\big)\big).$$

The case where we have atoms of the form $\varphi^{\vec{x}}_{\vec{t}}$ is a particular case of the previous analysis when one has in mind the following identities (that can be extended for more complex replacements):

- For C_0,

$$C_0(\varphi^y_t, x) = \big(C_7(\varphi,y) \,\&\, C_0(\varphi,x)\big) \bigvee \Big(\sim C_7(\varphi,y) \,\&\, C_2(\varphi,t,y) \,\&\, \Big(\big(C_4(x,y) \,\&\, C_0(\varphi,x) \,\&\, C_3(t,x)\big) \bigvee \big(C_4(x,y) \,\&\, C_0(\varphi,x) \,\&\, \sim C_3(t,x) \,\&\, C_5(\varphi,x,y)\big)\Big)\Big).$$

- For C_2,

$$C_2(\varphi^y_s, t, x) = \big(C_7(\varphi,y) \,\&\, C_2(\varphi,t,x)\big) \bigvee \Big(\sim C_7(\varphi,y) \,\&\, C_2(\varphi,s,y) \,\&\, \Big(\sim C_4(x,y) \bigvee \big(C_4(x,y) \,\&\, C_3(s,x) \,\&\, C_2(\varphi,t,x)\big) \bigvee \big(C_4(x,y) \,\&\, \sim C_3(s,x) \,\&\, C_2(\varphi,t,x) \,\&\, C_2(\varphi,t,y)\big)\Big)\Big).$$

- For C_5,

$$C_5(\varphi^x_s, z, y) = \big(C_7(\varphi,x) \,\&\, C_5(\varphi,z,y)\big) \bigvee \Big(\sim C_7(\varphi,x) \,\&\, C_2(\varphi,s,x) \,\&\, \Big(\sim C_4(x,y) \bigvee \big(C_4(x,y) \,\&\, C_3(s,y) \,\&\, C_5(\varphi,z,y)\big) \bigvee \big(C_4(x,y) \,\&\, \sim C_3(s,y) \,\&\, C_5(\varphi,z,y) \,\&\, C_5(\varphi,z,x)\big)\Big)\Big).$$

- For $\sim C_7$,

$$\sim C_7(\varphi^y_t, x) = \big(C_7(\varphi,y) \,\&\, \sim C_7(\varphi,x)\big) \bigvee \Big(\sim C_7(\varphi,y) \,\&\, C_2(\varphi,t,y) \,\&\, \Big(C_4(x,y) \,\&\, \Big(\sim C_7(\varphi,x) \bigvee \big(\sim C_3(t,x) \,\&\, C_8(\varphi,x,y)\big)\Big)\Big)\Big).$$

- For C_8,

$$C_8(\varphi^x_s, z, y) = \big(C_7(\varphi,x) \,\&\, C_8(\varphi,z,y)\big) \bigvee \Big(\sim C_7(\varphi,x) \,\&\, C_2(\varphi,s,x) \,\&\, \Big(C_4(x,y) \,\&\, C_3(s,y) \,\&\, C_8(\varphi,z,y)\Big) \bigvee \Big(C_4(x,y) \,\&\, \sim C_3(s,y) \,\&\, \big(C_8(\varphi,z,y) \bigvee C_8(\varphi,z,x)\big)\Big)\Big).$$

The result follows by this observation and the previous analysis (and by a version of disjunctive normal form for meta-connectives). ⊣

Theorem 3.3.2. *Given a schema S and a formula φ, it is decidable whether there is a concrete-substitution Σ such that $\Sigma(S) = \varphi$.*

Proof. Consider φ a formula and S a schema to which all the conventions were already applied. If the condition in S is *a priori* false, i.e. if using the rules of Convention 3.3.3

one can obtain $\perp$, then there is no concrete-substitution. Suppose now that S is of the form

$$F[\varphi_0,\ldots,\varphi_{n_0},t_0,\ldots,t_{n_1},v_0,\ldots,v_{n_2}]\&\bigvee_{i\in I}\mathop{\&}_{j\in J_i}\sim^{k_j^0}C_{k_j^1}(A_i,t_{k_j^2},v_{k_j^3}),$$

where the condition is not *a priori* false. It is not hard to see that one can computationally decide whether there is a substitution σ such that $\sigma(F)$ and φ have the same layout (the structure of parenthesis, implication signs, negation signs, and universal quantifier signs described before). The idea is the following:

1. If F and φ have different types, then reject (by different types we mean that, for instance, one is a negation and the other is an implication).
2. If they have the same type, then one goes to the layout to the left of both outermost implication signs.
3. If the layout is the same, then one goes to the right and does the same move throughout the process.
4. If one reaches an incompatibility—for example one implication sign versus one negation sign or universal quantification sign—one rejects.
5. One does the same for negations and universal quantifications.
6. Using attributions to the formula-variables, one locally matches the layout of φ.
7. One proceeds by going to the inside of the respective layouts until one reaches either a rejection or an equal layout (this procedure must eventually stop because the layout of the formula, just like the layout of a formula-structure, is finite).

In this briefly described way, one is not creating unnecessary changes in F, it is minimal in that sense. If F cannot be changed to have the same layout as φ, then there is no concrete-substitution (all this is decidable). Suppose that there is such a substitution σ that makes the layout the same. Consider σ in the mentioned minimal conditions, $F' := \sigma(F)$, and $S' := \sigma(S)$. Assume that Convention 3.3.6 was already applied to the schema, as well as Convention 3.3.2 (this will yield a finite number of schemata to which one should do the analysis that follows). Having in mind that F' and φ have the same layout, it is decidable whether there is Σ such that $\Sigma(F') = \varphi$. The idea is the following:

1. Just like what was done in 4 of the proof of Lemma 3.3.1, let us consider all the entries in the layout of φ (that is the same layout of F').
2. Now, consider the finite list of variables occurring in φ under the scope of universal quantifiers and the finite list of terms occurring in φ.
3. One checks if there is any incompatibility between the array of quantifiers and negation signs of each entry of F' and of φ.

4. If there is, one rejects.

5. Otherwise, one defines Σ for the variable-variables occurring in the quantifiers accordingly.

6. Starting from the first entry and going through all entries, one takes for each entry (where the corresponding entry in the formula-structure has a replacement) the respective formula where the occurrences of some terms are replaced by fresh variables—ones should analyze all possibilities. Then, one checks if it is possible to assign formulas, terms, and variables to the entries of F' in such a way that one obtains φ and they satisfy the (decidable) condition of the schema F'. One does this for all entries. For single occurrences of term-variables without replacements, one simply assigns the corresponding term that occurs in the actual term.

7. More precisely, one tests all possible substitutions by considering the formulas and the terms where the occurrences of some terms are replaced by fresh variables (one should vanish over all possibilities), one sees what the substituting term should look like, and one tests all the (finite number of) possibilities by making the fresh variables equal to some of the variables of the considered replacement. For each test one sees if the conditions of the schema are satisfied. For example, if one has the formula $(x+0)+z = y$ in the actual formula and the corresponding formula-structure $(t_0)^{x_0}_{s_0} = x_1$ in the schema, then one should assign x_1 to the variable y and, as in t_0 we are considering one replacement, one should consider the following possibilities for t_0:

   - t_0 as being x_0, with s_0 being $(x+0)+z$;
   - t_0 as being $(x+0)+x_0$, where s_0 is z;
   - t_0 as being x_0+z, where s_0 is $x+0$;
   - t_0 as being $(x_0+0)+z$, with s_0 being x;
   - t_0 as being $(x+x_0)+z$, with s_0 being 0;
   - t_0 as being $(x+0)+z$ with x_0 not occurring in t_0, where x_0 and s_0 are "arbitrary" in what t_0 is concerned; in practise, this means that either they are assigned in the next entries, or they are to be considered as not assigned, which means that one has just to further study them if they appear in the conditions—see the proof of Lemma 3.3.3 for a more detailed account.

   More generally, consider the case where a term-structure in the schema needs to be equal, under a concrete-substitution, to a certain term. Then, one needs to satisfy an equality similar to

$$\Sigma(((t_0)^{\vec{x_0}}_{\vec{s_0}} + \mathsf{S}(t_1)) \times t_2) = ((\mathsf{S}(x+y) + \mathsf{S}(\mathsf{S}(0))) + \mathsf{S}(\mathsf{S}(0) + (x \times z))) \times ((x \times y) + 0),$$

i.e.

$$\left(\Sigma(t_0)^{\Sigma(\vec{x_0})}_{\Sigma(\vec{s_0})}+\mathsf{S}(\Sigma(t_1))\right)\times\Sigma(t_2) = \\ ((\mathsf{S}(x+y)+\mathsf{S}(\mathsf{S}(0)))+\mathsf{S}(\mathsf{S}(0)+(x\times z)))\times((x\times y)+0).$$

This entails that

$$\begin{cases}\Sigma(t_0)^{\Sigma(\vec{x_0})}_{\Sigma(\vec{s_0})} & = (\mathsf{S}(x+y)+\mathsf{S}(\mathsf{S}(0)))\\ \mathsf{S}(\Sigma(t_1)) & = \mathsf{S}(\mathsf{S}(0)+(x\times z))\\ \Sigma(t_2) & = ((x\times y)+0)\end{cases}$$

Thus, the image of t_2 under the concrete-substitution that one wants to construct is fixed, as well as the image of t_1 (if they were already fixed one should test if a contradiction is obtained). This means that, for the desired equality, it only remains to be analyzed the equality $\Sigma(t_0)^{\Sigma(\vec{x_0})}_{\Sigma(\vec{s_0})} = (\mathsf{S}(x+y)+\mathsf{S}(\mathsf{S}(0)))$. For this equality, one makes a (finite) case analysis as before by means of fresh variables (this will yield a similar analysis for the term-structures $\vec{s_0}$). If $\Sigma(t_0)$ is already assigned to, for instance, $\Sigma(t_0) = \mathsf{S}(\Sigma(x_0)+y)+\mathsf{S}(\Sigma(x_i))$, then one substitutes the already attributed t_0 in the desired equality, namely

$$(\mathsf{S}(\Sigma(x_0)+y)+\mathsf{S}(\Sigma(x_i)))^{\Sigma(\vec{x_0})}_{\Sigma(\vec{s_0})} = \mathsf{S}(x+y)+\mathsf{S}(\mathsf{S}(0)),$$

which entails that

$$\mathsf{S}\left(\Sigma(x_0)^{\Sigma(\vec{x_0})}_{\Sigma(\vec{s_0})}+y^{\Sigma(\vec{x_0})}_{\Sigma(\vec{s_0})}\right)+\mathsf{S}\left(\Sigma(x_i)^{\Sigma(\vec{x_0})}_{\Sigma(\vec{s_0})}\right) = \mathsf{S}(x+y)+\mathsf{S}(\mathsf{S}(0)),$$

and so

$$\begin{cases}\Sigma(x_0)^{\Sigma(\vec{x_0})}_{\Sigma(\vec{s_0})} & = x\\ y^{\Sigma(\vec{x_0})}_{\Sigma(\vec{s_0})} & = y\\ \Sigma(x_i)^{\Sigma(\vec{x_0})}_{\Sigma(\vec{s_0})} & = \mathsf{S}(0);\end{cases}$$

something that can, once again, be easily solved through a case analysis. Throughout the process one should add the suitable conditions on the variables. After that, one substitutes the new information about the variables and one sees if any contradiction is reached.

8. We do the previous procedure for each entry of the layout and also for atoms—in fact, for formula-variables with replacements the fresh variables analysis remains valid: if one has $\varphi^{\vec{x}}_{\vec{s}}$, then one considers fresh term-variables t_0 and t_1, one considers φ as being $(t_0 = t_1)$, and one proceeds the analysis as before. For each case in the analysis of a given entry, one should consider all the sub-cases in the other entries for the choices that were made—this gives rise to a tree of possible cases; moving from one entry to another, either a new case analysis is created, or one reaches a contradiction, which, by

its turn, forces the considered case in the already established case analysis to change (in particular this yields that the algorithm as a whole halts). One can computationally check if there are any incompatibilities at any stage; if any incompatibility is detected, one should consider another case in the analysis, if one reaches an incompatibility with all cases, it means that there is no concrete-substitution in the desired conditions. Observe that, for each entry, there is a finite number of ways to do the considered procedure, which entails that in the whole schema there is also a finite number of ways to consider all the possible cases.

9. In the end of a case analysis, one should test to see whether the conditions of the schema are satisfied. This is achieved using Lemma 3.3.3. One should also test if Σ can be made in such a way that satisfies the Definition 3.3.5.

10. If the previous steps are not possible, one should reject.

From Lemma 3.3.2, if there is a concrete-substitution Σ' such that $\Sigma'(S') = \varphi$, then there is a concrete substitution such that $\Sigma(S) = \varphi$. All the mentioned construction yields that one can decide whether there is a concrete-substitution Σ such that $\Sigma(S) = \varphi$. ⊣

It is important to observe that in the former algorithm it is not fundamental that the Convention 3.3.6 is applied: if one has, for instance, the term-structure $(t_r^x)_s^y$ and one wants it to be equal to a certain term, then one proceeds by forcing T_s^y to be equal to that term using the considered analysis, and then one imposes $\Sigma(t_r^x) = \Sigma(T)$ and makes a similar analysis for that fact; this means that if Convention 3.3.6 was not applied, then one has to do several times the creation of the case analysis of stage 7 from the previous algorithm. We opted to firstly apply the convention because it simplifies the analysis and avoids having chained replacements.

3.3.3 Decidability of some proof-skeletons and k-provability

As mentioned in the introduction, the decidability of k-provability for PA with the usual instantiation schema is an open problem and the proof-skeleton problem is in general undecidable for that version of PA. Nevertheless, we will characterize some values of k for which k-provability is decidable and some proof-skeletons for which the corresponding proof-skeleton problem is decidable.

Definition 3.3.6. We say that a proof-skeleton $\mathcal{S}$ is *stable* (*for* T) if there is a finite list of provable-schemata $\mathcal{L}_\mathcal{S}$ such that:

Stability: A formula φ has a proof whose skeleton is $\mathcal{S}$ if, and only if, there are a schema S in $\mathcal{L}_\mathcal{S}$ and a concrete-substitution Σ that satisfy $\Sigma(S) = \varphi$.

We say that a number k is *stable* (*for* T) if all proof-skeletons with length at most k are stable.

For stable proof-skeleton, the corresponding proof-skeleton problem is decidable, as the next result confirms.

Theorem 3.3.3. *If a proof-skeleton $\mathcal{S}$ is stable, then, for any formula φ, it is decidable whether φ has a proof whose skeleton is $\mathcal{S}$.*

Proof. By definition, φ has a proof whose skeleton is $\mathcal{S}$ if, and only if, there are a schema S in $\mathcal{L}_{\mathcal{S}}$ and a concrete-substitution Σ that satisfy $\Sigma(S) = \varphi$. The decidability follows from the fact that $\mathcal{L}_{\mathcal{S}}$ is finite and from Theorem 3.3.2 (one tests computationally for each element of the finite list $\mathcal{L}_{\mathcal{S}}$). ⊣

If k is stable, then the respective k-provability is decidable, as the next result confirms.

Theorem 3.3.4. *If k is a stable number, then, for any formula φ, it is decidable whether $T \vdash_{k\ steps} \varphi$.*

Proof. The result follows from the fact that for each k there is a finite number of proof-skeletons with length k and from Theorem 3.3.3. ⊣

Theorem 3.3.5. *There is a maximum k stable number for* PA′.

Proof. Suppose that there is no maximum k stable number for PA′. As the fact that k is a stable number implies that, for all $s \leq k$, s is a stable number; it follows that all numbers are stable for PA′. Thus, all proof-skeletons are stable. In particular, the proof-skeletons from the proof of Theorem 6.1 from [31]—which is similar to the proof of 5.1 from [58]—are stable; from the previous Theorem, it follows that it is decidable whether a formula has a proof whose skeleton is the skeletons from the proof of Theorem 6.1 from [31], which contradicts that very Theorem (when one adapts the proof to the theory PA′). ⊣

Although Theorem 3.3.4 presents a characterization of some proof-skeletons that have the respective proof-skeleton problem decidable, we have no information on how the lists were created; furthermore, so far we have no general way to generate (some of the) stable proof-skeletons. We now proceed to develop a general way to generate stable proof-skeletons.

From [58], we know that second-order unification is in general undecidable. In the proof presented in that paper, the main idea was to represent, inside the context of second-order unification, numerals, addition, and multiplication; something that is delivered by, c.f. [58], equations that include:

$$s(\tau) = \tau(a/s(a)),$$
$$\tau(a/\sigma_1, b/s(b), c/a \circ (b \circ c)) = \sigma_2 \circ (\sigma_3 \circ \tau).$$

In the previous equations, we followed the notation of the considered paper, where s denotes a unary function-symbol, $\circ$ denotes a binary one, and σ_2, σ_3, τ term-variables. It is important to observe that, in the previous equations, one has in the two members of the equality sign occurrences of the same term-variable and occurrences of variables being replaced by something where that very variables occur. While this last feature can be avoided, using Convention 3.3.6, the first feature will be something that we will not allow in the algorithms that we are going to develop; for example, $\mathsf{S}(t^{x_0}_{x_1}) = t^{x_0}_{\mathsf{S}(x_1)}$ has an infinity of solutions, namely $t = \mathsf{S}^n(x_0)$, that cannot be written in a closed form like $t = \mathscr{E}$, where in $\mathscr{E}$ we do not have occurrences of natural numbers variables like $\mathsf{S}^n(x_0)$. We will create an algorithm that will be able to solve *some* systems of the form

$$\begin{cases} r_0 &= s_0 \\ &\vdots \\ r_n &= s_n, \end{cases}$$

where $r_0, \dots, r_n, s_0, \dots, s_n$ are term-structures (we will assume that the equations in the left do not have common variables with the equations on the right): by a solution we mean a common substitution to the left and the right side such that one achieves the equality between them maybe after the application of several of the conventions (in fact, we might achieve the equality after the application of several substitutions and several conventions)—for each substitution there might be several ways to apply the conventions that yield several solutions. The main move that we are going to make is to avoid the mentioned occurrences of equal term-variables while the algorithm is running, this offers no problem since we will not develop a general algorithm and since the reasoning of [58] cannot be applied—we force the algorithm not to have the needed conditions for the proof, and thus avoid the undecidability; we do this with the cost that the algorithm will reject or not halt systems that indeed have a solution.

Algorithm 3.3.1. We will describe an algorithm that, *for certain cases*, sees whether there is a substitution σ such that

$$\begin{cases} \sigma(r_0) &= \sigma(s_0) \\ &\vdots \\ \sigma(r_n) &= \sigma(s_n), \end{cases}$$

where $r_0, \dots, r_n, s_0, \dots, s_n$ are term-structures. Furthermore, such a σ can be found (for some cases) without the introduction of unnecessary complexity. Following Convention 3.3.4, we assume that the r_i's and s_j's do not have common meta-variables. Furthermore, we assume that Convention 3.3.6 was already applied. The algorithm is as follows:

1. Starting from $i = 0$, do the following to construct a list of equations:

   1.1. If r_i and s_i are term-structures without term-variables, then see if it is possible to identify the variable-variables in such a way that $\sigma(r_i) = \sigma(s_i)$ and add this fact to the list; if it is not possible, then reject.

1.2. If r_i or s_i is of the form $t_{\vec{s}}^{\vec{x}}$, then add to the list of equations $t_{\vec{s}}^{\vec{x}} = s_i$, when $r_i = t_{\vec{s}}^{\vec{x}}$, and $r_i = t_{\vec{s}}^{\vec{x}}$, for the other case.

1.3. If r_i and s_i are both term-structures with an outermost occurrence of a function-symbol, then test whether the function-symbol is the same.

1.4. If it is not, then reject.

1.5. If $r_i = \mathsf{S}(t_0)$ and $r_1 = \mathsf{S}(t_1)$, with t_0 and t_1 term-structures, then apply the previous procedure to t_0 and t_1 and add $t_0 = t_1$ to the list. Do the same for + and ×. This means that for each pair of terms r_i and s_i, one should see if $r_i = t_0 + t_1$ and $s_i = t_2 \times t_3$, or $r_i = t_2 \times t_3$ and $s_i = t_0 + t_1$. If it occurs, then reject; and for each pair of terms r_i and s_i with $r_i = t_0 \circ t_1$ and $s_i = t_2 \circ t_3$, do the previous procedure for t_0 with t_2, and t_1 with t_3, where $\circ$ is + or ×, and add $t_0 = t_2$ and $t_1 = t_3$ to the list.

1.6. Do the previous procedure until no further reductions are possible (this must stop after a finite number of steps). Thus, one should apply the procedure until one reaches either a rejection or has analyzed all possible cases.

1.7. Increment i until all the values for i were considered.

2. Let us now assume that a list was build without reaching a rejecting state.

3. As mentioned in 1.1, one has to make the suitable variable identification (for example using fresh variable-variables). For instance, if one has in the list an equation of the form

$$(\cdots + x) \times \cdots = (\cdots + y) \times \cdots,$$

where x and y occur in the same place of the layout of the function-symbols, then one has to assign x and y to a new fresh common variable-variable. With this, we get the true equality

$$(\cdots + \sigma(x)) \times \cdots = (\cdots + \sigma(y)) \times \cdots.$$

If any of them was already assigned, assign all the variables previously assigned to this new common fresh variable.

4. Proceed in a similar way with the occurrence of term-variables that do not occur under the scope of a replacement. This means that if one has an equation of the form

$$((t_0 \times t_1) + \mathsf{S}(0)) \times \cdots = ((t_2 \times y) + t_3) \times \cdots,$$

then $\sigma(t_0) = \sigma(t_2) = t$, with t fresh, $\sigma(t_1) = \sigma(y)$, and $\sigma(t_3) = \mathsf{S}(0)$.

5. Let us now briefly describe the most complex case.

6. Suppose we have in the list equations of the following form (observe that the analysis of the list can be reduced to the analysis of the next system), where we are implicitly considering in the left-side the r-part, and in the right-side the s-part:

$$n \text{ equations} \begin{cases} (t_0)^{\vec{x_0}}_{\vec{s_0}} & = (\cdots + (x+0)) \times (\cdots (y+x) \times 0) \\ (t_0)^{\vec{x_2}}_{\vec{s_2}} & = ((\cdots \times (x \times 0))) \times t_2 \\ (\cdots \times (x + \mathsf{S}(0))) + 0 & = (t_4)^{\vec{x_5}}_{\vec{s_5}} \\ \vdots & \\ (t_5)^{\vec{x_6}}_{\vec{s_6}} & = (\cdots + (y \times 0)) \times \mathsf{S}(\mathsf{S}(0)). \end{cases}$$

Do the following (the next procedure is a kind of co-recursion because one should apply the whole procedure to smaller parts of the very same procedure):

6.1. Firstly, substitute the already changed variable-variables and term-variables in the equations (one should do this step at every stage of the algorithm).

6.2. If at any stage one obtains an equality where in both members one has an occurrence of $t^{\vec{x}}_{\vec{s}}$, one should output **problem**.

6.3. If at any stage one obtains a non-trivial equality (i.e. not yet syntactically satisfied) where in both members one has an occurrence of the same term-variable one should output **problem**.

6.4. Starting from $i = 1$ (up to n), solve, if possible, the first equation in the following way:

6.4.1. Consider the term-structure that corresponds to the right-side (respectively left-side) of the equation where some occurrences of term-structures in the considered equation were replaced by fresh variables (analyze all the finitely many possibilities for this).

6.4.2. Analyze all the possibilities to identify the fresh variables with the variables-variables $\sigma(\vec{x_0})$ and see what the term-structures $\sigma(\vec{s_0})$ ought to be—keep in mind that all the conventions ought to be satisfied (this fact can be verified in a decidable way)[4]. If one has one of the variables $\sigma(\vec{x_0})$ occurring in the opposite side, one should reject, since that is an impossibility[5]. One should always substitute the already found solutions in the new equations.

6.4.3. In $\sigma(\vec{s_0})$ there might occur complex term-structure that would require an analysis similar to the one that we are considering (for example, s_0 could be $t^{x}_{t_0+t_1}$) and would yield a new system to be solved (to create the system we replicate the previous steps for the creation of the list and the fresh variables analysis). Without loss of generality, we might proceed our analysis

[4]This idea was also used at stage 7 in the second procedure of the proof of Theorem 3.3.2.

[5]In the context of (provable) schemata, one should add condition C_4 for the variables that are being considered as being different throughout the procedure (the analogous situation for the other conventions).

because the algorithm will account for all the cases due to its co-recursive nature. We allow the occurrence of equations of the form

$$(t_0)^{\vec{x_0}}_{\vec{s_0}} = ((x+y)\times 0) + (\mathsf{S}(t_1) + \mathsf{S}(0)).$$

6.4.4. For instance, in the case analysis of the previous stage, one might have $t^{x_0}_{r^{x_1}_s} = (x+y)\times z$. One should consider the cases as in 7 in the second procedure of the proof of Lemma 3.3.2 and as mentioned before. One of the cases is $\sigma(t) = \sigma(x_0)\times\sigma(z)$ with $\sigma(r^{x_1}_s) = \sigma(x+y)$. Then, for the previous equality, one should also carry a similar analysis. For example, $\sigma(r) = \sigma(x_1)+\sigma(y)$, with $\sigma(s) = \sigma(x)$. In all, for the considered case, we get

$$\sigma(t^{x_0}_{r^{x_1}_s}) = \sigma(t)^{\sigma(x_0)}_{\sigma(r)^{\sigma(x_1)}_{\sigma(s)}} = (\sigma(x_0)\times\sigma(z))^{\sigma(x_0)}_{\sigma(r)^{\sigma(x_1)}_{\sigma(s)}} = \sigma(r)^{\sigma(x_1)}_{\sigma(s)}\times\sigma(z) =$$
$$(\sigma(x_1)+\sigma(y))^{\sigma(x_1)}_{\sigma(s)}\times\sigma(z) = (\sigma(x)+\sigma(y))\times\sigma(z) = \sigma((x+y)\times z).$$

Observe that, in the previous analysis, we considered the variables as being different when the replacements are applied because they are fresh, they do not occur at all. Furthermore, observe that, for the considered case, $\sigma(z)$ cannot be identified with $\sigma(x_0)$ just like $\sigma(x_1)$ cannot be identified with $\sigma(y)$ (otherwise we would get a contradiction, in particular if we apply concrete-substitutions)—one should add the suitable conditions to express this fact. This means that the replacements that we are considering here act only on the fresh variables, without the possibility of overlapping the left with the right side, as mentioned.

6.4.5. If a term-variable occurs in the other side of the equation, one should also consider, together with the fresh variables construction, the case where for that entry one places a fresh term-variable in the term-structure that one is creating. For example, if one has

$$(t_0)^x_s = (\mathsf{S}(0)+y)+t_1,$$

one of the cases that one should analyze is the case where $\sigma(t_0) = (\sigma(x)+\sigma(y))+t$, with t fresh, which entails that $\sigma(t_1) = t^{\sigma(x)}_{\sigma(s)}$. This serves to cover the possibility, under the interpretation of the equality via concrete-substitutions, that the concrete term that results from t_1 was placed using the considered replacement to some other term in that very position in the function-symbols layout.

6.4.6. Take note of the possible ones—the admissible possibilities form a tree, in the sense that for one case one might have to analyze a great variety of sub-cases. Unlike the proof of Theorem 3.3.2 where the second algorithm only asks for a successful path in the tree, in this algorithm we want to study all admissible possibilities, since we want to find all the solutions to the system.

6.4.7. If one reaches an impossibility in the case analysis one should reject that case and consider the other remaining cases.

6.5. If, at any stage, there are no solutions, one should reject.

6.6. Suppose that the system was solved, if possible, for $i < n$. Then, do the following:

6.6.1. Consider the $(i+1)$-th equation, say $(t_5)^{\vec{x_6}}_{\vec{s_6}} = (\cdots + (y \times 0)) \times \mathsf{S}(\mathsf{S}(0))$.

6.6.2. Substitute the already found term-structures and, if necessary, create new systems for the equalities that emerge.

6.6.3. In particular, if t_5 was already found, then substitute it for the found solution, make a case analysis for the applications of the replacements, and solve the considered equation. If t_5 was not yet assigned, then solve the equation using the fresh variables analysis and with 6.2.5.

6.6.4. Substitute in the previously obtained solutions the new ones and if necessary solve the new equations that emerge.

7. Output all the found solutions.

We say that a system is *successful* if it halts and is not rejected as a whole (nevertheless, we allow that in the case analysis some cases are rejected), and if in the case analysis no **problem** situation appeared. What we described is not a complete description of the algorithm, but it has the main ideas that are necessary to develop the much more complex and involved complete description of the algorithm. We considered particular cases in the algorithm to emphasize the main ideas. For instance, in the hard case that we presented—the one with a new system—, a more general approach is needed to write the complete algorithm.

Let us give an example of the some steps of the previous procedure. Suppose that one is given the system

$$\begin{cases} (t_0)^{x_0\ x_1}_{s_0\ s_1} & = (x \times y) + t_1 \\ (t_0)^{x_3}_{s_3} & = (x \times x) + \mathsf{S}(0). \end{cases}$$

As the algorithm suggests, one should start by considering the first equation (the other cases should also be studied). One should consider all the sub-term-structures of $(x \times y) + t_1$ and replace some occurrences by fresh variables, and then unify them, using substitutions, with x_0 or x_1; furthermore, one should proceed with the term-variables as described in 6.4.5. For example, for $(x \times y) + t_1$, one could consider, after the suitable unifications, t_0 as being $(x_0 \times x_1) + t$, t_1 as being $t^{x_0\ x_1}_{s_0\ s_1}$, where t is fresh, s_0 as being x, and s_1 as being y. This constitutes a solution to the first equation.

Following the algorithm, then one should substitute the obtained solution in the second equation, giving $((x_0 \times x_1) + t)^{x_3}_{s_3} = (x \times x) + \mathsf{S}(0)$, i.e.

$$((x_0)^{x_3}_{s_3} \times (x_1)^{x_3}_{s_3}) + t^{x_3}_{s_3} = (x \times x) + \mathsf{S}(0).$$

This entails that

$$\begin{cases} (x_0)^{x_3}_{s_3} \times (x_1)^{x_3}_{s_3} & = x \times x \\ t^{x_3}_{s_3} & = \mathsf{S}(0). \end{cases}$$

So, after another case analysis, one of the cases is

$$\begin{cases} (x_0)^{x_3}_{s_3} & = (x_1)^{x_3}_{s_3} = x \\ t^{x_3}_{s_3} & = \mathsf{S}(0). \end{cases}$$

Hence, for this case, $x_0 = x$ and x_0 is not x_3, or $x_0 = x_3$ and $s_3 = x$; similarly for x_1. Let us now focus in the second equation from the previous system. We proceed as before with a case analysis, one of the cases yields that t as being $\mathsf{S}(0)$; after that, one substitutes t in t_0, yielding $(x_0 \times x_1) + \mathsf{S}(0)$, and in t_1, yielding $\mathsf{S}(0)$. Observe that in the considered cases no **problem** was identified (it is not hard to see that the previous system is successful).

In the end we get, if not all the cases are rejections and no **problem** was obtained, substitutions σ in the desired conditions. Furthermore, such substitutions are most general ones, in the sense that they do not introduce unnecessary complexity and unnecessary identifications (just like a most general unifier).

Convention 3.3.8. We assume that **A** is a generic algorithm that solves some unification (of term-structures) problems (for example by a case analysis) and without the introduction of unnecessary complexity—i.e. has as output a *finite* number of most general unifiers in the sense that we used previously—, with some situations where it might output **problem**. We assume that **A** works in a very similar way to Algorithm 3.3.1 (for instance, it might use a case analysis, it might use substitution of found solutions, it might use the creation of the systems, by considering all possible situation for the system, for some of them might output **problem**, etc). We will say that **A** is *successful* for a given system if it halts and no **problem** situations were identified during the computations, just like what was considered for the previous algorithm. We assume that **A** gives informations about the conditions needed for each step in the context of provable schemata (just like Algorithm 3.3.1) and that has the following feature:

Sub. property: If **A** is successful for the system

$$\begin{cases} r_0 & = s_0 \\ & \vdots \\ r_n & = s_n, \end{cases}$$

and there are concrete-substitutions Σ_0 and Σ_1 such that

$$\begin{cases} \Sigma_0(r_0) & = \Sigma_1(s_0) \\ & \vdots \\ \Sigma_0(r_n) & = \Sigma_1(s_n), \end{cases}$$

then there are a solution σ of the system that is constructed by the algorithm **A** and a concrete-substitution Σ such that

$$\begin{cases} \Sigma_0(r_0) = \Sigma(\sigma(r_0)) & = \Sigma(\sigma(s_0)) = \Sigma_1(s_0) \\ & \vdots \\ \Sigma_0(r_n) = \Sigma(\sigma(r_n)) & = \Sigma(\sigma(s_n)) = \Sigma_1(s_n). \end{cases}$$

We also assume that the previous property is compatible with the conventions that the algorithm gives as additional information and that if t is a term-structure, then $\Sigma(\sigma(t)) = \Sigma_i(t)$, for $i = 0, 1$, depending on the concrete-substitution for which the value of t defined (this last condition is useful when one is dealing with provable schemata).

It is important to observe that Algorithm 3.3.1 has the **Sub. property**: if the algorithm 3.3.1 is successful for a considered system and one is given concrete-substitutions as before, then, guided by the concrete-substitutions Σ_0 and Σ_1, one can use the Algorithm 3.3.1 to obtain a desired solution σ—here the concrete-substitutions can be used to identify the choices that one has to make while running the algorithm; moreover, the algorithm will be successful (by hypothesis); the existence of a concrete-substitution Σ with the mentioned properties follows from the fact that all the choices that were made for the construction of σ were guided by the concrete-substitutions Σ_0 and Σ_1.

Observe that one can conceive several algorithm that satisfy the **Sub. Property**; for instance, one can extend Algorithm 3.3.1 to account for other possibilities without outputting **problem** so often. For example, one could extend Algorithm 3.3.1 by allowing situations where one has occurrences of replacements in the two sides of the equality, like

$$(t_0)^{x_0}_{s_0} = (\mathsf{S}(\mathsf{S}(0)) + x_1) \times (t_1)^{x_1}_{s_1},$$

under the proviso that no variable-variable in the replacement in the opposite side can be identified with a variable-variable that occurs in the considered replacement. Observe that x_1 is a variable-variable in the replacement that occurs in the opposite side of the considered replacement, but, under the concrete-substitution interpretation, it cannot be x_0, since it occurs again in the right side, namely in $(\mathsf{S}(\mathsf{S}(0)) + x_1)$. Then, one makes the usual fresh variable analysis, but, similarly to stage 6.4.5, one also accounts for the possibility of internal application of the replacement; in the considered example, this means that one of the cases to be considered is the one where t_0 is $(x_0 + x_1) \times t^{x_1}_{s_1}$, with t fresh, s_0 is $\mathsf{S}(\mathsf{S}(0))$, and t_1 is $t^{x_0}_{s_0}$. Let us briefly justify what was described. Let us suppose now that t_0, t_1, s_0, and s_1 are concrete terms, and x_0, x_1 concrete variables. Assume that x_0 is not x_1 and that

$$(t_0)^{x_0}_{s_0} = (t_1)^{x_1}_{s_1},$$

where both variables are being replaced, i.e. x_0 occurs in t_0 and x_1 in t_1. It is not hard to conclude that x_0 cannot occur in s_1, just like x_1 cannot occur in s_0. Consider t as being the term obtained from $(t_0)^{x_0}_{s_0}$ (that is the same as $(t_1)^{x_1}_{s_1}$) by replacing the occurrences of s_0 that were placed using the replacement by x_0, and the same for s_1 and x_1. Then, t has x_0 and x_1 as variables. Furthermore, it follows that

$$\begin{cases} t^{x_0}_{s_0} = & t_1 \\ t^{x_1}_{s_1} = & t_0, \end{cases}$$

as desired. This justifies the described procedure that one can add to Algorithm 3.3.1. Observe that in the justification it was used the fact that x_0 is not x_1, otherwise the construction of t could fail; for example, for $t_0 = \mathsf{S}(\mathsf{S}(x_0))$, $s_0 = x$, $t_1 = \mathsf{S}(x_1)$, and $s_1 = \mathsf{S}(x)$ one has that $(t_0)^{x_0}_{s_0} = (t_1)^{x_1}_{s_1}$, but one cannot construct t as before, since s_0 is being placed in the same place as s_1. It is worth mentioning that the described impossibility is in the heart of the undecidability of second-order unification: keep in mind that, for example, $\mathsf{S}(t^{x_0}_{x_1}) = t^{x_0}_{\mathsf{S}(x_1)}$ has an infinity of solutions, namely $t = \mathsf{S}^n(x_0)$, that cannot be written in a closed form without using natural numbers, in fact a variation of this is used to represent the natural numbers inside the context of second-order unification in [58] (the idea of the proof of the undecidability is to represent natural numbers, addition, and multiplication inside second-order unification and apply Matijasevič's Theorem).

We believe that Algorithm 3.3.1 halts for every input, but we do not have a proof of that fact or a counter-example to it; the reason for this is that we believe that if a loop situation is reached using substitutions and replacements, then one must have an occurrence of the same term-variable in both sides of a given equation, something that is accounted for by the algorithm by just outputting **problem**. All this concern about the halting nature of algorithm 3.3.1 is not necessary for what we are going to develop because we want to account for other algorithms that might not halt on every input (that is why we assumed the previous convention), thus the halting nature of algorithm 3.3.1 is a side discussion to our goal.

We move to create the desired lists that give a more concrete inside of Theorem 3.3.3.

List 3.3.1. We now proceed to create lists, for each k and each algorithm $\mathbf{A}$, of provable schemata, $\mathscr{L}_{\mathbf{A},k}$, and proof-skeletons, $\mathfrak{L}_{\mathbf{A},k}$.

Basis case: The list $\mathscr{L}_{\mathbf{A},0}$, for the case $k = 0$, is simply the (finite) list of axioms. The list $\mathfrak{L}_{\mathbf{A},0}$ is the list of the numbers of the schemata that are axioms.

Inductive step: Suppose that the lists $\mathscr{L}_{\mathbf{A},s}$ and $\mathfrak{L}_{\mathbf{A},s}$, with $s \leq k$, where already created. Add all elements of $\mathscr{L}_{\mathbf{A},k}$ to $\mathscr{L}_{\mathbf{A},k+1}$, and all elements of $\mathfrak{L}_{\mathbf{A},k}$ to $\mathfrak{L}_{\mathbf{A},k+1}$. Consider the following cases:

Gen: If S is a schema in $\mathscr{L}_{\mathbf{A},k}$, then pick x a variable-variable and add $\forall x.S(\& C)$ to $\mathscr{L}_{\mathbf{A},k+1}$—the respective schema obtained by placing a universal quantifier, just like in Definition 3.3.4. If $\mathcal{S}$ is a proof-skeleton in $\mathfrak{L}_{\mathbf{A},k}$, then add Gen($\mathcal{S}$) to $\mathfrak{L}_{\mathbf{A},k+1}$.

MP: Take $k = k_0 + k_1$. Do the following:

1. Consider S_0 a schema in $\mathscr{L}_{\mathbf{A},k_0}$ and S_1 a schema in $\mathscr{L}_{\mathbf{A},k_1}$.
2. If S_1 is a universal quantification or a negation, then reject and consider another pair.
3. Suppose that
   - $S_0 := F[\varphi_0^0,\ldots,\varphi_{n_0}^0,t_0^0,\ldots,t_{n_1}^0,v_0^0,\ldots,v_{n_2}^0]\&\bigvee_{i\in I^0}\&_{j\in J_i^0}\sim^{k_j^{0,0}}C_{k_j^{0,1}}(A_i^0,t_{k_j^{0,2}},v_{k_j^{0,3}})$, and
   - $S_1 := G[\varphi_0^1,\ldots,\varphi_{n_3}^1,t_0^1,\ldots,t_{n_4}^1,v_0^1,\ldots,v_{n_5}^1]\to H[\varphi_0^2,\ldots,\varphi_{n_6}^2,t_0^2,\ldots,t_{n_7}^2,v_0^2,\ldots,v_{n_8}^2]\&\bigvee_{i\in I^1}\&_{j\in J_i^1}\sim^{k_j^{1,0}}C_{k_j^{1,1}}(A_i^1,t_{k_j^{1,2}},v_{k_j^{1,3}})$.
4. We now proceed to see whether F and G can be unified by means of a substitution σ using the algorithm **A** and the previously developed methods. We assume that the conventions were already applied to these schemata.
5. Using the ideas of the previous proofs (with the suitable adaptations), it is not hard to see that one can test whether F and G have a common layout.
6. If they do not have, then reject and consider another pair of schemata.
7. If they have, find the common layout in such a way that unnecessary complexity is avoided (follow the ideas of the previous proofs).
8. Starting from the first entry of the common layout, do the following:
   8.1. Proceed with the quantifiers and negation signs as before (for instance, like in the proof of Lemma 3.3.1).
   8.2. One should create a system for the cases where in both entries one has something of the form $X_0\cdots X_n r = s$, with r and s term-structures. Act accordingly with the quantifiers and the negation signs. If it is not possible, then reject and consider another pair of proof-skeletons.
   8.3. One should run algorithm **A** (for example Algorithm 3.3.1) for the needed equalities of term-structures that emerge.
   8.4. If **A** is not successful, then reject and consider another pair of schemata.
   8.5. Assume for the rest of the procedure that the algorithm **A** is successful. Take note of all solutions.
   8.6. Consider the case where one has in one entry $X_0^0\cdots X_{\ell_0}^0(\varphi_0)_{\vec{s_0}}^{\vec{x_0}}$ and in the other one has $X_0^1\cdots X_{\ell_1}^1 r = s$. Firstly, act accordingly with the quantifiers and negation signs (see, for example, 4.1.1 of the proof of Lemma 3.3.1). If it is not possible, then reject and consider another pair of schemata.
   8.7. If any of the $\vec{x_0}$ appears in $r = s$, then reject and consider another pair of schemata. If φ_0 was already assigned to something of the form $X_0^2\cdots X_{\ell_2}^2 r' = s'$, then apply the considered replacement (this yields a case analysis). Apply the conventions and force it to be $X_0^1\cdots X_{\ell_1}^1 r = s$; this will give rise to another system that one should solve using

algorithm **A**. If the system is not successful, then reject and consider another pair of schemata. If it is, save the solutions. If φ_0 was not yet assigned to such a structure, then consider φ_0 as being $t = t'$, with t and t' fresh term-variables. Run the algorithm **A** to solve the systems that emerge and proceed only in the case where the algorithm is successful (all the cases analyzed should include the suitable information about the conditions that are needed at each stage).

8.8. One now considers all entries were ones has something of the form $X_0 \cdots X_n(\varphi_0)_{\vec{s_0}}^{\vec{x_0}}$ in both entries.

8.9. We assume, due to our conventions, that in the considered case we do not have the same formula-variable occurring. Act accordingly with the quantifiers and the negation signs. If it is not possible, then reject and consider another pair of schemata.

8.10. If both formula-variables were already assigned to something of the form $X_0 \cdots X_{\ell_0} r = s$, then apply the replacements, apply the conventions, and, for each case of the conventions, solve the obtained system in the previously mentioned ways. If any of the systems is not successful, reject and consider another pair of schemata.

8.11. If only one of them was mapped to the mentioned structure, adapt step 8.6.

8.12. Suppose now that non of the formula-variables was assigned. Then, one should unify them like what was done for the case of Algorithm 3.3.1 for term-variables. If one needs to satisfy an equality where in both members of the equality one has something of the form $\varphi_{\vec{s}}^{\vec{x}}$, then reject and consider another pair of schemata; otherwise proceed as before by substituting the already attributed values. For example, if one has the equations $\varphi_0 = \varphi_1$, $\varphi_1 = \varphi_2$, then one assigns all those formula-variables to a common fresh formula-variable, say φ; if one has $\varphi_0 = \varphi_s^x$ and $\varphi_2 = \varphi_0$, then one assigns φ_2 to φ_s^x.

8.13. Apply the previous steps to all entries.

9. If we do not get a rejection for the considered schemata in the previous procedure, then apply to each of the final results the conventions and then add the all resulting provable schemata to the list $\mathscr{L}_{\mathbf{A},k+1}$:

$$
\begin{aligned}
H[\sigma(\varphi_0^2),\ldots,\sigma(\varphi_{n_6}^2),\sigma(t_0^2),\ldots,\sigma(t_{n_7}^2),\sigma(v_0^2),\ldots,\sigma(v_{n_8}^2)]\& \\
\bigvee_{i\in I^0} \mathop{\&}_{j\in J_i^0} \sim^{k_j^{0,0}} C_{k_j^{0,1}}(\sigma(A_i^0),\sigma(t_{k_j^{0,2}}),\sigma(v_{k_j^{0,3}}))\& \\
\bigvee_{i\in I^1} \mathop{\&}_{j\in J_i^1} \sim^{k_j^{1,0}} C_{k_j^{1,1}}(\sigma(A_i^1),\sigma(t_{k_j^{1,2}}),\sigma(v_{k_j^{1,3}}))\Big(\mathop{\&} C\Big),
\end{aligned}
$$

where C are, possibly, the conditions that appear from the conventions and the solutions of the considered systems.

10. Consider $\mathcal{S}_0$ a proof-skeleton in $\mathfrak{L}_{\mathbf{A},k_0}$ and $\mathcal{S}_1$ a proof-skeleton in $\mathfrak{L}_{\mathbf{A},k_1}$.
11. If for all schemata S_0 in $\mathscr{L}_{\mathbf{A},k_0}$ with skeleton $\mathcal{S}_0$ and S_1 in $\mathscr{L}_{\mathbf{A},k_1}$ with skeleton $\mathcal{S}_1$, the previous procedure does not yield a rejection, then one should add $\mathrm{MP}(\mathcal{S}_0, \mathcal{S}_1)$ to $\mathfrak{L}_{\mathbf{A},k+1}$; if for any of them one rejects, then one should consider another pair of proof-skeletons and do the same move.

It is important to observe that the previous constructions do not contradict the undecidability of second-order unification problem (see, for instance, [58]): we are considering different types of terms, namely term-structures; we are considering a different type of substitutions σ; and in Algorithm 3.3.1 we do not allow the occurrence of a term-variable in both sides of an equation, something that is indispensable in the proof of the undecidability of second-order unification in [58]. One should keep in mind that the algorithms that we are considering are all necessarily partial—they cannot solve successfully all systems.

Observe that for each k and $\mathbf{A}$, the lists $\mathscr{L}_{k,\mathbf{A}}$ and $\mathfrak{L}_{k,\mathbf{A}}$ are finite—this follows by construction, in particular from the fact that for each convention there is a finite number of ways to apply it, and from the fact that the systems that are successful have a finite number of (most general) solutions for the considered algorithm. Although we presented an inductive construction of the lists, they are not necessarily computable uniformly in k; nevertheless, for each fixed k, $\mathscr{L}_{\mathbf{A},k}$ and $\mathfrak{L}_{\mathbf{A},k}$ are computable due to the fact that they are finite (and all finite lists are computable). The computable uniformity of the lists would entail that, for small values of k, one could, for the considered algorithm $\mathbf{A}$, computationally decide if a given system is successful for $\mathbf{A}$ (a feature that fails for most algorithms).

Definition 3.3.7. We say that a proof-skeleton $\mathcal{S}$ is *grounded for* $\mathbf{A}$ if $\mathcal{S}$ is in $\mathfrak{L}_{\mathbf{A},k}$, where k is the number of steps of $\mathcal{S}$. We say that k is a *grounded number for* $\mathbf{A}$ if all proof-skeleton whose number of steps is at most k are grounded for $\mathbf{A}$. We will consider $\mathscr{L}_{\mathbf{A}} := \cup_k \mathscr{L}_{\mathbf{A},k}$ and $\mathfrak{L}_{\mathbf{A}} := \cup_k \mathfrak{L}_{\mathbf{A},k}$.

If one makes some assumptions about the way the lists were generated—if one assumes that there is a general way to create them, if one assumes that in the construction one does not include unnecessary complexity, etc—, for most cases, the stable proof-skeletons are also grounded for some algorithm $\mathbf{A}$.

The intuition behind grounded proof-skeletons is that, to such skeletons, one can apply the intuitive reasoning made in the beginning of Section 3.3.1 for the analysis of the proofs whose general structure is given by the skeleton MP([L2],MP([L2],[L1])). Furthermore, a grounded number is a number such that all proof-skeletons of proofs that have that very same number as the maximum number of steps are grounded.

Theorem 3.3.6. *Given $\mathcal{S}$ a grounded proof-skeleton for* **A** *in $\mathfrak{L}_{\mathbf{A},k}$, if φ has a proof whose skeleton is $\mathcal{S}$, then there are a schema S in $\mathscr{L}_{\mathbf{A},k}$ with skeleton $\mathcal{S}$, generated by the algorithms, and a concrete-substitution Σ such that $\Sigma(S) = \varphi$.*

Proof. Let us prove the result by induction on k. If $\mathcal{S}$ is in $\mathfrak{L}_{\mathbf{A},0}$, then $\mathcal{S}$ is the number of an axiom; thus, if φ has a proof whose skeleton is $\mathcal{S}$, then there are a schema S in $\mathscr{L}_{\mathbf{A},0}$ with skeleton $\mathcal{S}$, generated by the algorithms, and a concrete-substitution Σ such that $\Sigma(S) = \varphi$. Suppose, by induction hypothesis, that the result holds for $s \leq k$. Suppose that $\mathcal{S}$ is in $\mathfrak{L}_{\mathbf{A},k+1}$ and that φ has a proof whose skeleton is $\mathcal{S}$. Consider the following cases:

$\mathcal{S} = \mathrm{Gen}(\mathcal{S}_0)$**:** In this case, one must have $\varphi = \forall x.\psi$ and ψ should have a proof whose skeleton is $\mathcal{S}_0$. By construction, $\mathcal{S}_0$ must be in $\mathfrak{L}_{\mathbf{A},k}$. By induction hypothesis, there are a schema S_0 in $\mathscr{L}_{\mathbf{A},k}$ with skeleton $\mathcal{S}_0$, generated by the algorithms, and a concrete-substitution Σ such that $\Sigma(S_0) = \psi$. It is clear that the schema $\forall v.S_0$, obtained by S_0 using the generalisation rule, is in $\mathscr{L}_{\mathbf{A},k+1}$; furthermore, $\forall v.S_0$ has skeleton $\mathcal{S}$. Moreover, one can extend Σ in such a way that $\Sigma(\forall v.S_0) = \forall x.\Sigma(S_0) = \forall x.\psi = \varphi$.

$\mathcal{S} = \mathrm{MP}(\mathcal{S}_0, \mathcal{S}_1)$**:** In this case, there must be a ψ such that ψ has a proof whose skeleton is $\mathcal{S}_0$ and $\psi \to \varphi$ has a proof whose skeleton is $\mathcal{S}_1$. So, one has $\mathcal{S}_0$ in $\mathfrak{L}_{\mathbf{A},k_0}$ and $\mathcal{S}_1$ in $\mathfrak{L}_{\mathbf{A},k_1}$, with $k = k_0 + k_1$. By induction hypothesis, there are schemata S_0 in $\mathscr{L}_{\mathbf{A},k_0}$ with skeleton $\mathcal{S}_0$ and S_1 in $\mathscr{L}_{\mathbf{A},k_1}$ with skeleton $\mathcal{S}_1$, and concrete-substitutions Σ_0 and Σ_1 such that $\Sigma_0(S_0) = \psi$ and $\Sigma_1(S_1) = \psi \to \varphi$. Lemma 3.3.1 guarantees that there is a substitution that unifies F and G; furthermore, guided by the concrete-substitutions Σ_0 and Σ_1, one can use the previous algorithms to obtain a minimal unifier σ for the algorithm **A** (the concrete-substitutions can be used to see what choices should be done while running the algorithm)—this follows, more directly, from the **Sub. property** of Convention 3.3.8 (and from the fact that a version of **Sub. Property** holds for formula-variables); moreover, the algorithm will be successful because $\mathcal{S}$ is grounded. Thus, using σ from the algorithm, $\sigma(F) = \sigma(G)$. Clearly, the schema $\sigma(H)$—this schema includes the suitable conventions—is in $\mathscr{L}_{\mathbf{A},k+1}$. Moreover, using the reasoning of the proof of Lemma 3.3.1 and the **Sub. property**, one can guarantee the existence a concrete-substitution Σ such that $\Sigma(\sigma(H)) = \varphi$.

The result follows by induction. ⊣

Corollary 3.3.1. *Given $\mathcal{S}$ a grounded proof-skeleton for* **A***, for each formula φ, it is decidable whether there is a proof of φ whose skeleton is $\mathcal{S}$.*

Proof. Consider $\mathcal{S}$ a grounded proof-skeleton for **A**. Generate, using the previous procedures for the algorithm **A**, the schemata that have $\mathcal{S}$ as proof-skeleton, let us call this finite list $\mathscr{R}$. These obtained schemata are minimal in the sense that there was added no unnecessary complexity to the substitutions. Let us prove that φ has a proof whose skeleton is $\mathcal{S}$ if, and only if, there are a concrete-substitution Σ and S in $\mathscr{R}$ such that $\Sigma(S) = \varphi$. It is clear that if there are a concrete-substitution Σ and S in $\mathscr{R}$ such that

$\Sigma(R) = \varphi$, then φ has a proof whose skeleton is $\mathcal{S}$ (this can be proved by induction on the definition of proof-skeleton). Let us prove the other direction. Suppose that φ has a proof whose skeleton is $\mathcal{S}$. Then, by the previous result, there are a schema S with skeleton $\mathcal{S}$, generated by the algorithms, and a concrete-substitution Σ such that $\Sigma(S) = \varphi$. Clearly S is in $\mathcal{R}$, so the desired result holds.

The decidability follows from the fact the the list $\mathcal{R}$ is finite and from Theorem 3.3.2. ⊣

From the proof of the previous result we can conclude that every grounded skeleton is stable. Thus, every grounded number is stable.

Theorem 3.3.7. *For k a grounded number for* $\mathbf{A}$*, $T \vdash_{k\ steps} \varphi$ if, and only if, there are a schema S in $\mathcal{L}_{\mathbf{A},k}$ and a concrete-substitution Σ such that $\Sigma(S) = \varphi$.*

Proof. Clearly, if there is a schema S in $\mathcal{L}_{\mathbf{A},k}$ and a concrete-substitution Σ such that $\Sigma(S) = \varphi$, then $T \vdash_{k \text{ steps}} \varphi$ (this follows by a simple induction argument). Let us prove the other direction by induction on k. It is clear that the result holds for $k = 0$, the axiom case. Suppose, by induction hypothesis, that the result holds for all $s \leq k$. Suppose that $k + 1$ is a grounded number for $\mathbf{A}$ and consider φ such that $T \vdash_{k+1 \text{ steps}} \varphi$. We have two cases:

Last step uses Gen: In this case, we have that $\varphi = \forall x.\psi$ and $T \vdash_{k \text{ steps}} \psi$. It follows from the definition that k is a grounded number for $\mathbf{A}$. By induction hypothesis, there are a schema S in $\mathcal{L}_{\mathbf{A},k}$ and a concrete-substitution Σ such that $\Sigma(S) = \psi$. Consider $\forall v.S$ the schema that is obtained from S by the generalisation rule. Clearly, $\forall v.S$ is in $\mathcal{L}_{\mathbf{A},k+1}$. Furthermore, we can extend Σ in such a way that $\Sigma(\forall v.S) = \forall x.\psi = \varphi$.

Last step uses MP: In this case, there is a formula ψ such that $T \vdash_{k_0 \text{ steps}} \psi$ and $T \vdash_{k_1 \text{ steps}} \psi \to \varphi$, with $k = k_0 + k_1$. It follows that k_0 and k_1 are grounded numbers for $\mathbf{A}$. By induction hypothesis, there are schemata F in $\mathcal{L}_{\mathbf{A},k_0}$ and $G \to H$ in $\mathcal{L}_{\mathbf{A},k_1}$, and concrete-substitutions Σ_0 and Σ_1 such that $\Sigma_0(F) = \psi$ and $\Sigma_1(G \to H) = \psi \to \varphi$. By the reasoning of the proof of Theorem 3.3.6, we can guarantee the existence of a suitable substitution σ delivered by the algorithm such that $\sigma(F) = \sigma(G)$. As $k+1$ is a grounded number for $\mathbf{A}$, the algorithm must be successful and thus $\sigma(H)$ is in $\mathcal{L}_{\mathbf{A},k+1}$. As σ is minimal in the sense of introduction of unnecessary complexity and by the reasoning of the proof of Theorem 3.3.6 (the fact that σ was chosen using the concrete-substitutions), there must be a concrete-substitution Σ such that $\Sigma(\sigma(H)) = \varphi$.

The result follows by induction. One could also prove the result using the proof of Corollary 3.3.1. ⊣

Corollary 3.3.2. *Given k a grounded number for* $\mathbf{A}$*, it is decidable whether $T \vdash_{k\ steps} \varphi$.*

Proof. One considers the finite list $\mathscr{L}_{\mathbf{A},k}$. By the previous result, it is enough to see whether there is a concrete-substitution Σ and S in $\mathscr{L}_{\mathbf{A},k}$ such that $\Sigma(S) = \varphi$, something that is decidable from the fact that $\mathscr{L}_{\mathbf{A},k}$ is finite and from Theorem 3.3.2. This also follows from Theorem 3.3.4. ⊣

Corollary 3.3.3. *Given k a grounded number for* **A** *in* PA′, *it is decidable whether* $\mathsf{PA}' \vdash_{k\ steps} \varphi$.

Proof. Follows from the previous result when one has in mind that PA′ is one of the considered theories. ⊣

Corollary 3.3.4. *Given k a grounded number for* **A** *in* PA′, *it is decidable whether* $\mathsf{PA} \vdash_{k\ steps} \varphi$.

Proof. Follows from the previous Corollary and Theorem 3.2.3. ⊣

Theorem 3.3.8. *Given* **A**, *there is an* infinite number *of schemata in* $\mathscr{L}_{\mathbf{A}}$.

Proof. It is not hard to see that all schemata that are constructed using only the propositional logic axioms, L1–L3 in the initial list, are in $\mathscr{L}_{\mathbf{A}}$. Furthermore, besides these propositional schemata, there are much more schemata in $\mathscr{L}_{\mathbf{A}}$: the only restriction that ones has is that they do not yield the **problem** cases in their construction and the algorithm halts without rejecting. ⊣

Theorem 3.3.9. *Given an algorithm* **A**, *there is an algorithm* $\mathbf{H}_{\mathbf{A}}$ *such that, for every grounded proof-skeleton* $\mathcal{S}$ *and every formula* φ, *the algorithm halts and accepts for* $\mathcal{S}$ *and* φ *if, and only if,* φ *has a proof whose skeleton is* $\mathcal{S}$.

Proof. Fix an algorithm **A**. For a proof-skeleton $\mathcal{S}$, the algorithm $\mathbf{H}_{\mathbf{A}}$—using the construction of the schemata in $\mathscr{L}_{\mathbf{A}}$—tries to generate all the provable schemata in $\mathscr{L}_{\mathbf{A}}$ whose skeleton is $\mathcal{S}$. Observe that in the construction of the lists $\mathscr{L}_{\mathbf{A},k}$ one needed to make assumptions about the successfulness of the algorithm **A** for certain systems, but in this algorithm we do not make those (possibly non-computable) assumptions; the rest of the process remains the same, but one only focus on the construction of the schemata that potentially have skeleton $\mathcal{S}$. This yields no problem because the construction of the lists is computable with possible exception of the successfulness conditions. For non-grounded proof-skeletons (not in $\mathscr{L}_{\mathbf{A}}$) the algorithm might not halt. Suppose that $\mathcal{S}$ is a grounded proof-skeleton. Then, the algorithm $\mathbf{H}_{\mathbf{A}}$ can successfully construct the lists $\mathscr{R}$ from the proof of Corollary 3.3.1. Thus, using Theorem 3.3.2, the algorithm can decide whether φ has a proof whose skeleton is $\mathcal{S}$. ⊣

The algorithms $\mathbf{H}_{\mathbf{A}}$ have, in a sense, implemented the idea of the analysis made to the skeleton MP([L2)],MP([L2)],[L1)])) in the beginning of Section 3.3.1.

Theorem 3.3.10. *If* T *is such that* $T \vdash_{k\ steps} \varphi$ *is uniformly decidable in* k, *then there is a recursive function* $f(k,\varphi)$ *such that*

$$T \vdash_{k\ steps} \varphi \implies T \vdash_{f(k,\varphi)\ symbols} \varphi.$$

Proof. Assume $T \vdash_{k \text{ steps}} \varphi$ is uniformly decidable in k. Consider the (partial) recursive functions

$$c(k,\varphi) := \mu n[n \text{ is the code of a proof of } \varphi \text{ in } T \text{ with at most } k \text{ steps}],$$

and

$$sym(s) := \begin{cases} \text{number of symbols in} & \\ \text{the proof whose code is } s, & s \text{ is the code of a proof in } T \\ 0, & \text{otherwise.} \end{cases}$$

By hypothesis, we can decide uniformly in k if $T \vdash_{k \text{ steps}} \varphi$ holds or not. Thus, the function

$$f(k,\varphi) := \begin{cases} sym(c(k,\varphi)), & T \vdash_{k \text{ steps}} \varphi \\ 0, & \text{otherwise.} \end{cases}$$

is, by construction, a total recursive-function that satisfies the desired property. ⊣

Theorem 3.3.11. *If T is such that there is a recursive function $f(k,\varphi)$ such that*

$$T \vdash_{k\ steps} \varphi \implies T \vdash_{k\ steps\ and\ f(k,\varphi)\ symbols} \varphi,$$

then $T \vdash_{k\ steps} \varphi$ is uniformly decidable in k.

Proof. Assume there is a recursive-function $f(k,\varphi)$ satisfying the considered property. Let us consider the following algorithm.

1. Input: k and φ.

2. Compute $f(k,\varphi)$. If there is a proof of φ in k steps, then there is a proof of φ with at most $f(k,\varphi)$ symbols; such a proof would use at most $f(k,\varphi)$ variables, furthermore it does not matter the choice of variables that one makes in the sense that if one changes all the occurrences of a given variable in the proof one continues to have a sound proof. Take a finite list of at most $f(k,\varphi)$ variables.

3. Consider a finite list $\mathcal{I}_0$ of symbols consisting of: the logical symbols ('$\forall$', '$\neg$', and '$\rightarrow$'), '=', '(', ')', 'S', '+', '$\times$', '0' and the previously mentioned finite list of variables. Consider also a blank symbol '**B**' (just to separate the candidate formulas in a proof to be) not in $\mathcal{I}_0$.

4. Using only symbols from $\mathcal{I}_0$ and '**B**', generate a list $\mathcal{I}$ of all the (finitely many) possible lists of symbols which contain at most $f(k,\varphi)$ ones from $\mathcal{I}_0$ that have φ as the last element of the list.

5. Test if any element of $\mathcal{I}$ is a proof in T with k steps (clearly, this can be done in a computational manner): output 1 in affirmative case, and 0 in the negative case.

It is not hard to see that the previous algorithm decides uniformly in k the relation $T \vdash_{k \text{ steps}} \varphi$. ⊣

Consider PA^a as being any formulation of PA considered in [31] and proved to have a decidable k-provability. The next result is a solution to Problem 20 of [18] for these formulations (and not in general), a problem proposed by Krajíček.

Corollary 3.3.5. *There is a recursive-function $f(k, \varphi)$ such that*

$$\mathsf{PA}^a \vdash_{k\ steps} \varphi \implies \mathsf{PA}^a \vdash_{f(k,\varphi)\ symbols} \varphi.$$

Proof. Follows from Theorem 3.3.10 and the fact that $\mathsf{PA}^a \vdash_{k \text{ steps}} \varphi$ was proved to be decidable uniformly in k (see [31]). ⊣

Chapter 4

Montagna's Problem and Kreisel's Conjecture

4.1 Preliminaries

For a (partial) recursive function h, the notation $T \vdash_{\leq h} \varphi$ expresses that $h(\#\varphi)$ is defined and φ is provable in T with a proof whose code is at most $h(\#\varphi)$. This notion generalizes the approach followed in [65, p. 33–35]. Clearly, $\vdash_{\leq h}$ depends heavily on the chosen Gödel-numbering: different codings give rise to different notions. For the rest of this chapter, the concrete Gödel-numbering is assumed to be a fixed one; moreover, we assume that T is a r.e. extension of PA.

Given T, the theory K_T extends T by the following axiom schema:

Axiom K. *If f is a total recursive function such that, for all $n \in \mathbb{N}$, $f(n) \neq 0$, and $\mathsf{R}(x,y)$ is a formula that strongly-represents f in T, then $K_T \vdash \forall x.\neg\mathsf{R}(x,0)$.*

This schema can be restricted to a smaller class of functions in such a way that K_T might be recursively enumerable.

4.2 Introduction

According to [35], Kreisel's Conjecture is the statement:

If, for all $n \in \mathbb{N}$, $\mathsf{PA} \vdash_{k \text{ steps}} \varphi(\overline{n})$, then $\mathsf{PA} \vdash \forall x.\varphi(x)$. **[Kreisel's Conjecture]**

Kreisel's Conjecture has been studied for different system, with partial solutions for specific theories of arithmetic besides PA—see, for instance, [73], [68], and [1]; for a detailed account on the Conjecture we refer to [15].

In this chapter, we present results similar to Kreisel's Conjecture for $\vdash_{\leq h}$, which are not restricted to PA. Furthermore, we present general results concerning notions of provability and study the following problem proposed by Montagna in [18, p. 9]:

Does $\mathsf{PA} \vdash_{k \text{ steps}} \mathsf{Pr}_{\mathsf{PA}}(\ulcorner \varphi \urcorner) \rightarrow \varphi$ imply $\mathsf{PA} \vdash_{k \text{ steps}} \varphi$? **[Montagna's Problem]**

Besides studying the original problem, we also study the following variant of this question:

Does $T \vdash_{\leq h} \mathsf{Pr}_T(\ulcorner \mathsf{Pr}_T(\ulcorner \varphi \urcorner) \rightarrow \varphi \urcorner)$ imply $T \vdash_{\leq h} \varphi$? [Adapted Montagna's Problem]

A part of this chapter gave rise to the following paper: [86].

It is an interesting features of our approach on Kreisel's Conjecture that it does not depend so heavily on the particular axiomatization of T that one chooses. In some sense, it can be seen as a uniform approach, since it applies to any consistent theory T that is a recursively enumerable extension of PA.

For h a total recursive function, the *adapted Kreisel's conjecture for* $\vdash_{\leq h}$ is:

If, for all $n \in \mathbb{N}$, $T \vdash_{\leq h} \varphi(\overline{n})$, then $T \vdash \forall x.\varphi(x)$. [Adapted Kreisel's Conjecture]

Theorem 4.2.1. *The adapted Kreisel's Conjecture for* $\vdash_{\leq h}$ *is false.*

Proof. Let $\mathsf{Prf}_T(x, y)$ be as in Chapter 2. Let h be the function defined by:

$$h(m) := \begin{cases} \mu k[k \text{ is the code of a proof of the formula coded by } m \text{ in } T], \\ \qquad\qquad\qquad\qquad \text{if } m = \neg \mathsf{Prf}_T(\overline{n}, \ulcorner \bot \urcorner), \text{for some } n, \\ 0, \qquad\qquad\qquad\quad \text{otherwise.} \end{cases}$$

where μ denotes the minimization function (see [91, p. 833] for further details on minimization). It is clear that h is a (total) recursive function. By construction, for all $n \in \mathbb{N}$, $T \vdash_{\leq h} \neg \mathsf{Prf}_T(\overline{n}, \ulcorner \bot \urcorner)$. If the adapted Kreisel's Conjecture for $\vdash_{\leq h}$ was true, it would follow $T \vdash \forall x. \neg \mathsf{Prf}_T(x, \ulcorner \bot \urcorner)$, contradicting the Second Incompleteness Theorem (see [91, p. 828]). ⊣

It is not known whether the previous result still holds if one restricts oneself to primitive-recursive functions (or any other proper class of the recursive functions).

The result is in accordance with [74], where several reasons are given to believe that Kreisel's Conjecture is, in fact, false.

Even though the adapted Kreisel's Conjecture for $\vdash_{\leq h}$ is false, it is worth studying variants and weakenings of it. For example, one could ask for an extension T^h of T such that Kreisel's Conjecture holds adapted to T^h: given a total recursive function h, if, for all $n \in \mathbb{N}$, $T \vdash_{\leq h} \varphi(\overline{n})$, then $T^h \vdash \forall x.\varphi(x)$.

One immediate solution would be to add the true sentence $\forall x.\varphi(x)$ as an axiom to T. We, however, construct a theory T^h, avoiding the trivial *a priori* addition of $\forall x.\varphi(x)$ as an axiom. The approach is of interest, since it allows to establish relations between different concepts.

We also study versions of the conjecture for theories that satisfy certain derivability conditions and, in a sense, we parametrize variants of the conjecture to the satisfiability of the derivability conditions. We exhibit conditions for a theory to satisfy the following implication:

$$T \vdash \forall x.\mathsf{P}^h_T(\ulcorner \varphi(\dot{x}) \urcorner) \implies T \vdash \forall x.\varphi(x).$$

This corresponds to an arithmetization of Kreisel's Conjecture, where $\mathsf{P}^h_T(\cdot)$ represents $\vdash_{\leq h} \cdot$ inside T.

Finally, we prove, for certain theories, the existence of a total recursive function h such that $\vdash_{k \text{ steps}} \subseteq \vdash_{\leq h}$.

4.3 General Facts on Bounded Notions of Provability

In this section, we present two general results that express limitations of the use of notions of provability. We say that τ is a numeration of the axioms of T if τ identifies all the axioms of T; we use Pr_τ to denote the standard provability predicate (for τ). Pr_τ plays the role of Pr_T, but here we decided to emphasize the considered numeration τ (see Chapter 5 for details).

Theorem 4.3.1. *Let T be such that there are $R_0(x,y)$ a $\Sigma_1(T)$-formula, $R_1(x,y)$ a $\Pi_1(T)$-formula, and τ a numeration of the axioms of T that satisfy:*

1. *For all formulas φ, $T \vdash \mathsf{Pr}_\tau(\ulcorner \varphi \urcorner) \rightarrow \exists x.R_0(\ulcorner \varphi \urcorner, x)$;*

2. *For all formulas φ, $T \vdash \neg\mathsf{Pr}_\tau(\ulcorner \varphi \urcorner) \rightarrow \forall x.\neg R_1(\ulcorner \varphi \urcorner, x)$.*

Then, there is φ such that $T \vdash \exists x.(R_0(\ulcorner \mathsf{Pr}_\tau(\ulcorner \varphi \urcorner) \urcorner, x) \wedge \neg R_1(\ulcorner \varphi \urcorner, x))$.

Proof. Consider φ given by the Diagonalization Lemma [65, p. 15] such that

$$T \vdash \varphi \leftrightarrow \exists x.(R_0(\ulcorner \mathsf{Pr}_\tau(\ulcorner \varphi \urcorner) \urcorner, x) \wedge \neg R_1(\ulcorner \varphi \urcorner, x)).$$

It is clear that φ is a $\Sigma_1(T)$-formula, so[1] $T \vdash \varphi \rightarrow \mathsf{Pr}_\tau(\ulcorner \varphi \urcorner)$. Furthermore, from 1 and 2, $T \vdash \mathsf{Pr}_\tau(\ulcorner \mathsf{Pr}_\tau(\ulcorner \varphi \urcorner) \urcorner) \wedge \neg\mathsf{Pr}_\tau(\ulcorner \varphi \urcorner) \rightarrow \varphi$, and so $T \vdash \mathsf{Pr}_\tau(\ulcorner \mathsf{Pr}_\tau(\ulcorner \varphi \urcorner) \urcorner) \wedge \neg\varphi \rightarrow \mathsf{Pr}_\tau(\ulcorner \varphi \urcorner)$; this fact together with what was concluded yields that $T \vdash \mathsf{Pr}_\tau(\ulcorner \mathsf{Pr}_\tau(\ulcorner \varphi \urcorner) \urcorner) \rightarrow \mathsf{Pr}_\tau(\ulcorner \varphi \urcorner)$. By Löb's Theorem, $T \vdash \mathsf{Pr}_\tau(\ulcorner \varphi \urcorner)$, and so $T \vdash \varphi$, as wanted. ⊣

From the previous result, we conclude that if R_0 and R_1 are notions of provability with the feature that $T \vdash \mathsf{Pr}_\tau(\ulcorner \varphi \urcorner) \leftrightarrow \exists x.R_i(\ulcorner \varphi \urcorner, x)$, with $i = 0, 1$, then one can guarantee the existence of φ such that $T \vdash \exists x.(R_0(\ulcorner \mathsf{Pr}_\tau(\ulcorner \varphi \urcorner) \urcorner, x) \wedge \neg R_1(\ulcorner \varphi \urcorner, x))$—this formula expresses that, for some parameter n, it is the case that $\mathbb{N} \models R_0(\ulcorner \mathsf{Pr}_\tau(\ulcorner \varphi \urcorner) \urcorner, \overline{n})$, but $\mathbb{N} \models \neg R_1(\ulcorner \varphi \urcorner, \overline{n})$.

[1]This step uses the internal Σ_1-completeness of T.

In a more intuitive way, if one can uniformly represent two notions of provability in T via standard provability predicates, then, for some parameter n, $\mathsf{Pr}_\tau(\ulcorner\varphi\urcorner)$ will hold for the first notion of provability with parameter n, but φ will not hold for the second; thus, either the notions of provability are not representable in terms of standard provability predicates, or one has the non-intuitive fact that $\mathsf{Pr}_\tau(\ulcorner\varphi\urcorner)$ is "simpler" than φ, for the considered notions of provability; this notions can be, for instance, the syntactical representation of k-provability (see Chapter 3 for further details on this notion of provability).

The next results are about limitations on the uniformity of theories with respect to certain notions of provability.

Corollary 4.3.1. *If there is a $\Delta_1(T)$-formula $L(x,y)$ such that*[2]*:*

1. *For all formulas φ and all k, $\mathbb{N} \models L(\ulcorner\varphi\urcorner,\overline{k})$ if, and only if, $T \vdash_{k \text{ steps (symbols)}} \varphi$;*
2. *There is a numeration τ of the axioms of T such that , for all formulas φ, $T \vdash \mathsf{Pr}_\tau(\ulcorner\varphi\urcorner) \leftrightarrow \exists x.L(\ulcorner\varphi\urcorner,x)$.*

Then, there is a formula φ and k such that $T \vdash_{k \text{ steps (symbols)}} \mathsf{Pr}_\tau(\ulcorner\varphi\urcorner)$, but $T \nvdash_{k \text{ steps (symbols)}} \varphi$.

Proof. From Theorem 4.3.1 by considering $R_0(x,y), R_1(x,y) := L(x,y)$ we can guarantee the existence of φ such that $T \vdash \exists x.(L(\ulcorner\mathsf{Pr}_\tau(\ulcorner\varphi\urcorner)\urcorner,x) \wedge \neg L(\ulcorner\varphi\urcorner,x))$. So, there is k such that $\mathbb{N} \models L(\ulcorner\mathsf{Pr}_\tau(\ulcorner\varphi\urcorner)\urcorner,\overline{k}) \wedge \neg L(\ulcorner\varphi\urcorner,\overline{k})$, i.e. such that $T \vdash_{k \text{ steps (symbols)}} \mathsf{Pr}_\tau(\ulcorner\varphi\urcorner)$, but $T \nvdash_{k \text{ steps (symbols)}} \varphi$. ⊣

Observe that such a formula $L(x,y)$ in the conditions of the previous corollary constitutes a notion of provability. The previous result states that if one can represent, in a decidable way, the relation $T \vdash_{k \text{ steps}} \varphi$ inside T, then one must have the counter-intuitive fact $T \vdash_{k \text{ steps}} \mathsf{Pr}_\tau(\ulcorner\varphi\urcorner)$, but $T \nvdash_{k \text{ steps}} \varphi$, for certain k and φ.

The next result shows that if k-provability is decidable in T, then T cannot have a stronger version of Löb's rule for k-provability.

Theorem 4.3.2. *Suppose that k-provability can be expressed in T using a $\Delta_1(T)$-formula $L(x,y)$, i.e. $T \vdash L(\ulcorner\varphi\urcorner,\overline{k}) \iff T \vdash_{k \text{ steps}} \varphi$. Then, the rule*

$$\frac{L(\ulcorner\varphi\urcorner,\overline{k}) \to \varphi}{\varphi}$$

is not valid in T.

Proof. Consider any k. Using the Diagonalization Lemma [90, p. 169], one can guarantee the existence of a formula φ such that

$$T \vdash \varphi \leftrightarrow L(\ulcorner\varphi\urcorner,\overline{k}) \wedge \neg L(\ulcorner\varphi\urcorner,\overline{k}).$$

[2]The notation $\vdash_{k \text{ steps (symbols)}}$ is used to mention either $\vdash_{k \text{ steps}}$ or $\vdash_{k \text{ symbols}}$.

Clearly, φ is false, and so $\neg L(\ulcorner\varphi\urcorner,\overline{k})$ is true, but as $\neg L(\ulcorner\varphi\urcorner,\overline{k})$ is a $\Sigma_1(T)$-formula, from the Σ_1-completeness of T, one concludes that $T \vdash \neg L(\ulcorner\varphi\urcorner,\overline{k})$. Hence, $T \vdash \varphi \leftrightarrow L(\ulcorner\varphi\urcorner,\overline{k})$, in particular $T \vdash L(\ulcorner\varphi\urcorner,\overline{k}) \to \varphi$. If the rule was valid, it would follows that $T \vdash \varphi$, i.e. $T \vdash \perp$, that would be a contradiction. ⊣

4.4 On the notion $\vdash_{\leq h}$

In this section, we study the notion $\vdash_{\leq h}$ and some of its properties. We start with a result that guarantees that $\vdash_{\leq h}$ is representable in T.

Theorem 4.4.1. *Given a total recursive function h, there is $\mathsf{P}^h_T(\cdot)$ that represents $\vdash_{\leq h} \cdot$ in T such that if, for all $n \in \mathbb{N}$, $T \vdash_{\leq h} \alpha(\overline{n})$, then $K_T \vdash \forall x.\mathsf{P}^h_T(\ulcorner\alpha(\dot{x})\urcorner)$.*

Proof. Let h be an arbitrary, but fixed total recursive function. We define f_h by[3]:

$$f_h(n) := \begin{cases} \mu m \leq h(n)[m \text{ is the code of a proof of the formula coded by } n], \\ \qquad \text{if } n \text{ is a code of a formula and there is such an } m, \\ 0, \quad \text{otherwise.} \end{cases}$$

f_h is a total recursive function, thus f_h can be strongly-representable[4] by a formula $\mathsf{R}_h(x,y)$ in T. Given $n,m \in \mathbb{N}$, it is clear that $m \leq h(n)$ is the smallest code of a proof of the formula coded by n if, and only if, $T \vdash \mathsf{R}_h(\overline{n},\overline{m}) \wedge \overline{m} \neq 0$. Thus, we can define[5] $\mathsf{P}^h_T(x) := \exists y \neq 0.\mathsf{R}_h(x,y)$.

Assume that, for all $n \in \mathbb{N}$, $T \vdash_{\leq h} \alpha(\overline{n})$. Let g_h be the function defined by $g_h(n) := f_h(\#\alpha(\overline{n}))$. g_h is a total recursive function. Furthermore, g_h is strongly-representable by the formula $\mathsf{S}_h(x,y) := \mathsf{R}_h(\ulcorner\alpha(\dot{x})\urcorner,y)$ since:

(i) If $g_h(n) = m$, then $f_h(\#\alpha(\overline{n})) = m$, and thus $T \vdash \mathsf{R}_h(\ulcorner\alpha(\overline{n})\urcorner,\overline{m})$, i.e., $T \vdash \mathsf{S}_h(\overline{n},\overline{m})$;

(ii) As $T \vdash \forall x.\exists! y.\mathsf{R}_h(x,y)$ it follows that $T \vdash \forall x.\exists! y.\mathsf{R}_h(\ulcorner\alpha(\dot{x})\urcorner,y)$, i.e., $T \vdash \forall x.\exists! y.\mathsf{S}_h(x,y)$.

By hypothesis, for all $n \in \mathbb{N}$, there is $m \leq h(\#\alpha(\overline{n}))$ such that m is the code of a proof of $\alpha(\overline{n})$ in T. Hence, for all $n \in \mathbb{N}$, $g_h(n) \neq 0$. As $\mathsf{S}_h(x,y)$ strongly-represents g_h, we have by hypothesis that $K_T \vdash \forall x.\neg\mathsf{S}_h(x,0)$. From $T \vdash \forall x.\exists! y.\mathsf{S}_h(x,y)$ follows that $T \vdash \forall x.\exists y.\mathsf{S}_h(x,y)$. Together with $K_T \vdash \forall x.\neg\mathsf{S}_h(x,0)$, it follows that $K_T \vdash \forall x.\exists y \neq 0.\mathsf{S}_h(x,y)$, i.e., $K_T \vdash \forall x.\exists y \neq 0.$ $\mathsf{R}_h(\ulcorner\alpha(\dot{x})\urcorner,y)$. So, $K_T \vdash \forall x.\mathsf{P}^h_T(\ulcorner\alpha(\dot{x})\urcorner)$.

We now show that, for all formulas φ, $T \vdash_{\leq h} \varphi \iff T \vdash \mathsf{P}^h_T(\ulcorner\varphi\urcorner)$. Suppose that $T \vdash_{\leq h} \varphi$. For $m := f_h(\#\varphi)$, we have that $m \neq 0$. Thus, $T \vdash \mathsf{R}_h(\ulcorner\varphi\urcorner,\overline{m}) \wedge \overline{m} \neq 0$, and so $T \vdash \exists y \neq 0.\mathsf{R}_h(\ulcorner\varphi\urcorner,y)$. Hence, $T \vdash \mathsf{P}^h_T(\ulcorner\varphi\urcorner)$. Now suppose that $T \vdash \mathsf{P}^h_T(\ulcorner\varphi\urcorner)$. Let $m := f_h(\#\varphi)$. If $m \neq 0$, then $T \vdash_{\leq h} \varphi$. Suppose, towards a contradiction, that $m = 0$. We

[3]We are assuming that the Gödel-numbers are defined in such a way that 0 is never the code of a proof.

[4]See Theorem 2.1.1.

[5]Keep in mind that P^h_T is very distinct from Pr_T.

have that $T \vdash \exists y \neq 0.\mathsf{R}_h(\ulcorner\varphi\urcorner, y)$ and $T \vdash \mathsf{R}_h(\ulcorner\varphi\urcorner, 0)$. As $T \vdash \forall x.\exists! y.\mathsf{R}_h(x,y)$ we arrive at a contradiction. So, $m \neq 0$, as desired. ⊣

The next result follows immediately from the proof of the previous result.

Corollary 4.4.1. *Given a total recursive function h and a formula φ, we have that $T \vdash_{\leq h} \varphi \iff T \vdash \mathsf{P}^h_T(\ulcorner\varphi\urcorner)$.*

P^h_T is provably decidable, in the following sense:

Theorem 4.4.2. *Given a total recursive function h, for every formula φ, we have that $T \vdash \mathsf{P}^h_T(\ulcorner\varphi\urcorner)$ or $T \vdash \neg\mathsf{P}^h_T(\ulcorner\varphi\urcorner)$.*

Proof. Suppose that $T \nvdash \mathsf{P}^h_T(\ulcorner\varphi\urcorner)$. By the previous result, $T \nvdash_{\leq h} \varphi$. This means that $f_h(\#\varphi) = 0$. As $\mathsf{R}_h(x,y)$ strongly-represents the function f_h, it follows that $T \vdash \mathsf{R}_h(\ulcorner\varphi\urcorner, 0)$. Since $T \vdash \forall x.\exists! y.\mathsf{R}_h(x,y)$, we can conclude that $T \vdash \forall y.(\mathsf{R}_h(\ulcorner\varphi\urcorner, y) \to y = 0)$, and so, $T \vdash \neg\exists y.(y \neq 0 \wedge \mathsf{R}_h(\ulcorner\varphi\urcorner, y))$, i.e., $T \vdash \neg\mathsf{P}^h_T(\ulcorner\varphi\urcorner)$. ⊣

The next result corresponds to an arithmetization of the previous statement.

Theorem 4.4.3. *Given a provability predicate $P(x)$ and a total recursive function h, we have for every formula φ that $T \vdash P(\ulcorner\mathsf{P}^h_T(\ulcorner\varphi\urcorner)\urcorner) \vee P(\ulcorner\neg\mathsf{P}^h_T(\ulcorner\varphi\urcorner)\urcorner)$.*

Proof. From Theorem 4.4.2, it follows $T \vdash \mathsf{P}^h_T(\ulcorner\varphi\urcorner)$ or $T \vdash \neg\mathsf{P}^h_T(\ulcorner\varphi\urcorner)$. If $T \vdash \mathsf{P}^h_T(\ulcorner\varphi\urcorner)$, then $T \vdash P(\ulcorner\mathsf{P}^h_T(\ulcorner\varphi\urcorner)\urcorner)$, and so $T \vdash P(\ulcorner\mathsf{P}^h_T(\ulcorner\varphi\urcorner)\urcorner) \vee P(\ulcorner\neg\mathsf{P}^h_T(\ulcorner\varphi\urcorner)\urcorner)$. If $T \vdash \neg\mathsf{P}^h_T(\ulcorner\varphi\urcorner)$, then $T \vdash P(\ulcorner\neg\mathsf{P}^h_T(\ulcorner\varphi\urcorner)\urcorner)$, hence $T \vdash P(\ulcorner\mathsf{P}^h_T(\ulcorner\varphi\urcorner)\urcorner) \vee P(\ulcorner\neg\mathsf{P}^h_T(\ulcorner\varphi\urcorner)\urcorner)$. In sum, $T \vdash P(\ulcorner\mathsf{P}^h_T(\ulcorner\varphi\urcorner)\urcorner) \vee P(\ulcorner\neg\mathsf{P}^h_T(\ulcorner\varphi\urcorner)\urcorner)$. ⊣

We now prove that there is no single h such that $\vdash_{\leq h}$ coincides with $\vdash$.

Theorem 4.4.4. *For every total recursive function h, there is a formula φ such that $T \vdash \varphi$, but $T \nvdash_{\leq h} \varphi$.*

Proof. Let h be a fixed total recursive function. Let φ be the sentence obtained from the application of the Diagonalization Lemma to the formula $\neg\mathsf{P}^h_T(x)$. Then,

$$T \vdash \varphi \leftrightarrow \neg\mathsf{P}^h_T(\ulcorner\varphi\urcorner). \tag{I}$$

Suppose, towards a contradiction, that $T \vdash_{\leq h} \varphi$. So, $T \vdash \varphi$. By Corollary 4.4.1 we have that $T \vdash \mathsf{P}^h_T(\ulcorner\varphi\urcorner)$, and so, by (I), $T \vdash \neg\varphi$, which contradicts $T \vdash \varphi$. Hence, $T \nvdash_{\leq h} \varphi$. From Theorem 4.4.2 it follows that $T \vdash \neg\mathsf{P}^h_T(\ulcorner\varphi\urcorner)$, i.e., $T \vdash \varphi$. ⊣

The next fact will play a major role in the discussion of Kreisel's Conjecture, and it is similar to a reflection principle.

Theorem 4.4.5. *Given a total recursive function h, for every formula φ, $T \vdash \mathsf{P}^h_T(\ulcorner\varphi\urcorner) \to \varphi$.*

Proof. Let φ be an arbitrary formula. From Theorem 4.4.2, $T \vdash \mathsf{P}^h_T(\ulcorner\varphi\urcorner)$ or $T \vdash \neg\mathsf{P}^h_T(\ulcorner\varphi\urcorner)$. Firstly, suppose that $T \vdash \mathsf{P}^h_T(\ulcorner\varphi\urcorner)$. Then, by Corollary 4.4.1, we conclude $T \vdash_{\leq h} \varphi$, from where we get $T \vdash \varphi$. Thus, $T \vdash \mathsf{P}^h_T(\ulcorner\varphi\urcorner) \to \varphi$. Secondly, suppose that $T \vdash \neg\mathsf{P}^h_T(\ulcorner\varphi\urcorner)$. Then, by logic, $T \vdash \mathsf{P}^h_T(\ulcorner\varphi\urcorner) \to \varphi$. In all, $T \vdash \mathsf{P}^h_T(\ulcorner\varphi\urcorner) \to \varphi$. ⊣

Theorem 4.4.6. *Let h be a primitive-recursive function and $P(x)$ be a provability predicate such that:*

C1: *For all $\Sigma_1(T)$-formulas φ, $T \vdash \varphi \to P(\ulcorner\varphi\urcorner)$;*

C2: *For all formulas φ and ψ, $T \vdash \varphi \leftrightarrow \psi \implies T \vdash P(\ulcorner\varphi\urcorner) \leftrightarrow P(\ulcorner\psi\urcorner)$.*

Then, for every formula φ, $T \vdash \neg P(\ulcorner\neg\mathsf{P}^h_T(\ulcorner\varphi\urcorner)\urcorner) \to \mathsf{P}^h_T(\ulcorner\varphi\urcorner)$.

Proof. If h is primitive-recursive, then $\mathsf{R}_h(x,y)$ can be picked as being a $\Sigma_1(T)$-formula. Clearly, $T \vdash \neg\mathsf{P}^h_T(\ulcorner\varphi\urcorner) \leftrightarrow \mathsf{R}_h(\ulcorner\varphi\urcorner, 0)$. From **C2**, $T \vdash P(\ulcorner\neg\mathsf{P}^h_T(\ulcorner\varphi\urcorner)\urcorner) \leftrightarrow P(\ulcorner\mathsf{R}_h(\ulcorner\varphi\urcorner, 0)\urcorner)$. From **C1**, $T \vdash \mathsf{R}_h(\ulcorner\varphi\urcorner, 0) \to P(\ulcorner\mathsf{R}_h(\ulcorner\varphi\urcorner, 0)\urcorner)$, so $T \vdash \neg\mathsf{P}^h_T(\ulcorner\varphi\urcorner) \to P(\ulcorner\neg\mathsf{P}^h_T(\ulcorner\varphi\urcorner)\urcorner)$, as wanted. ⊣

Theorem 4.4.7. *Given h a primitive-recursive function, for every formula φ, $T + \mathsf{Con}_T \vdash \mathsf{Pr}_T(\ulcorner\mathsf{P}^h_T(\ulcorner\varphi\urcorner)\urcorner) \to \mathsf{P}^h_T(\ulcorner\varphi\urcorner)$.*

Proof. It is clear that $T \vdash \mathsf{Pr}_T(\ulcorner\mathsf{P}^h_T(\ulcorner\varphi\urcorner)\urcorner) \wedge \mathsf{Pr}_T(\ulcorner\neg\mathsf{P}^h_T(\ulcorner\varphi\urcorner)\urcorner) \to \mathsf{Pr}_T(\ulcorner\bot\urcorner)$. Thus, $T + \mathsf{Con}_T \vdash \mathsf{Pr}_T(\ulcorner\mathsf{P}^h_T(\ulcorner\varphi\urcorner)\urcorner) \wedge \mathsf{Pr}_T(\ulcorner\neg\mathsf{P}^h_T(\ulcorner\varphi\urcorner)\urcorner) \to \bot$. Hence, $T + \mathsf{Con}_T \vdash \neg\mathsf{Pr}_T(\ulcorner\mathsf{P}^h_T(\ulcorner\varphi\urcorner)\urcorner) \vee \neg\mathsf{Pr}_T(\ulcorner\neg\mathsf{P}^h_T(\ulcorner\varphi\urcorner)\urcorner)$, i.e., $T + \mathsf{Con}_T \vdash \mathsf{Pr}_T(\ulcorner\mathsf{P}^h_T(\ulcorner\varphi\urcorner)\urcorner) \to \neg\mathsf{Pr}_T(\ulcorner\neg\mathsf{P}^h_T(\ulcorner\varphi\urcorner)\urcorner)$. By the previous result we conclude that $T + \mathsf{Con}_T \vdash \mathsf{Pr}_T(\ulcorner\mathsf{P}^h_T(\ulcorner\varphi\urcorner)\urcorner) \to \mathsf{P}^h_T(\ulcorner\varphi\urcorner)$. ⊣

4.5 Montagna's conjecture

Löb's Theorem [65, pp. 28, 29] expresses that, for all formulas φ, if $T \vdash \mathsf{Pr}_T(\ulcorner\varphi\urcorner) \to \varphi$, then $T \vdash \varphi$. More generally, for all formulas φ,

$$T \vdash \mathsf{Pr}_T(\ulcorner\mathsf{Pr}_T(\ulcorner\varphi\urcorner) \to \varphi\urcorner) \leftrightarrow \mathsf{Pr}_T(\ulcorner\varphi\urcorner).$$

If one analyzes the proof of Löb's Theorem, it gives the impression that one can prove $\mathsf{Pr}_T(\ulcorner\varphi\urcorner) \to \varphi$ only if one has already proved φ. It indicates, moreover, that $\mathsf{Pr}_T(\ulcorner\mathsf{Pr}_T(\ulcorner\varphi\urcorner) \to \varphi\urcorner)$ is only provable, if φ is established in the first place. This intuition can be related to the problem proposed by Montagna in [18, p. 9] mentioned in Section 4.2.

Theorem 4.5.1 (Negative Adapted Montagna's Problem). *For every primitive-recursive function $g(x,y)$ with $g(x,y) > y$, there are a sentence φ and a number $n_0 \in \mathbb{N}$ such that $T \vdash_{\leq h} \mathsf{Pr}_T(\ulcorner\mathsf{Pr}_T(\ulcorner\varphi\urcorner) \to \varphi\urcorner)$, but $T \nvdash_{\leq h} \varphi$, where $h := \lambda x. g(x, n_0)$.*

Proof. We follow closely the proof of Theorem 14 from [65, p. 34]. Let g be a function-symbol that represents the primitive-recursive function g. By the Diagonalization Lemma, there is a sentence φ such that

$$T \vdash \varphi \leftrightarrow \exists y.(\mathtt{Prf}_T(\ulcorner \mathtt{Pr}_T(\ulcorner \mathtt{Pr}_T(\ulcorner \varphi \urcorner) \to \varphi \urcorner)\urcorner, y) \\ \wedge \forall z \leq \mathtt{g}(\ulcorner \varphi \urcorner, y).\neg \mathtt{Prf}_T(\ulcorner \varphi \urcorner, z)).$$

By construction, $T + \mathtt{Pr}_T(\ulcorner \mathtt{Pr}_T(\ulcorner \mathtt{Pr}_T(\ulcorner \varphi \urcorner) \to \varphi \urcorner)\urcorner) + \neg \mathtt{Pr}_T(\ulcorner \varphi \urcorner) \vdash \varphi$. As T is Σ_1-complete, $T + \varphi \vdash \mathtt{Pr}_T(\ulcorner \varphi \urcorner)$. Thus, we can conclude that $T + \mathtt{Pr}_T(\ulcorner \mathtt{Pr}_T(\ulcorner \mathtt{Pr}_T(\ulcorner \varphi \urcorner) \to \varphi \urcorner)\urcorner) \vdash \mathtt{Pr}_T(\ulcorner \varphi \urcorner)$. By Löb's Theorem, it follows that $T \vdash \mathtt{Pr}_T(\ulcorner \mathtt{Pr}_T(\ulcorner \mathtt{Pr}_T(\ulcorner \varphi \urcorner) \to \varphi \urcorner)\urcorner) \leftrightarrow \mathtt{Pr}_T(\ulcorner \mathtt{Pr}_T(\ulcorner \varphi \urcorner)\urcorner)$. Hence, $T + \mathtt{Pr}_T(\ulcorner \mathtt{Pr}_T(\ulcorner \varphi \urcorner)\urcorner) \vdash \mathtt{Pr}_T(\ulcorner \varphi \urcorner)$. Again by Löb's Theorem, it follows that $T \vdash \mathtt{Pr}_T(\ulcorner \varphi \urcorner)$, and consequently $T \vdash \varphi$ and $T \vdash \mathtt{Pr}_T(\ulcorner \mathtt{Pr}_T(\ulcorner \varphi \urcorner) \to \varphi \urcorner)$. This means that φ is true. Let n_0 satisfy the true existential property of φ. Then, n_0 is the code of a proof of $\mathtt{Pr}_T(\ulcorner \mathtt{Pr}_T(\ulcorner \varphi \urcorner) \to \varphi \urcorner)$. By hypothesis on g, it follows that $n_0 < g(\#\mathtt{Pr}_T(\ulcorner \mathtt{Pr}_T(\ulcorner \varphi \urcorner) \to \varphi \urcorner), n_0) = h(\#\mathtt{Pr}_T(\ulcorner \mathtt{Pr}_T(\ulcorner \varphi \urcorner) \to \varphi \urcorner))$, *ergo* $T \vdash_{\leq h} \mathtt{Pr}_T(\ulcorner \mathtt{Pr}_T(\ulcorner \varphi \urcorner) \to \varphi \urcorner)$. From the fact that φ is true one can conclude that for all $z \leq g(\#\varphi, n_0)$, z is not the code of a proof of φ. This means that $T \not\vdash_{\leq h} \varphi$. ⊣

If a formula φ is provable in T, we define

$$\|\varphi\|_T := \mathtt{min}\{n | n \text{ is the code of a proof of } \varphi \text{ in } T\}.$$

Moreover, if φ and ψ are formulas, we stipulate that $\varphi <_T \psi$ if $T \vdash \varphi \wedge \psi$ and $\|\varphi\|_T < \|\psi\|_T$. The following result confirms that the mentioned intuition that a proof of $\mathtt{Pr}_T(\ulcorner \mathtt{Pr}_T(\ulcorner \varphi \urcorner) \to \varphi \urcorner)$ should encompass, in a way, a proof of φ fails.

Theorem 4.5.2. *There is a formula φ such that*

$$\mathtt{Pr}_T(\ulcorner \mathtt{Pr}_T(\ulcorner \varphi \urcorner) \to \varphi \urcorner) <_T \varphi.$$

Proof. By the Diagonalization Lemma, there is a sentence φ such that

$$T \vdash \varphi \leftrightarrow \exists y.(\mathtt{Prf}_T(\ulcorner \mathtt{Pr}_T(\ulcorner \mathtt{Pr}_T(\ulcorner \varphi \urcorner) \to \varphi \urcorner)\urcorner, y) \\ \wedge \forall z \leq y.\neg \mathtt{Prf}_T(\ulcorner \varphi \urcorner, z)).$$

Applying the same reasoning as in the previous proof, it follows that $T \vdash \mathtt{Pr}_T(\ulcorner \mathtt{Pr}_T(\ulcorner \varphi \urcorner) \to \varphi \urcorner) \wedge \varphi$; in particular φ is true. Take n_0 as being the natural number that is guaranteed to exist from the true formula φ. It is straightforward that $\|\mathtt{Pr}_T(\ulcorner \mathtt{Pr}_T(\ulcorner \varphi \urcorner) \to \varphi \urcorner)\|_T \leq n_0$. As φ is true, it follows that for all $z \leq n_0$, z is not the code of a proof of φ. Hence, $n_0 < \|\varphi\|_T$, and so $\|\mathtt{Pr}_T(\ulcorner \mathtt{Pr}_T(\ulcorner \varphi \urcorner) \to \varphi \urcorner)\|_T < \|\varphi\|_T$. In all, $\mathtt{Pr}_T(\ulcorner \mathtt{Pr}_T(\ulcorner \varphi \urcorner) \to \varphi \urcorner) <_T \varphi$. ⊣

Theorem 4.5.3 (Negative Montagna's Problem 1)**.** *There are an axiomatization of Peano arithmetic in a Hilbert-style system, PA_c, a formula φ of PA_c, and k such that $\mathsf{PA}_c \vdash_{k\ steps} \mathtt{Pr}_{\mathsf{PA}_c}(\ulcorner \varphi \urcorner) \to \varphi$, but $\mathsf{PA}_c \nvdash_{k\ steps} \varphi$.*

Proof. Consider the Hilbert-style axiomatization of Peano Arithmetic, PA_c (c stands for 'counter-example'), that is obtained from the usual one by replacing the usual axioms by the following axioms[6]:

A'1. $\varphi \to (\psi \to \psi)$;

A'2. $(\varphi \to (\psi \to \mu)) \to ((\varphi \to \psi) \to (\varphi \to \mu))$;

A'3. $(\neg\varphi \to \neg\psi) \to (\psi \to \varphi)$;

A'4. $(\varphi \to (\psi \to \mu)) \to (\psi \to (\varphi \to \mu))$.

Using the automated theorem prover ProverX[7], we were able to confirm that the previous set of axioms constitutes an alternative basis to the propositional axioms of the usual axiomatization[8]. We used a technical trick to implement the desired formulas, since ProverX works in first-order logic—we used `P(x)` to denote a meta-provability predicate, `imp(x,y)` to denote implication and `neg(x)` to denote negation[9]:

```
   formulas(assumptions).
P(x) & P(imp(x,y)) -> P(y). % Modus Ponens
P(imp(x,imp(y,y))). %A'1
P(imp(imp(x,imp(y,z)),imp(imp(x,y),imp(x,z)))). %A'2
P(imp(imp(neg(x),neg(y)),imp(y,x))). %A'3
P(imp(imp(x,imp(y,z)),imp(y,imp(x,z)))). %A'4
end_of_list.

   formulas(goals).
P(imp(x,imp(y,x))).%A1%The other axioms are already satisfied.
end_of_list.
```

[6]One could also substitute, for instance, by the following list

A"1. $\varphi \to (\psi \to \psi)$;

A"2. $(\varphi \to (\psi \to \mu)) \to ((\varphi \to \psi) \to (\varphi \to \mu))$;

A"3. $(\neg\varphi \to \neg\psi) \to (\psi \to \varphi)$;

A"4. $\varphi \to (\psi \to \varphi \wedge \psi)$;

A"5. $\varphi \wedge \psi \to \varphi$;

A"6. $\varphi \wedge \psi \to \psi$;

A"7. $(\varphi \to \psi) \to ((\mu \to \varphi) \to (\mu \to \psi))$.

[7]ProverX is an extension of the automated theorem prover Prover9, see `http://proverx.com` for details.

[8]It is a basis in the sense that they generate the same provable formulas and that they are mutually independent. This step is crucial; it is not enough to see that the new formula is simply provable in the old system, it needs to generate all the same propositional theorems. ProverX, in particular Mace4, was especially useful prove the independence of the axioms.

[9]It is not hard to find a proof without using ProverX.

Take, for example, $\varphi := (0 = 0 \to 0 = 0)$. We have that $\mathsf{Pr}_{\mathsf{PA}_c}(\ulcorner 0 = 0 \to 0 = 0 \urcorner) \to (0 = 0 \to 0 = 0)$ is an A'1 axiom, but $0 = 0 \to 0 = 0$ is not an axiom of PA_c. This means that $\mathsf{PA}_c \vdash_{0 \text{ steps}} \mathsf{Pr}_{\mathsf{PA}_c}(\ulcorner 0 = 0 \to 0 = 0 \urcorner) \to (0 = 0 \to 0 = 0)$, but $\mathsf{PA}_c \nvdash_{0 \text{ steps}} 0 = 0 \to 0 = 0$. ⊣

Observe that the provability predicate played no role in the previous proof, this means that we can immediately strengthen the former result.

Theorem 4.5.4 (Negative Montagna's Problem 2)**.** *There are an axiomatization of Peano Arithmetic in a Hilbert-style system,* PA_c*, a formula* φ *of* PA_c*, and k such that for all provability predicates P,* $\mathsf{PA}_c \vdash_{k\ steps} P(\ulcorner\varphi\urcorner) \to \varphi$*, but* $\mathsf{PA}_c \nvdash_{k\ steps} \varphi$*.*

Proof. The proof of this result is the same as the proof of the previous theorem when one has in mind the previously made observation. ⊣

For the two previous results we changed the axiomatization of the propositional part of the usual axiomatization—this constituted a surprise to us because usually, in proof theory, the axiomatization of the propositional part of first-order logic plays no role in results about the considered theory. Those results presented a single counter-example, the next results will present a uniform negative answer, in the sense that we will describe families of axiomatizations of PA for which Montagna problem has a negative answer.

Theorem 4.5.5 (Negative Montagna's Problem 3)**.** *Let* PA_c *be an axiomatization of Peano Arithmetic such that* $\mathsf{PA}_c \nvdash_{1\ step} \exists x.\exists y.\mathsf{Prf}_{\mathsf{PA}_c}(x, y)$ *and that has the two following schemata as axioms:*

Axiom 1: $(\varphi \to \neg\psi) \to (\psi \to \neg\varphi)$*;*

Axiom 2: $(\forall x.\varphi) \to \varphi^x_t$*, where t is substitutable for x in* φ*.*

Then, there are a formula φ *and k such that* $\mathsf{PA}_c \vdash_{k\ steps} \mathsf{Pr}_{\mathsf{PA}_c}(\ulcorner\varphi\urcorner) \to \varphi$*, but* $\mathsf{PA}_c \nvdash_{k\ steps} \varphi$*.*

Proof. Consider $\varphi := \exists x.\exists y.\mathsf{Prf}_{\mathsf{PA}_c}(x, y)$. It is clear that $\forall x.\neg\mathsf{Pr}_{\mathsf{PA}_c}(x) \to \neg\mathsf{Pr}_{\mathsf{PA}_c}(\ulcorner\varphi\urcorner)$ is an **Axiom2**. Furthermore, $(\forall x.\neg\mathsf{Pr}_{\mathsf{PA}_c}(x) \to \neg\mathsf{Pr}_{\mathsf{PA}_c}(\ulcorner\varphi\urcorner)) \to (\mathsf{Pr}_{\mathsf{PA}_c}(\ulcorner\varphi\urcorner) \to \neg\forall x.\neg\mathsf{Pr}_{\mathsf{PA}_c}(x))$ is an **Axiom 1**. Thus, we can conclude $\mathsf{PA}_c \vdash_{1 \text{ step}} \mathsf{Pr}_{\mathsf{PA}_c}(\ulcorner\varphi\urcorner) \to \neg\forall x.\neg\mathsf{Pr}_{\mathsf{PA}_c}(x)$, i.e. $\mathsf{PA}_c \vdash_{1 \text{ step}} \mathsf{Pr}_{\mathsf{PA}_c}(\ulcorner\varphi\urcorner) \to \varphi$. By hypothesis, $\mathsf{PA}_c \nvdash_{1 \text{ step}} \varphi$. ⊣

Theorem 4.5.6 (Negative Montagna's Problem 4)**.** *Let* PA_c *be an axiomatization of Peano Arithmetic such that* $\exists x.\exists y.\mathsf{Prf}_{\mathsf{PA}_c}(x, y)$ *is not an axiom, but such that it has the following axiom schema:*

Axiom 1: $\varphi^x_t \to \exists x.\varphi$*, where t is substitutable for x in* φ*.*

Then, there are a formula φ *and k such that* $\mathsf{PA}_c \vdash_{k\ steps} \mathsf{Pr}_{\mathsf{PA}_c}(\ulcorner\varphi\urcorner) \to \varphi$*, but* $\mathsf{PA}_c \nvdash_{k\ steps} \varphi$*.*

Proof. Similarly to what was done on the previous result, consider $\varphi := \exists x.\exists y.\mathsf{Prf}_{\mathsf{PA}_c}(x, y)$. In this case, $\mathsf{Pr}_{\mathsf{PA}_c}(\ulcorner\varphi\urcorner) \to \exists x.\mathsf{Pr}_{\mathsf{PA}_c}(x)$ is an **Axiom 1**. So, $\mathsf{PA}_c \vdash_{0 \text{ steps}} \mathsf{Pr}_{\mathsf{PA}_c}(\ulcorner\varphi\urcorner) \to \varphi$, but $\mathsf{PA}_c \nvdash_{0 \text{ steps}} \varphi$. ⊣

4.6 Variants of Kreisel's Conjecture

In this section, we present some partial results related to Kreisel's Conjecture, namely variants of the conjecture for provability predicates in the presence of different derivability conditions. (For the next result T does not need to be r.e.)

Theorem 4.6.1. *Let h be a primitive-recursive function and T be such that there is a provability predicate $P(x)$ satisfying:*

C1: *If $\varphi(x)$ is a $\Sigma_n(T)$-formula, then $T \vdash \varphi(x) \to P(\ulcorner \varphi(\dot{x}) \urcorner)$;*

C2: $T \vdash \mathsf{P}^h_T(\ulcorner \varphi(\dot{x}) \urcorner) \to P(\ulcorner \varphi(\dot{x}) \urcorner)$;

C3: *For all formulas $\varphi(x)$ and $\psi(x)$, $T \vdash P(\ulcorner \varphi(\dot{x}) \to \psi(\dot{x}) \urcorner) \to (P(\ulcorner \varphi(\dot{x}) \urcorner) \to P(\ulcorner \psi(\dot{x}) \urcorner))$.*

If $\varphi(x)$ is a $\Pi_n(T)$-formula such that $T \vdash \forall x.\mathsf{P}^h_T(\ulcorner \varphi(\dot{x}) \urcorner)$, then $T + \mathsf{Con}_P \vdash \forall x.\varphi(x)$.

Proof. As $\varphi(x)$ is $\Pi_n(T)$, by **C1**, we have $T \vdash \neg\varphi(x) \to P(\ulcorner \neg\varphi(\dot{x}) \urcorner)$. Thus, $T \vdash \exists x.\neg\varphi(x) \to \exists x.P(\ulcorner \neg\varphi(\dot{x}) \urcorner)$. By **C2**, $T \vdash \mathsf{P}^h_T(\ulcorner \varphi(\dot{x}) \urcorner) \to P(\ulcorner \varphi(\dot{x}) \urcorner)$. So, $T \vdash \forall x.\mathsf{P}^h_T(\ulcorner \varphi(\dot{x}) \urcorner) \to \forall x.P(\ulcorner \varphi(\dot{x}) \urcorner)$. So, $T + \forall x.\mathsf{P}^h_T(\ulcorner \varphi(\dot{x}) \urcorner) \wedge \exists x.\neg\varphi(x) \vdash \exists x.P(\ulcorner \neg\varphi(\dot{x}) \urcorner) \wedge \forall x.P(\ulcorner \varphi(\dot{x}) \urcorner)$. By **C3**, $T + \forall x.\mathsf{P}^h_T(\ulcorner \varphi(\dot{x}) \urcorner) \wedge \exists x.\neg\varphi(x) \vdash P(\ulcorner \bot \urcorner)$, i.e., $T + \mathsf{Con}_P \vdash \forall x.\mathsf{P}^h_T(\ulcorner \varphi(\dot{x}) \urcorner) \to \forall x.\varphi(x)$. ⊣

The condition $T \vdash \forall x.\mathsf{P}^h_T(\ulcorner \varphi(\dot{x}) \urcorner)$ corresponds to an arithmetization of the antecedent of a version of Kreisel's Conjecture. Thus, the result is weaker than Kreisel's Conjecture. If $T \vdash \mathsf{Con}_P$, then the previous result can be proved inside T. It is important to observe that for $n > 1$ and T including $\mathsf{Th}_{\Sigma_n}(\mathbb{N})$ (the set of the Σ_n-sentences that are true in $\mathbb{N}$), Proposition 2.4 of [55] shows that condition **C1** can, in fact, be satisfied; despite this fact, we are especially interested in the Σ_1-case, due to its connection to r.e. theories.

The next result is a particular case of the previous theorem.

Corollary 4.6.1. *Let h be a primitive-recursive function. If $\varphi(x)$ is a $\Pi_1(\mathsf{PA})$-formula such that $T \vdash \forall x.P^h_{\mathsf{PA}}(\ulcorner \varphi(\dot{x}) \urcorner)$, then $\mathsf{PA} + \mathsf{Con}_{\mathsf{PA}} \vdash \forall x.\varphi(x)$.*

Proof. The Corollary follows immediately from the fact that $\mathsf{Pr}_{\mathsf{PA}}$ satisfies **C1** and **C2** of the previous Theorem [62]. ⊣

Theorem 4.6.2. *Let h be a primitive-recursive function and T be such that there is a provability predicate $P(x)$ satisfying:*

C1: *For all formulas $\varphi(x)$, $T \vdash P(\ulcorner \neg\varphi(\dot{x}) \urcorner) \to P(\ulcorner \neg \mathsf{P}^h_T(\ulcorner \varphi(\dot{\dot{x}}) \urcorner) \urcorner)$;*

C2: *If $\varphi(x)$ is a $\Sigma_n(T)$-formula, then $T \vdash \varphi(x) \to P(\ulcorner \varphi(\dot{x}) \urcorner)$;*

C3: *For all formulas $\varphi(x)$ and $\psi(x)$, $T \vdash P(\ulcorner \varphi(\dot{x}) \to \psi(\dot{x}) \urcorner) \to (P(\ulcorner \varphi(\dot{x}) \urcorner) \to P(\ulcorner \psi(\dot{x}) \urcorner))$.*

If $\varphi(x)$ is a $\Pi_n(T)$-formula such that $T \vdash \forall x.\mathsf{P}^h_T(\ulcorner \varphi(\dot{x}) \urcorner)$, then $T + \mathsf{Con}_P \vdash \forall x.\varphi(x)$.

Proof. As $\neg\varphi(x)$ is $\Sigma_n(T)$, by **C2** $T \vdash \neg\varphi(x) \to P(\ulcorner\neg\varphi(\dot{x})\urcorner)$. Thus, $T \vdash \neg P(\ulcorner\neg\varphi(\dot{x})\urcorner) \to \varphi(x)$. By **C3**, we have that $T + P(\ulcorner \mathsf{P}^h_T(\ulcorner\varphi(\dot{x})\urcorner)\urcorner) \wedge P(\ulcorner\neg\mathsf{P}^h_T(\ulcorner\varphi(\dot{x})\urcorner)\urcorner) \vdash P(\ulcorner\bot\urcorner)$, since $\neg\varphi := \varphi \to \bot$. Hence, $T + \mathsf{Con}_P \vdash P(\ulcorner\mathsf{P}^h_T(\ulcorner\varphi(\dot{x})\urcorner)\urcorner) \to \neg P(\ulcorner\neg\mathsf{P}^h_T(\ulcorner\varphi(\dot{x})\urcorner)\urcorner)$. Together with **C1** we get that $T + \mathsf{Con}_P \vdash P(\ulcorner\mathsf{P}^h_T(\ulcorner\varphi(\dot{x})\urcorner)\urcorner) \to \neg P(\ulcorner\neg\varphi(\dot{x})\urcorner)$, but, by what was previously concluded, one gets that $T + \mathsf{Con}_P \vdash P(\ulcorner\mathsf{P}^h_T(\ulcorner\varphi(\dot{x})\urcorner)\urcorner) \to \varphi(x)$. Suppose that $T \vdash \forall x.\mathsf{P}^h_T(\ulcorner\varphi(\dot{x})\urcorner)$. As h is primitive-recursive, we have that $\mathsf{P}^h_T(x)$ is $\Sigma_1(T)$. *Ergo*, by **C2**, $T \vdash \mathsf{P}^h_T(\ulcorner\varphi(\dot{x})\urcorner) \to P(\ulcorner\mathsf{P}^h_T(\ulcorner\varphi(\dot{x})\urcorner)\urcorner)$. By assumption, it follows that $T \vdash \forall x.P(\ulcorner\mathsf{P}^h_T(\ulcorner\varphi(\dot{x})\urcorner)\urcorner)$, therefore, $T \vdash \forall x.\varphi(x)$. ⊣

In the next result, we drop the assumption that h is primitive-recursive[10], but we need to strengthen condition **C1**.

Theorem 4.6.3. *Let T be such that there is a provability predicate $P(x)$ satisfying:*

C1: *For all formulas $\varphi(x)$, $T \vdash \mathsf{P}^h_T(\ulcorner\varphi(\dot{x})\urcorner) \to P(\ulcorner\varphi(\dot{x})\urcorner)$;*

C2: *If $\varphi(x)$ is a $\Sigma_n(T)$-formula, then $T \vdash \varphi(x) \to P(\ulcorner\varphi(\dot{x})\urcorner)$;*

C3: *For all formulas $\varphi(x)$ and $\psi(x)$, $T \vdash P(\ulcorner\varphi(\dot{x}) \to \psi(\dot{x})\urcorner) \to (P(\ulcorner\varphi(\dot{x})\urcorner) \to P(\ulcorner\psi(\dot{x})\urcorner))$.*

If $\varphi(x)$ is a $\Pi_n(T)$-formula such that $T \vdash \forall x.\mathsf{P}^h_T(\ulcorner\varphi(\dot{x})\urcorner)$, then $T + \mathsf{Con}_P \vdash \forall x.\varphi(x)$.

Proof. As $\varphi(x)$ is $\Pi_n(T)$, by **C2**, $T \vdash \exists x.\neg\varphi(x) \to \exists x.P(\ulcorner\neg\varphi(\dot{x})\urcorner)$. As $T \vdash \forall x.\mathsf{P}^h_T(\ulcorner\varphi(\dot{x})\urcorner)$, it follows, by **C1**, that $T \vdash \forall x.P(\ulcorner\varphi(\dot{x})\urcorner)$. This, together with the fact that $T + \exists x.\neg\varphi(x) \vdash \exists x.P(\ulcorner\neg\varphi(\dot{x})\urcorner)$, yields $T + \exists x.\neg\varphi(x) \vdash \exists x.P(\ulcorner\neg\varphi(\dot{x})\urcorner) \wedge P(\ulcorner\varphi(\dot{x})\urcorner)$. As $\neg\varphi := \varphi \to \bot$, it follows by **C3** that $T + \exists x.\neg\varphi(x) \vdash \exists x.P(\ulcorner\bot\urcorner)$, i.e., $T + \exists x.\neg\varphi(x) \vdash P(\ulcorner\bot\urcorner)$. Hence, $T + \mathsf{Con}_P \vdash \forall x.\varphi(x)$. ⊣

Corollary 4.6.2. *Let T be such that there is a provability predicate $P(x)$ satisfying:*

C1: *For all formulas $\varphi(x)$, $T \vdash \mathsf{P}^h_T(\ulcorner\varphi(\dot{x})\urcorner) \to P(\ulcorner\varphi(\dot{x})\urcorner)$;*

C2: *If $\varphi(x)$ is a $\Sigma_1(T)$-formula, then $T \vdash \varphi(x) \to P(\ulcorner\varphi(\dot{x})\urcorner)$;*

C3: *For all formulas $\varphi(x)$ and $\psi(x)$, $T \vdash P(\ulcorner\varphi(\dot{x}) \to \psi(\dot{x})\urcorner) \to (P(\ulcorner\varphi(\dot{x})\urcorner) \to P(\ulcorner\psi(\dot{x})\urcorner))$;*

C4: $T \vdash \mathsf{Con}_P$.

If $\varphi(x)$ is a $\Pi_1(T)$-formula such that $T \vdash \forall x.\mathsf{P}^h_T(\ulcorner\varphi(\dot{x})\urcorner)$, then $T \vdash \forall x.\varphi(x)$.

Proof. Follows immediately from the previous Theorem. ⊣

By [61] and [33], there is a provability predicate that satisfies **C2**, **C3**, and **C4** of the previous Theorem. Furthermore, if $P(x)$ is a provability predicate that satisfies **C2** and **C4**, then $P'(x) := \mathsf{P}^h_T(x) \vee P(x)$ is a provability predicate that satisfies **C1**, **C2**, and **C4**.

[10] And so P^h_T is not necessarily Σ_1.

For this reason, we believe that any sufficiently strong theory T satisfies all the previous conditions.

Using the theory K_T we can go even further:

Corollary 4.6.3. *Let T be a theory in the conditions of the previous result. If $\varphi(x)$ is a $\Pi_1(T)$-formula such that, for all $n \in \mathbb{N}$, $T \vdash_{\leq h} \varphi(\overline{n})$, then $K_T \vdash \forall x.\varphi(x)$.*

Proof. By the proof of Corollary 4.6.2, it can be concluded that $T \vdash \forall x.\mathsf{P}^h_T(\ulcorner \varphi(\dot{x})\urcorner) \to \forall x.\varphi(x)$. Thus, the result follows from Theorem 4.4.1. $\dashv$

A result similar to Theorem 4.6.3 for some Σ-formulas, holds in the presence of the stronger schema $\omega\mathsf{Con}^n_P$:

$$P(\ulcorner \exists x.\varphi(x,\dot{y})\urcorner) \to \exists x.\neg P(\ulcorner \neg\varphi(\dot{x},\dot{y})\urcorner), \qquad \varphi(x) \text{ is a } \Pi_n(T)\text{-formula.}$$

Theorem 4.6.4. *Let T be such that there is a provability predicate $P(x)$ satisfying:*

C1: *For all formulas $\varphi(x)$, $T \vdash \mathsf{P}^h_T(\ulcorner \varphi(\dot{x})\urcorner) \to P(\ulcorner \varphi(\dot{x})\urcorner)$;*

C2: *For all $\Sigma_n(T)$-formulas $\varphi(x,y)$, $T \vdash \varphi(x,y) \to P(\ulcorner \varphi(\dot{x},\dot{y})\urcorner)$.*

Suppose that $\varphi(x)$ is a $\Pi_n(T)$-formula. If $T \vdash \forall y.\mathsf{P}^h_T(\ulcorner \exists x.\varphi(x,\dot{y})\urcorner)$, then we have $T + \omega\mathsf{Con}^n_P \vdash \forall y.\exists x.\varphi(x,y)$.

Proof. Suppose that $T \vdash \forall y.\mathsf{P}^h_T(\ulcorner \exists x.\varphi(x,\dot{y})\urcorner)$. By **C1**, we have $T \vdash \forall y.P(\ulcorner \exists x.\varphi(x,\dot{y})\urcorner)$. Hence, $T + \omega\mathsf{Con}^n_P \vdash \forall y.\exists x.\neg P(\ulcorner \neg\varphi(\dot{x},\dot{y})\urcorner)$, i.e., $T + \omega\mathsf{Con}^n_P \vdash \neg\exists y.\forall x.P(\ulcorner \neg\varphi(\dot{x},\dot{y})\urcorner)$. Furthermore, by **C2**, we have $T + \exists y.\forall x.\neg\varphi(x,y) \vdash \exists y.\forall x.P(\ulcorner \neg\varphi(\dot{x},\dot{y})\urcorner)$. Therefore, $T + \omega\mathsf{Con}^n_P \vdash \neg\exists y.\forall x.\neg\varphi(x,y)$, and so, $T + \omega\mathsf{Con}^n_P \vdash \forall y.\exists x.\varphi(x,y)$. $\dashv$

If $T \vdash \omega\mathsf{Con}^n_P$, then everything is provable in T. We can yet get a stronger result, but, like before, we need a stronger schema. Let $\omega\mathsf{Con}^{3,n}_P$ be the following schema:

$$P(\ulcorner \exists y.\varphi(\dot{x},y,\dot{z})\urcorner) \to \exists y.\neg P(\ulcorner \neg\varphi(\dot{x},\dot{y},\dot{z})\urcorner), \qquad \varphi(x) \text{ is a } \Pi_n(T)\text{-formula.}$$

Theorem 4.6.5. *Let T be such that there is a provability predicate $P(x)$ satisfying:*

C1: *For all formulas $\varphi(x)$, $T \vdash \mathsf{P}^h_T(\ulcorner \varphi(\dot{x})\urcorner) \to P(\ulcorner \varphi(\dot{x})\urcorner)$;*

C2: *For all $\Pi_n(T)$-formulas $\varphi(x,y,z)$, $T \vdash P(\ulcorner \forall x.\exists y.\varphi(x,y,\dot{z})\urcorner) \to \forall x.P(\ulcorner \exists y.\varphi(\dot{x},y,\dot{z})\urcorner)$;*

C3: *For all $\Sigma_n(T)$-formulas $\varphi(x,y,z)$, $T \vdash \varphi(x,y,z) \to P(\ulcorner \varphi(\dot{x},\dot{y},\dot{z})\urcorner)$.*

Suppose that $\varphi(x)$ is a $\Pi_n(T)$-formula. If $T \vdash \forall z.\mathsf{P}^h_T(\ulcorner \forall x.\exists y.\varphi(x,y,\dot{z})\urcorner)$, then $T + \omega\mathsf{Con}^{3,n}_P \vdash \forall z.\forall x.\exists y.\varphi(x,y,z)$.

Proof. $T + \forall z.\mathsf{P}^h_T(\ulcorner \forall x.\exists y.\varphi(x,y,\dot{z})\urcorner) \vdash \forall z.P(\ulcorner \forall x.\exists y.\varphi(x,y,\dot{z})\urcorner)$, by **C1**. From **C2**, we obtain $T + \forall z.\mathsf{P}^h_T(\ulcorner \forall x.\exists y.\varphi(x,y,\dot{z})\urcorner) \vdash \forall z.\forall x.P(\ulcorner \exists y.\varphi(\dot{x},y,\dot{z})\urcorner)$. This means that $T + \omega\mathsf{Con}^{3,n}_P \vdash \forall z.\forall x.\exists y.\neg P(\ulcorner \neg\varphi(\dot{x},\dot{y},\dot{z})\urcorner)$, i.e., $T + \omega\mathsf{Con}^{3,n}_P \vdash \neg\exists z.\exists x.\forall y.P(\ulcorner \neg\varphi(\dot{x},\dot{y},\dot{z})\urcorner)$. As $\varphi(x,y,z)$ is $\Pi_n(T)$, by **C3**, $T + \exists z.\exists x.\forall y.\neg\varphi(x,y,z) \vdash \exists z.\exists x.\forall y.P(\ulcorner \neg\varphi(\dot{x},\dot{y},\dot{z})\urcorner)$. Altogether, $T + \omega\mathsf{Con}^{3,n}_P \vdash \neg\exists z.\exists x.\forall y.\neg\varphi(x,y,z)$. $\dashv$

By Theorem 4.4.5, we have that the Local Reflection Principle (see [91, p. 845]) of $\mathsf{P}^h_T(x)$ is provable in T, i.e., $T \vdash \mathsf{P}^h_T(\ulcorner \varphi \urcorner) \to \varphi$. In fact, we have the following result.

Theorem 4.6.6. *Suppose that h is primitive-recursive. Let T be such that there is a provability predicate $P(x)$ satisfying:*

C1: *For all formulas $\varphi(x)$, $T \vdash \mathsf{P}^h_T(\ulcorner \varphi(\dot{x}) \urcorner) \to P(\ulcorner \varphi(\dot{x}) \urcorner)$;*

C2: *For all $\Sigma_1(T)$-formulas $\varphi(x)$, $T \vdash \varphi(x) \to P(\ulcorner \varphi(\dot{x}) \urcorner)$;*

C3: *For all formulas $\varphi(x)$ and $\psi(x)$, $T \vdash P(\ulcorner \varphi(\dot{x}) \to \psi(\dot{x}) \urcorner) \to (P(\ulcorner \varphi(\dot{x}) \urcorner) \to P(\ulcorner \psi(\dot{x}) \urcorner))$;*

C4: *$T \vdash \mathsf{Con}_P$;*

C5: *For all formulas $\varphi(x)$, $T \vdash \varphi(x) \implies T \vdash P(\ulcorner \varphi(\dot{x}) \urcorner)$.*

Then, $T \vdash \forall x. P(\ulcorner \mathsf{P}^h_T(\ulcorner \varphi(\dot{\dot{x}}) \urcorner) \to \varphi(\dot{x}) \urcorner)$.

Proof. As h is primitive-recursive, we know that $\mathsf{P}^h_T(x)$ and $\neg\mathsf{P}^h_T(x)$ are $\Sigma_1(T)$-formulas. By **C2**, $T \vdash \mathsf{P}^h_T(\ulcorner \varphi(\dot{x}) \urcorner) \to P(\ulcorner \mathsf{P}^h_T(\ulcorner \varphi(\dot{\dot{x}}) \urcorner) \urcorner)$, so $T \vdash \neg P(\ulcorner \mathsf{P}^h_T(\ulcorner \varphi(\dot{\dot{x}}) \urcorner) \urcorner) \to \neg\mathsf{P}^h_T(\ulcorner \varphi(\dot{x}) \urcorner)$. It holds that $T \vdash \neg\mathsf{P}^h_T(\ulcorner \varphi(\dot{x}) \urcorner) \to P(\ulcorner \neg\mathsf{P}^h_T(\ulcorner \varphi(\dot{\dot{x}}) \urcorner) \urcorner)$. So, $T \vdash \neg P(\ulcorner \mathsf{P}^h_T(\ulcorner \varphi(\dot{\dot{x}}) \urcorner) \urcorner) \to P(\ulcorner \neg\mathsf{P}^h_T(\ulcorner \varphi(\dot{\dot{x}}) \urcorner) \urcorner)$, i.e., $T \vdash P(\ulcorner \mathsf{P}^h_T(\ulcorner \varphi(\dot{\dot{x}}) \urcorner) \urcorner) \vee P(\ulcorner \neg\mathsf{P}^h_T(\ulcorner \varphi(\dot{\dot{x}}) \urcorner) \urcorner)$.

From logic, $T \vdash \neg\mathsf{P}^h_T(\ulcorner \varphi(\dot{x}) \urcorner) \to (\mathsf{P}^h_T(\ulcorner \varphi(\dot{x}) \urcorner) \to \varphi(x))$. So, by **C5**, $T \vdash P(\ulcorner \neg\mathsf{P}^h_T(\ulcorner \varphi(\dot{\dot{x}}) \urcorner) \to (\mathsf{P}^h_T(\ulcorner \varphi(\dot{\dot{x}}) \urcorner) \to \varphi(\dot{x})) \urcorner)$, thus, by **C3**, $T \vdash P(\ulcorner \neg\mathsf{P}^h_T(\ulcorner \varphi(\dot{\dot{x}}) \urcorner) \urcorner) \to P(\ulcorner \mathsf{P}^h_T(\ulcorner \varphi(\dot{\dot{x}}) \urcorner) \to \varphi(\dot{x}) \urcorner)$.

From **C1**, $T + \neg P(\ulcorner \varphi(\dot{x}) \urcorner) \vdash \neg\mathsf{P}^h_T(\ulcorner \varphi(\dot{x}) \urcorner)$. By **C2**, $T + \neg P(\ulcorner \varphi(\dot{x}) \urcorner) \vdash P(\ulcorner \neg\mathsf{P}^h_T(\ulcorner \varphi(\dot{\dot{x}}) \urcorner) \urcorner)$. *Ergo* we have $T + P(\ulcorner \mathsf{P}^h_T(\ulcorner \varphi(\dot{\dot{x}}) \urcorner) \urcorner) + \neg P(\ulcorner \varphi(\dot{x}) \urcorner) \vdash P(\ulcorner \neg\mathsf{P}^h_T(\ulcorner \varphi(\dot{\dot{x}}) \urcorner) \urcorner) \wedge P(\ulcorner \mathsf{P}^h_T(\ulcorner \varphi(\dot{\dot{x}}) \urcorner) \urcorner)$. From **C3** and **C4**, it follows that $T + P(\ulcorner \mathsf{P}^h_T(\ulcorner \varphi(\dot{\dot{x}}) \urcorner) \urcorner) + \neg P(\ulcorner \varphi(\dot{x}) \urcorner) \vdash \perp$, i.e., $T \vdash P(\ulcorner \mathsf{P}^h_T(\ulcorner \varphi(\dot{\dot{x}}) \urcorner) \urcorner) \to P(\ulcorner \varphi(\dot{x}) \urcorner)$. From logic, $T \vdash \varphi(x) \to (\mathsf{P}^h_T(\ulcorner \varphi(\dot{x}) \urcorner) \to \varphi(x))$; so, by **C5** and **C3**, $T \vdash P(\ulcorner \varphi(\dot{x}) \urcorner) \to P(\ulcorner \mathsf{P}^h_T(\ulcorner \varphi(\dot{\dot{x}}) \urcorner) \to \varphi(\dot{x}) \urcorner)$. Hence, $T \vdash P(\ulcorner \mathsf{P}^h_T(\ulcorner \varphi(\dot{\dot{x}}) \urcorner) \urcorner) \to P(\ulcorner \mathsf{P}^h_T(\ulcorner \varphi(\dot{\dot{x}}) \urcorner) \to \varphi(\dot{x}) \urcorner)$.

So we have $T \vdash P(\ulcorner \mathsf{P}^h_T(\ulcorner \varphi(\dot{\dot{x}}) \urcorner) \urcorner) \vee P(\ulcorner \neg\mathsf{P}^h_T(\ulcorner \varphi(\dot{\dot{x}}) \urcorner) \urcorner)$ and also $T \vdash P(\ulcorner \neg\mathsf{P}^h_T(\ulcorner \varphi(\dot{\dot{x}}) \urcorner) \urcorner) \to P(\ulcorner \mathsf{P}^h_T(\ulcorner \varphi(\dot{\dot{x}}) \urcorner) \to \varphi(\dot{x}) \urcorner)$. From before, we have $T \vdash P(\ulcorner \mathsf{P}^h_T(\ulcorner \varphi(\dot{\dot{x}}) \urcorner) \urcorner) \to P(\ulcorner \mathsf{P}^h_T(\ulcorner \varphi(\dot{\dot{x}}) \urcorner) \to \varphi(\dot{x}) \urcorner)$, and thus the result follows. ⊣

Inspired by the previous fact, one can consider the *uniform reflection principle* schema, $\mathrm{RFN}^h(T)$, for the provability notion $\mathsf{P}^h_T(x)$ (see [91, p. 845]):

$$\forall x. (\mathsf{P}^h_T(\ulcorner \varphi(\dot{x}) \urcorner) \to \varphi(x)), \qquad \varphi(x) \text{ has only } x \text{ free.}$$

With Γ being an arbitrary class of formulas (for example Σ_n, Π_n, or even Δ_n), we denote by $\mathrm{RFN}^h_\Gamma(T)$ the previous schema restricted to Γ-formulas and define $T^h_\Gamma := K_T + \mathrm{RFN}^h_\Gamma(T)$. There is a deep relation between ω-consistency and reflection principles [91, p. 853]: Restrictions to Π-formulas of the uniform reflexion principle for the standard provability predicate are equivalent to restrictions of the schema $\omega\mathsf{Con}^n_{\mathsf{Pr}}$ from above to Σ-formulas.

Note that we are adding ω-consistency and not ω-completeness, hence Kreisel's Conjecture—which follows immediate from ω-completeness—is not being trivialized.

Now we presented another adapted version of Kreisel's Conjecture.

Theorem 4.6.7. *Let h be a total recursive function and $\varphi(x)$ be a Γ-formula such that, for all $n \in \mathbb{N}$, $T \vdash_{\leq h} \varphi(\overline{n})$. Then, $T^h_\Gamma \vdash \forall x.\varphi(x)$.*

Proof. Let h be a total recursive function and $\varphi(x)$ be a Γ-formula such that, for all $n \in \mathbb{N}$, $\mathsf{PA} \vdash_{\leq h} \varphi(\overline{n})$. By Theorem 4.4.1, we have that $K_T \vdash \forall x.\mathsf{P}^h_T(\ulcorner \varphi(\dot{x}) \urcorner)$. Thus, by $\mathrm{RFN}^h_\Gamma(T)$, it follows that $T^h_\Gamma \vdash \forall x.\varphi(x)$. ⊣

Note that there are no particular reasons to believe that the theory K_T is effectively axiomatisable. This is something worth studying.

Furthermore, one could consider a modal logic with modalities $\Box$ (interpreted by $\mathsf{Pr}_{\mathsf{PA}}(\cdot)$) and $\Box_{\leq h}$ (with $\mathsf{P}^h_T(\cdot)$ as an intended interpretation) and, at least, the usual axioms of $\Box$ and the properties of $\mathsf{P}^h_T(\cdot)$. For example, as modal versions of the Theorems 4.4.3, 4.4.5, and 4.4.6, one could add the following axioms:

Ax.1 $(\Box\Box_{\leq h}A) \vee (\Box\neg\Box_{\leq h}A)$;

Ax.2 $\Box_{\leq h}A \to A$;

Ax.3 $\neg\Box\neg\Box_{\leq h}A \to \Box_{\leq h}A$.

4.7 On $\vdash_{k\text{ steps}}$ and $\vdash_{\leq h}$

From [18, p. 8], we know the following fact:

Theorem 4.7.1. *If T is a finitely axiomatized theory, then there is a total recursive function $f(k,\#\varphi)$ such that*

$$T \vdash_{k\ steps} \varphi \implies T \vdash_{f(k,\#\varphi)\ symbols} \varphi.$$

With this Theorem, one can establish a relation between $\vdash_{k\text{ steps}}$ and $\vdash_{\leq h}$.

Theorem 4.7.2. *Given k, if T is a finitely axiomatized theory, then the function*

$$g_k(\#\varphi) := \begin{cases} 1, & T \vdash_{k\ steps} \varphi \\ 0, & otherwise \end{cases}$$

is recursive.

Proof. Let k be fixed. We will intuitively describe the algorithm that computes the function g_k. Consider the input $\#\varphi$. Compute, by Theorem 4.7.1, $f(k,\#\varphi)$. If φ is provable with at most k steps, then it must be provable using at most $f(k,\#\varphi)$ symbols. In such a hypothetical proof, clearly there are, at most, $f(k,\#\varphi)$ different variables. Furthermore, the variables, besides the ones that occur in φ, can be arbitrarily chosen, i.e., if one performs a change of variables in the proof without changing the variables occurring in φ, one maintains the soundness of the proof and the number of steps in it. This means that one can consider a finite set of variables consisting of: the variables in φ and $f(k,\#\varphi)$ other variables. Then, the algorithm considers all possible finite strings constructed using the finite set consisting of: the logical connectives, quantifiers, parenthesis, a blank symbol (to separate the steps in a proof), and the variables of the finite set that was mentioned. By vanishing over all the (finite) possible strings with at most $f(k,\#)$-symbols, the algorithm tests if any of them is a proof of φ with at most k steps. If there is any, it outputs 1, otherwise it ought to output 0. Thus, the algorithm outputs 1 exactly when φ is provable with at most k steps. ⊣

Theorem 4.7.3. *Given k, if T is a finitely axiomatized theory, then there is a total recursive function h_k such that*

$$T \vdash_{k\ steps} \varphi \implies T \vdash_{\leq h_k} \varphi.$$

Proof. Let g_k be as in Theorem 4.7.2. It is immediate that the function

$$h_k(n) := \begin{cases} m, & \text{if } g_k(n) = 1 \text{ and } m \text{ is the smallest code of a proof of the} \\ & \text{formula coded by } n \text{ with at most } k \text{ steps,} \\ 0, & \text{otherwise} \end{cases}$$

is total recursive. We show that $T \vdash_{k \text{ steps}} \varphi \implies T \vdash_{\leq h_k} \varphi$. If $T \vdash_{k \text{ steps}} \varphi$, then $g_k(\#\varphi) = 1$ and so $h_k(\#\varphi)$ is the code of a proof of φ with at most k steps; by definition, $T \vdash_{\leq h_k} \varphi$. ⊣

There are two immediate consequences of the previous result.

Corollary 4.7.1. *Suppose that T is a finitely axiomatized theory satisfying the conditions of Corollary 4.6.3 for the function h_k and that $\varphi(x)$ is a $\Pi_1(T)$-formula. If for all $n \in \mathbb{N}$, $T \vdash_{k\ steps} \varphi(\overline{n})$, then $K_T \vdash \forall x.\varphi(x)$.*

Proof. Follows from the previous Theorem and from Corollary 4.6.3. ⊣

Corollary 4.7.2. *Suppose that T is a finitely axiomatized theory and that $\varphi(x)$ be a Γ-formula. If, for all $n \in \mathbb{N}$, $T \vdash_{k\ steps} \varphi(\overline{n})$, then $T_\Gamma^{h_k} \vdash \forall x.\varphi(x)$.*

Proof. Follows from Theorems 4.7.3 and 4.6.7. ⊣

We finish with an open problem.

Problem. *Is there a total recursive function h such that, for all formulas φ, $\mathsf{PA} \vdash_{k\ steps} \varphi \implies \mathsf{PA} \vdash_{\leq h} \varphi$?*

4.8 Conclusion

Kreisel's Conjecture is a fundamental problem of k-steps-provability. As mentioned in the introduction, there are some solutions under specific conditions. Usually they rely on properties of the considered formulas or properties of the theory. In this chapter, we presented a novel approach to the conjecture, where we abstracted from the concrete formalization.

We introduced a notion of provability $\vdash_{\leq h}$ expressing that $T \vdash_{\leq h} \varphi$ holds if there is a proof of φ in T whose code is at most $h(\#\varphi)$. This is clearly an intensional notion. We studied the representation of $\vdash_{\leq h}$ inside the theory T using the formula $\mathsf{P}^h_T(x)$ and several of its properties. Montagna's conjecture ("Does $\mathsf{PA} \vdash_{k \text{ steps}} \mathsf{Pr}_{\mathsf{PA}}(\ulcorner \varphi \urcorner) \to \varphi$ imply $\mathsf{PA} \vdash_{k \text{ steps}} \varphi$?") was analyzed for the notion $\vdash_{\leq h}$.

We also considered variants of Kreisel's Conjecture for provability predicates with different derivability conditions. From the results, we like to highlight Theorem 5.4 that, using a form of ω-consistency ($\omega\mathsf{Con}^n_P$) and under certain derivability conditions, allows to conclude $T + \omega\mathsf{Con}^n_P \vdash \forall y. \exists x. \varphi(x, y)$ from $T \vdash \forall y. \mathsf{P}^h_T(\ulcorner \exists x. \varphi(x, \dot{y}) \urcorner)$.

The chapter finishes with connections between $\vdash_{k \text{ steps}}$ and $\vdash_{\leq h}$, in particular, two forms of Kreisel's Conjecture for $\vdash_{\leq h}$ (Corollaries 6.1 and 6.2).

Chapter 5

Numeral Completeness

PART A: The Mathematical Results

5.1 Introduction

Numeralwise representability [91, p. 838] is a standard notion in proof theory. It leads to studies on (numeral forms of) completeness and consistency, as [33], [97], and [62]. For significant results on partial consistency statements, we refer the reader to [76], [77], [12], [59], and [57].

In this chapter, we study numeral forms of completeness and consistency for weak theories of arithmetic, like the theory S^1_2 of bounded arithmetic and the elementary arithmetic EA—see [10], [39], and [44] for details on these theories.

Notions like proof and provability (with respect to a numeration τ) are formalized via predicates *Prf* and Pr_τ introduced in Section 5.2. In Section 5.3, we prove numeral forms of completeness and provable consistency. These results are proved for the theories S^1_2 and EA, based on derivability conditions established in the preceding section. Numeral Completeness (see Theorem 5.3.1), states that whenever one is given a *true* sentence $\vec{Q}\vec{x}.\varphi(\vec{x})$, with $\varphi(\vec{x})$ a $\Sigma^b_1(\mathsf{S}^1_2)$-formula, one is able to construct a provability predicate Pr_τ such that $\mathsf{S}^1_2 \vdash \vec{Q}\vec{x}.\mathsf{Pr}_\tau(\ulcorner\varphi(\dot{\vec{x}})\urcorner)$. This fact, in an intuitive setting, expresses that the fact that $\vec{Q}\vec{x}.\varphi(\vec{x})$ is true can be instance-wise (using numerals) captured in S^1_2 using the constructed numeration τ. We improve this result (c.f. Proposition 5.3.1) to the general case of sentences that might not be true, but we can no longer guarantee that Pr_τ is a provability predicate (in fact, we cannot guarantee that Pr_τ defines T-provability in $\mathbb{N}$).

In Section 5.4, we study the derivability condition stating "provability implies provable provability". Firstly, we confirm that, for T not necessarily satisfying the mentioned derivability condition, G2 still holds[1], see Theorem 5.4.3. Secondly, using Numeral Completeness, we construct a provability predicate Pr_τ such that weak theories T prove the previously mentioned derivability condition, *id est* $T \vdash \mathsf{Pr}_\tau(x) \to \mathsf{Pr}_\tau(\ulcorner \mathsf{Pr}_\tau(\dot{x})\urcorner)$, for a certain constructed τ. These are clearly distinct results: the first one shows that G2 does not need the mentioned condition, the second one presents a particular case for which the condition holds; these are not conflicting goals. We also explore the derivability conditions needed to state our result on numeral completeness—we show that an internal form of Σ-completeness (Σ_1^b or Σ_1-completeness) is not needed.

In addition to what we mentioned, in Section 5.5, we present characterizations of the provability predicates that satisfy our forms of provable completeness and consistency: in a sense to be made precise in the mentioned section, they are exactly the provability predicates whose consistency statements imply the considered sentence.

In Section 5.6, we study representability of function computing proofs, in particular proofs of partial consistency statements. We confirm that Numeral Completeness can be used to improve a known result (see Corollary 5.6.1). Finally, in Section 5.7, as a side remark to what was done in the preceding section, we study *bounded notions of provability*, i.e. formulas $B(x,y)$ such that $T \vdash \varphi \iff \exists n \in \mathbb{N}. T \vdash B(\ulcorner\varphi\urcorner, \overline{n})$. We exhibit bounded notions of provability, $T \overset{k \text{ steps}}{\vdash} \cdot$ and $T \overset{n}{\vdash} \cdot$, and we present a general negative result for them.

5.2 Preliminaries

Throughout this chapter, S and T stand for consistent and recursively enumerable (r.e.) theories of arithmetic that are extensions of S^1_2; S and T, unless otherwise mentioned, are not necessarily sound, i.e. $\mathbb{N} \models S, T$ does not need to occur. Moreover, S^1_2 and EA serve interchangeably as basis to our investigations.

5.2.1 Notation and definitions

Given a class of formulas Γ, an $\exists\Gamma$-formula is a formula of the form $\exists\vec{x}.\varphi(\vec{x},\vec{y})$, where $\varphi(\vec{x},\vec{y})$ is a Γ-formula; a similar definition is assumed for $\forall\Gamma$-formulas; the classes Σ_n and all standard notation are taken from [39, pp. 13–18, 62]. $\vec{Q}$ abbreviates an arbitrary sequence of quantifiers. We say that T is Γ-sound if, for every Γ-formula φ, whenever we have $T \vdash \varphi$, we also have $\mathbb{N} \models \varphi$ (see [56] for necessary and sufficient conditions for Σ_n-sounds and, besides that paper, see also [82] for an account of Incompleteness for Σ_n-sound theories).

We use the efficient numerals of bounded arithmetic [10, p. 29], the corresponding sequence functions and pairing functions [10, p. 48], and all the metamathematical notions from [10]. We adopt the notation $\#\varphi$ for the Gödel-number of the formula φ (we consider

[1]See [102] for further recent results on G2, and [99] for an exploration of G2 without the need for arithmetization "as a hidden parameter".

a fixed feasible way of coding), we set $\ulcorner\varphi\urcorner := \overline{\#\varphi}$ for the numeral of the Gödel-number of φ, and we use Feferman's dot notation $\ulcorner\varphi(\dot{\vec{x}})\urcorner$ (see [91, p. 837] and [10, p. 135]) to formalize numeral instantiation: this notation uses num, representing in S^1_2 the function $n \mapsto \#\overline{n}$ that associates with each n the Gödel-number of the numeral $\overline{n}$, it involves also the internal substitution function sub that satisfies $\mathsf{S}^1_2 \vdash \mathsf{sub}(\ulcorner\varphi\urcorner, \ulcorner t\urcorner) = \ulcorner\varphi(t)\urcorner$, for each formula φ and each term t, and can be defined by $\ulcorner\varphi(\dot{x})\urcorner := \mathsf{sub}(\ulcorner\varphi\urcorner, \mathsf{num}(x))$. Intuitively speaking, $\ulcorner\varphi(\dot{x})\urcorner$ represents the formula $\ulcorner\varphi\urcorner$ instantiated by the numeral of x; thus, Feferman's dot notation allows to represent instances of formulas by natural numbers inside a theory of (bounded) arithmetic.

We also consider $\dot{\neg}$ such that $\mathsf{S}^1_2 \vdash \dot{\neg}\ulcorner\varphi\urcorner = \ulcorner\neg\varphi\urcorner$, and similarly for the other connectives, namely $\mathsf{S}^1_2 \vdash \ulcorner\varphi\urcorner \dot{\to} \ulcorner\psi\urcorner = \ulcorner\varphi \to \psi\urcorner$. As usual, $\mathsf{S}^1_2 \vdash 0 \dot{-} \overline{1} = 0$, and for $n \geq 1$, $\mathsf{S}^1_2 \vdash \overline{n} \dot{-} \overline{1} = \overline{n-1}$ (see [10, pp. 36, 42]); Sq is a Δ^b_1-formula representing sequences in S^1_2; Fm is a Δ^b_1-formula representing formulas; and L is a function-symbol representing the length of a sequence.

Consider the following well-established definitions:

Definition 5.2.1.

1. $Prf(x,y)$ is a *proof predicate* (*for* T) when $\mathsf{S}^1_2 \vdash Prf(\overline{n},\overline{m})$ if, and only if, m is (the code of) a T-proof of the formula (whose code is) n;

2. $P(x)$ is a *provability predicate* (*for* T *over* S) if $S \vdash P(\overline{n})$ holds exactly when n is (the code of) a T-provable formula[2];

3. A theory T *is numerated by a formula* $\tau(v)$ *in* S if the set $\mathcal{S}_\tau := \{n \in \mathbb{N} | S \vdash \tau(\overline{n})\}$ coincides with the set of all (codes of) axioms (of some axiomatization) of T;

4. Given a theory T numerated by formulas θ and τ in S, θ *is included in* τ if $S \vdash \theta \to \tau$.

In the case of item 3 above, we also say that τ *is a numeration of* T *in* S, or that τ *numerates* T *in* S. Let us introduce predicates for proof and provability with respect to a considered numeration.

Definition 5.2.2 (Standard predicates)**.** Given a numeration τ:

1. The *standard proof predicate for* τ is

$$\begin{aligned}\mathtt{Prf}_\tau(x,y) :=& \mathsf{Sq}(y) \wedge \neg\mathsf{L}(y) = 0 \wedge (\forall u < \mathsf{L}(y).\mathsf{Fm}((y)_u) \wedge \\ & (\tau((y)_u) \vee \exists v < u.\exists w < u.(y)_v = (y)_w \dot{\to} (y)_u)) \\ & \wedge x = (y)_{\mathsf{L}(y)\dot{-}\overline{1}};\end{aligned}$$

2. The *standard provability predicate for* τ is $\mathtt{Pr}_\tau(x) := \exists y.\mathtt{Prf}_\tau(x,y)$.

[2]This corresponds to Kreisel's condition from [102].

We define the standard proof predicate, for a numeration τ, as Feferman did in [33]. However, Feferman emphasized the logical axioms using another formula, let us say LAxiom. In our work, the logical axioms are codified in τ, *videlicet* LAxiom is included in each considered τ. It is important to observe that whenever τ defines the T-axioms in $\mathbb{N}$, then Pr_τ is a provability predicate (this assertion assumes soundness of the basis theory); as we are going to see in Proposition 5.3.1, there are numerations τ of the axioms of T such that Pr_τ is no longer guaranteed to be a provability predicate.

Definition 5.2.3. Let Γ be a set of formulas.

1. $\Gamma(T)$ is the set of formulas that are T-provably equivalent to some formula in Γ;
2. Γ-*numeration* is a numeration by a formula in Γ;
3. A theory T is Γ-*definable in S* if there is a Γ-numeration of T in S.

5.2.2 Derivability conditions

We summarize some results, concerning derivability conditions, that are relevant to our work.

Fact 5.2.1. (A) Derivability conditions for S^1_2: *The following conditions hold when there is θ a $\Delta^b_1(\mathsf{S}^1_2)$-numeration of T in S^1_2 included in τ:*

C1. *If $T \vdash \varphi(\vec{x})$, then $\mathsf{S}^1_2 \vdash \mathsf{Pr}_\tau(\ulcorner \varphi(\dot{\vec{x}}) \urcorner)$;*

C2. $\mathsf{S}^1_2 \vdash \mathsf{Pr}_\tau(\ulcorner \varphi(\dot{\vec{x}}) \to \psi(\dot{\vec{x}}) \urcorner) \to (\mathsf{Pr}_\tau(\ulcorner \varphi(\dot{\vec{x}}) \urcorner) \to \mathsf{Pr}_\tau(\ulcorner \psi(\dot{\vec{x}}) \urcorner))$;

C3. *For all $\Sigma^b_1(\mathsf{S}^1_2)$-formulas $\varphi(\vec{x})$, $\mathsf{S}^1_2 \vdash \varphi(\vec{x}) \to \mathsf{Pr}_\tau(\ulcorner \varphi(\dot{\vec{x}}) \urcorner)$;*

C4. $\mathsf{S}^1_2 \vdash \tau(x) \to \mathsf{Pr}_\tau(x)$;

C5. *If* $\mathsf{S}^1_2 \vdash (\forall x.\xi(x) \to \tau(x)) \to (\forall x.\mathsf{Pr}_\xi(x) \to \mathsf{Pr}_\tau(x))$.

(B) Derivability conditions for EA: *The following conditions hold when there is θ a $\Sigma_1(\mathsf{EA})$-numeration of T in* EA *included in τ: conditions C1–C5 above, with 'Σ^b_1' and 'S^1_2' replaced by 'Σ_1' and '*EA*', respectively.*

We focus on (A) to briefly explain why the conditions hold. For more details we recommend [10, pp. 133–149], [11, pp. 117, 118], [33], [65, p. 14], [62], [42], and [63].

C1 follows from the Σ^b_1-completeness of S^1_2 (*conferatur* [10, p. 135]) and the fact that Pr_τ commutes with universal quantifications.

C2 follows from the definition of Prf_τ and from the fact that concatenation of arbitrary sequences is defined in S^1_2.

A proof of condition C3 can be found in [10, pp. 135–141]; there, Buss proves it for a proof predicate whose underlying system is a Gentzen-style proof system. Despite here Prf_τ is being defined for a Hilbert-style notion of proof [33], this does not constitute a problem because one can reconstruct Buss's proof with small changes for our Hilbert-style formulation; alternatively, one could simply translate, internally in S^1_2, Buss's Gentzen system to Feferman's Hilbert-style system—this approach works because one can make such a translation in a polynomially bounded way, see Theorem 5.5.1.a from [21] for further details[3]. It is worth mentioning that the condition (B) C3 is very sensitive to the formulas one considers to be Σ_1-related: in Remark 6.17 of [98], Visser presented a numeration σ of EA and a $\Sigma_{1,1}$-sentence such that the condition fails[4].

C4 is immediate from the fact that the function $n \mapsto \langle n \rangle$ is defined in S^1_2.

Finally, C5 is obtainable from logic alone.

5.3 Numeral Completeness and Consistency

Recall that T stands for a consistent r.e. theory extending S^1_2.

5.3.1 Numeral Completeness

Theorem 5.3.1 (Numeral Completeness of S^1_2)**.** *Given a $\Delta^b_1(\mathsf{S}^1_2)$-definable theory T (in S^1_2), for every true sentence $\vec{Q}\vec{x}.\varphi(\vec{x})$, with $\varphi(\vec{x})$ a $\Sigma^b_1(\mathsf{S}^1_2)$-formula, there is a numeration τ of T in S^1_2 such that Pr_τ is a provability predicate for T in S^1_2 and $\mathsf{S}^1_2 \vdash \vec{Q}\vec{x}.\mathsf{Pr}_\tau(\ulcorner \varphi(\dot{\vec{x}})\urcorner)$.*

Informally, from the true sentence $\vec{Q}\vec{x}.\varphi(\vec{x})$, one is able to prove the sentence $\vec{Q}\vec{x}.\mathsf{Pr}_\tau(\ulcorner \varphi(\dot{\vec{x}})\urcorner)$, a sentence that expresses, under the scope of $\vec{Q}\vec{x}$, that S^1_2 can prove that $\varphi(\vec{x})$ is provable in T, and consequently *true* (this is why this result constitutes a form of provable numeral *completeness*). The numerals correspond to a formal counter-part of the natural numbers; if one establishes a result for numerals, one has the guarantee that it holds for natural numbers; this entails that, in a sense, the previously mentioned result guarantees that if $\vec{Q}\vec{x}.\varphi(\vec{x})$ is true, then this fact is provable instance-wise in S^1_2 for each numeral (for a suitable numeration τ). This fact has a similar flavour to the *outside big, inside small principle* from Theorem 2.1 of [52] that states "$T \vdash \forall x \Box_T\, x \in I$", where I is a T-cut, and '$\Box_T$' denotes provability in T (for some fixed $\Delta_0(T)$-definition of the axioms).

Proof. (of the Numeral Completeness of S^1_2) T is $\Delta^b_1(\mathsf{S}^1_2)$-definable, thus there is a $\Delta^b_1(\mathsf{S}^1_2)$-formula $\psi(v)$ such that ψ numerates T in S^1_2. Define

$$\tau(v) := \psi(v) \vee \neg(\vec{Q}\vec{x}.\varphi(\vec{x})).$$

To prove that $\tau(v)$ is a numeration of T in S^1_2, it suffices to show that, for all $n \in \mathbb{N}$, $\mathsf{S}^1_2 \vdash \psi(\overline{n}) \iff \mathsf{S}^1_2 \vdash \tau(\overline{n})$. The implication $[\implies]$ is immediate. Let us prove the converse direction. Assume $\mathsf{S}^1_2 \vdash \tau(\overline{n})$. Then, using the soundness of S^1_2, $\mathbb{N} \models \tau(\overline{n})$, i.e. $\mathbb{N} \models \psi(\overline{n}) \vee$

[3] One can find strengthenings of this result in [22], [59], and [9].

[4] We recall that the class $\Sigma_{1,0}$ consists of the formulas of the form $\exists\vec{x}.S_0(\vec{x},\vec{y})$, with S_0 a Δ_0-formula; the class $\Sigma_{1,n+1}$ consists of the formulas $\exists\vec{x}.\forall\vec{y} < \vec{t}.S_0(\vec{x},\vec{y})$, where S_0 is $\Sigma_{1,n}$, see [101]. Before, $\forall\vec{y} < \vec{t}$ stands for $\forall y_0 < t_0. \cdots \forall y_{n-1} < t_{n-1}$. As Visser observed in [98], this means that, in EA, there are $\Sigma_{1,1}$-sentences that are not equivalent to Σ_1-sentences.

$\neg(\vec{Q}\vec{x}.\varphi(\vec{x}))$. As, by assumption, $\mathbb{N} \models \vec{Q}\vec{x}.\varphi(\vec{x})$, it follows that $\mathbb{N} \models \psi(\overline{n})$. As S^1_2 is Σ^b_1-complete and ψ is $\Delta^b_1(\mathsf{S}^1_2)$, we get that $\mathsf{S}^1_2 \vdash \psi(\overline{n})$.

ψ is a $\Delta^b_1(\mathsf{S}^1_2)$-numeration of T included in τ, so the derivability conditions of Fact 5.2.1(A) hold. To show that $\mathsf{S}^1_2 \vdash \vec{Q}\vec{x}.\mathsf{Pr}_\tau(\ulcorner\varphi(\dot{\vec{x}})\urcorner)$, one uses the conditions C1–C4: (i) C4 states that $\mathsf{S}^1_2 \vdash \tau(x) \to \mathsf{Pr}_\tau(x)$, consequently $\mathsf{S}^1_2 \vdash \neg\mathsf{Pr}_\tau(\ulcorner\bot\urcorner) \to \neg\tau(\ulcorner\bot\urcorner)$. Hence, $\mathsf{S}^1_2 + \neg\mathsf{Pr}_\tau(\ulcorner\bot\urcorner) \vdash \neg\psi(\ulcorner\bot\urcorner) \wedge \vec{Q}\vec{x}.\varphi(\vec{x})$, so $\mathsf{S}^1_2 + \neg\mathsf{Pr}_\tau(\ulcorner\bot\urcorner) \vdash \vec{Q}\vec{x}.\varphi(\vec{x})$. Thus, $\mathsf{S}^1_2 \vdash \neg\mathsf{Pr}_\tau(\ulcorner\bot\urcorner) \to \vec{Q}\vec{x}.\varphi(\vec{x})$; (ii) From logic, $\mathsf{S}^1_2 \vdash \bot \to \varphi(\vec{x})$, thus from C1 and C2, we therefore conclude $\mathsf{S}^1_2 \vdash \mathsf{Pr}_\tau(\ulcorner\bot\urcorner) \to \mathsf{Pr}_\tau(\ulcorner\varphi(\dot{\vec{x}})\urcorner)$. Using (i), $\mathsf{S}^1_2 \vdash \neg(\vec{Q}\vec{x}.\varphi(\vec{x})) \to \mathsf{Pr}_\tau(\ulcorner\varphi(\dot{\vec{x}})\urcorner)$. Hence, $\mathsf{S}^1_2 \vdash \neg(\vec{Q}\vec{x}.\varphi(\vec{x})) \to \vec{Q}\vec{x}.\mathsf{Pr}_\tau(\ulcorner\varphi(\dot{\vec{x}})\urcorner)$; (iii) As $\varphi(\vec{x})$ is a $\Sigma^b_1(\mathsf{S}^1_2)$-formula, from C3, we have $\mathsf{S}^1_2 \vdash \varphi(\vec{x}) \to \mathsf{Pr}_\tau(\ulcorner\varphi(\dot{\vec{x}})\urcorner)$. So, from logic, $\mathsf{S}^1_2 \vdash \vec{Q}\vec{x}.\varphi(\vec{x}) \to \vec{Q}\vec{x}.\mathsf{Pr}_\tau(\ulcorner\varphi(\dot{\vec{x}})\urcorner)$; Finally, from (ii) and (iii), $\mathsf{S}^1_2 \vdash \neg(\vec{Q}\vec{x}.\varphi(\vec{x})) \vee (\vec{Q}\vec{x}.\varphi(\vec{x})) \to \vec{Q}\vec{x}.\mathsf{Pr}_\tau(\ulcorner\varphi(\dot{\vec{x}})\urcorner)$, *ergo* $\mathsf{S}^1_2 \vdash \vec{Q}\vec{x}.\mathsf{Pr}_\tau(\ulcorner\varphi(\dot{\vec{x}})\urcorner)$.

It is easy to see that Pr_τ defines T-provability in $\mathbb{N}$, since τ defines the T-axioms in $\mathbb{N}$, and consequently Pr_τ is a provability predicate (this assertion requires the soundness of S^1_2). ⊣

It is important to emphasize that we guarantee a numeration for each sentence, but there is *no* single numeration for all sentences, otherwise truth would be recursively enumerable. We state an adapted version of the previous Theorem for EA.

Theorem 5.3.2 (Numeral Completeness of EA). *Given a theory T, for every true sentence of the form $\vec{Q}\vec{x}.\varphi(\vec{x})$, with $\varphi(\vec{x})$ a Σ_1(EA)-formula, there is a numeration τ of T in* EA *such that* Pr_τ *is a provability predicate for T in* EA *and* $\mathsf{EA} \vdash \vec{Q}\vec{x}.\mathsf{Pr}_\tau(\ulcorner\varphi(\dot{\vec{x}})\urcorner)$.

Proof. Every theory T (in our conditions) has a Σ_1-numeration τ [27] in EA. The proof of $\mathsf{EA} \vdash \vec{Q}\vec{x}.\mathsf{Pr}_\tau(\ulcorner\varphi(\dot{\vec{x}})\urcorner)$ is analogous to the correspondent result in the previous Theorem, based on the derivability conditions of Fact 5.2.1(B). ⊣

The Numeral Completeness of S^1_2 has several consequences; for instance, it yields a *small reflection principle* for $\Sigma^b_1(\mathsf{S}^1_2)$-formulas, similar to the one established in [97], namely for a $\Delta^b_1(\mathsf{S}^1_2)$-numeration τ and any formula φ, $\mathrm{I}\Delta_0 + \Omega_1 \vdash \forall x.\mathsf{Pr}_\tau(\ulcorner\forall y \leq \dot{x}.\mathsf{Prf}_\tau(\ulcorner\varphi\urcorner, y) \to \varphi\urcorner)$.

Corollary 5.3.1 (Small Reflection Principle for S^1_2). *Given a $\Delta^b_1(\mathsf{S}^1_2)$-definable theory T, for any $\Delta^b_1(\mathsf{S}^1_2)$ proof predicate Prf and any $\Sigma^b_1(\mathsf{S}^1_2)$-formula φ, there is a numeration τ of T in S^1_2 such that Pr_τ is a provability predicate for T in S^1_2 and $\mathsf{S}^1_2 \vdash \forall x.\mathsf{Pr}_\tau(\ulcorner Prf(\ulcorner\varphi\urcorner, \dot{x}) \to \varphi\urcorner)$.*

Proof. This Corollary follows from the Numeral Completeness of S^1_2 when one has in mind that $Prf(\ulcorner\varphi\urcorner, x) \to \varphi$ is a true $\Sigma^b_1(\mathsf{S}^1_2)$-formula. ⊣

The next result is also an immediate consequence of the Numeral Completeness of S^1_2; in this result, we specify the arithmetical complexity of the obtained numeration for a particular setting.

Corollary 5.3.2. *Given a $\Delta_1^b(\mathsf{S}_2^1)$-definable theory T, for every $\forall\vec{x}.\varphi(\vec{x})$, a true $\forall\Delta_1^b(\mathsf{S}_2^1)$-sentence, there is an $\exists\Delta_1^b(\mathsf{S}_2^1)$-numeration τ of T in S_2^1 such that Pr_τ is a provability predicate for T in S_2^1 and $\mathsf{S}_2^1 \vdash \forall\vec{x}.\mathsf{Pr}_\tau(\ulcorner\varphi(\dot{\vec{x}})\urcorner)$.*

Proof. The proof of the Numeral Completeness of S_2^1 guarantees that the constructed τ is an $\exists\Delta_1^b(\mathsf{S}_2^1)$-numeration satisfying the desired conditions. ⊣

It is interesting to notice that Numeral Completeness can be generalized to any kind of sentences (even to false sentences), but in this case we can no longer guarantee that the predicate we obtain defines T-provability in $\mathbb{N}$; thus, we cannot guarantee that the sentence of the result is capturing a form of truth. To prove this generalization, we use a different proof technique, to wit we make use of the Diagonalization Lemma[5]; moreover we do not make use of the soundness of S_2^1 to prove this general result.

Proposition 5.3.1 (Revisited Numeral Completeness of S_2^1). *Given a $\Delta_1^b(\mathsf{S}_2^1)$-definable theory T (in S_2^1), for every sentence (not necessarily true) $\vec{Q}\,\vec{x}.\varphi(\vec{x})$, with $\varphi(\vec{x})$ a $\Sigma_1^b(\mathsf{S}_2^1)$-formula, there is a numeration τ of T in S_2^1 such that $\mathsf{S}_2^1 \vdash \vec{Q}\,\vec{x}.\mathsf{Pr}_\tau(\ulcorner\varphi(\dot{\vec{x}})\urcorner)$.*

Proof. In this proof we use a proof predicate for S_2^1, by denoting by $\mathsf{Pr}_{\mathsf{S}_2^1}$ the predicate Pr_ξ, for some fixed $\Delta_1^b(\mathsf{S}_2^1)$-numeration ξ of the axioms of S_2^1 in S_2^1. Let $\psi(v)$ be a $\Delta_1^b(\mathsf{S}_2^1)$-numeration of T in S_2^1. We may therefore assume $\mathsf{S}_2^1 \vdash \mathsf{Pr}_{\mathsf{S}_2^1} \to \mathsf{Pr}_\psi$. By the Diagonalization Lemma (see [39, p. 158] and [84, p. 29]), we obtain a formula $\tau(v)$ such that

$$\mathsf{S}_2^1 \vdash \tau(v) \leftrightarrow \Big(\psi(v) \vee (\neg(\vec{Q}\,\vec{x}.\varphi(\vec{x})) \wedge \neg\mathsf{Pr}_{\mathsf{S}_2^1}(\ulcorner\tau(\dot{v})\urcorner))\Big).$$

Firstly, let us prove that τ is a numeration of T in S_2^1, *scilicet* that, for all $n \in \mathbb{N}$, $\mathsf{S}_2^1 \vdash \psi(\overline{n}) \iff \mathsf{S}_2^1 \vdash \tau(\overline{n})$. Clearly, the implication $\implies$ is immediate; let us argue for the other implication. Suppose $\mathsf{S}_2^1 \vdash \tau(\overline{n})$. Then, from C1, $\mathsf{S}_2^1 \vdash \mathsf{Pr}_{\mathsf{S}_2^1}(\ulcorner\tau(\overline{n})\urcorner)$, and so $\mathsf{S}_2^1 \vdash \psi(\overline{n})$.

Just like in the proof of Theorem 5.3.1, using conditions C1–C4, we can establish $\mathsf{S}_2^1 \vdash \vec{Q}\,\vec{x}.\varphi(\vec{x}) \to \vec{Q}\,\vec{x}.\mathsf{Pr}_\tau(\ulcorner\varphi(\dot{\vec{x}})\urcorner)$ and $\mathsf{S}_2^1 \vdash \mathsf{Pr}_\tau(\ulcorner\bot\urcorner) \to \mathsf{Pr}_\tau(\ulcorner\varphi(\dot{\vec{x}})\urcorner)$; thus, it suffices to prove $\mathsf{S}_2^1 \vdash \neg\mathsf{Pr}_\tau(\ulcorner\bot\urcorner) \to \vec{Q}\,\vec{x}.\varphi(\vec{x})$. As $T \nvdash \bot$ and ψ is $\Delta_1^b(\mathsf{S}_2^1)$, we have that $\mathsf{S}_2^1 \vdash \neg\psi(\ulcorner\bot\urcorner)$. So,

$$\tau(\ulcorner\bot\urcorner) \leftrightarrow (\neg(\vec{Q}\,\vec{x}.\varphi(\vec{x})) \wedge \neg\mathsf{Pr}_{\mathsf{S}_2^1}(\ulcorner\tau(\ulcorner\bot\urcorner)\urcorner)), \tag{I}$$

in particular $\mathsf{S}_2^1 \vdash \tau(\ulcorner\bot\urcorner) \to \neg\mathsf{Pr}_{\mathsf{S}_2^1}(\ulcorner\tau(\ulcorner\bot\urcorner)\urcorner)$. Moreover, we also conclude

$$\mathsf{S}_2^1 \vdash \mathsf{Pr}_{\mathsf{S}_2^1}(\ulcorner\tau(\ulcorner\bot\urcorner)\urcorner) \to \mathsf{Pr}_{\mathsf{S}_2^1}(\ulcorner\neg\mathsf{Pr}_{\mathsf{S}_2^1}(\ulcorner\tau(\ulcorner\bot\urcorner)\urcorner)\urcorner).$$

From C3, we know that $\mathsf{S}_2^1 \vdash \mathsf{Pr}_{\mathsf{S}_2^1}(\ulcorner\tau(\ulcorner\bot\urcorner)\urcorner) \to \mathsf{Pr}_{\mathsf{S}_2^1}(\ulcorner\mathsf{Pr}_{\mathsf{S}_2^1}(\ulcorner\tau(\ulcorner\bot\urcorner)\urcorner)\urcorner)$, and consequently $\mathsf{S}_2^1 \vdash \mathsf{Pr}_{\mathsf{S}_2^1}(\ulcorner\tau(\ulcorner\bot\urcorner)\urcorner) \to \mathsf{Pr}_{\mathsf{S}_2^1}(\ulcorner\bot\urcorner)$. As $\mathsf{S}_2^1 \vdash \mathsf{Pr}_{\mathsf{S}_2^1} \to \mathsf{Pr}_\tau$, we conclude

$$\mathsf{S}_2^1 \vdash \mathsf{Pr}_{\mathsf{S}_2^1}(\ulcorner\tau(\ulcorner\bot\urcorner)\urcorner) \to \mathsf{Pr}_\tau(\ulcorner\bot\urcorner). \tag{II}$$

[5] The use of the Diagonalization Lemma simply increases the complexity of the proof, but not its applicability, since this result is provable in the very weak Q (see [39, p. 158]).

From the fact that $\mathsf{S}^1_2 \vdash \tau(x) \to \mathsf{Pr}_\tau(x)$ we get $\mathsf{S}^1_2 \vdash \neg\mathsf{Pr}_\tau(\ulcorner\bot\urcorner) \to \neg\tau(\ulcorner\bot\urcorner)$. By (I), $\mathsf{S}^1_2 \vdash \neg\mathsf{Pr}_\tau(\ulcorner\bot\urcorner) \to ((\vec{Q}\,\vec{x}.\varphi(\vec{x})) \vee \mathsf{Pr}_{\mathsf{S}^1_2}(\ulcorner\tau(\ulcorner\bot\urcorner)\urcorner))$, and so, using (II), $\mathsf{S}^1_2 \vdash \neg\mathsf{Pr}_\tau(\ulcorner\bot\urcorner) \to \vec{Q}\,\vec{x}.\varphi(\vec{x})$. ⊣

As we observed, this result generalizes Theorem 5.3.1 to any kind of sentence (not necessarily true), but we can no longer guarantee that Pr_τ expresses provability, put in another way, we cannot guarantee that it is a provability predicate. Pr_τ constructed from an initial false sentence does not *intentionally* express a form of completeness; although τ numerates the axioms of T, it might be the case that Pr_τ does not define T-provability in $\mathbb{N}$: this is yet another strange facet of the incompleteness phenomenon. As the reader can see, we did not use soundness, the proof we gave is purely syntactical; this entails that the result also holds for unsound theories T, namely theories with false principles of arithmetic.

Although this result generalizes Theorem 5.3.1, in this chapter we will mainly use and discuss Theorem 5.3.1 and the idea used for the numeration in its proof; this is the case mostly because in the general setting we have no guarantee of dealing with provability predicates (not only that, but also the majority of the discussion we present in this chapter fails for that general case).

We can state a similar result for the theory EA.

Proposition 5.3.2 (Revisited Numeral Completeness of EA). *Given a $\Delta_1(\mathsf{EA})$-definable theory T (in* EA*), for every sentence $\vec{Q}\,\vec{x}.\varphi(\vec{x})$ (not necessarily true), with $\varphi(\vec{x})$ a $\Sigma_1(\mathsf{EA})$-formula, there is a numeration τ of T in* EA *such that $\mathsf{S}^1_2 \vdash \vec{Q}\,\vec{x}.\mathsf{Pr}_\tau(\ulcorner\varphi(\dot{\vec{x}})\urcorner)$.*

Proof. The proof of is similar to the previous one (we obviously make use of the derivability conditions of Fact 5.2.1(B)). ⊣

5.3.2 Numeral Consistency

We are now in conditions of presenting the following result concerning finitist consistency as a particular case of the Numeral Completeness of S^1_2. Informally, it states in S^1_2 that given a proof predicate *Prf* for T, a numeral can never be the code of a proof of $\bot$ in T.

Theorem 5.3.3 (Numeral Consistency of S^1_2). *Let Prf be a $\Delta^b_1(\mathsf{S}^1_2)$ proof predicate for a $\Delta^b_1(\mathsf{S}^1_2)$-definable theory T. There is an $\exists\Delta^b_1(\mathsf{S}^1_2)$-numeration τ of S^1_2 in S^1_2 such that Pr_τ is a provability predicate for T in S^1_2 and $\mathsf{S}^1_2 \vdash \forall x.\mathsf{Pr}_\tau(\ulcorner\neg Prf(\ulcorner\bot\urcorner,\dot{x})\urcorner)$.*

Proof. The result follows from the Corollary 5.3.2 when one has in mind that the sentence $\forall x.\neg Prf(\ulcorner\bot\urcorner, x)$ is true and $\forall\Delta^b_1(\mathsf{S}^1_2)$, by assumption, and by considering $T = \mathsf{S}^1_2$. ⊣

The previous result generalizes Proposition 7 from [10, p. 155], where Buss guarantees the existence of a bounded, consistent, theory Q (not to be confused with Robinson's arithmetic Q) satisfying, in the author's notation, "$Q \vdash (\forall x)[Q \vdash^{BD} \mathsf{Con}_Q(I_x)]$", where '$I_x$' denotes the numeral of x.

In addition to Theorem 5.3.3 above, we establish a similar result for EA.

Theorem 5.3.4 (Numeral Consistency of EA). *Let Prf be a* $\Delta_1(\mathsf{EA})$ *proof predicate for a theory* T*. Then there is a numeration* τ *of* EA *in* EA *such that* Pr_τ *is a provability predicate for* T *in* EA *and* $\mathsf{EA} \vdash \forall x.\mathsf{Pr}_\tau(\ulcorner \neg Prf(\ulcorner \bot \urcorner, \dot{x}) \urcorner)$.

Proof. Similar to the previous results, in the sense that it is an adapted carbon copy of the previous proofs by means of considering the suitable conditions from Fact 5.2.1 (B). ⊣

5.4 Gödel's Second Incompleteness Theorem and "Provability Implies Provable Provability"

Gödel's Second Incompleteness theorem (G2) states, in an intuitive way, that consistency is not provable in T. More precisely, a standard way to state G2 is due to Feferman [33]: there is a numeration τ of T such that $T \nvdash \neg\mathsf{Pr}_\tau(\ulcorner \bot \urcorner)$. Visser in [98] constructs a numeration σ of EA in EA that does not satisfy, for a sentence G, $\mathsf{Pr}_\sigma(\ulcorner G \urcorner) \to \mathsf{Pr}_\sigma(\ulcorner \mathsf{Pr}_\sigma(\ulcorner G \urcorner) \urcorner)$, and that does not prove the formalized G2 for σ, i.e. $\neg\mathsf{Pr}_\sigma(\ulcorner \bot \urcorner) \to \neg\mathsf{Pr}_\sigma(\ulcorner \neg\mathsf{Pr}_\sigma(\ulcorner \bot \urcorner) \urcorner)$.

The standard proof of G2 (see, for example, [39, p. 164]) requires the use of the derivability condition stating "provability implies provable provability"—corresponding to C3 in our context.

The derivability condition "provability implies provable provability", to wit $\mathsf{Pr}_\tau(x) \to \mathsf{Pr}_\tau(\ulcorner \mathsf{Pr}_\tau(\dot{x}) \urcorner)$, is very sensitive to the considered basis theory; for instance, for $\mathsf{I}\Delta_0$ it is still not known if it holds in general for $\Delta_0(\mathsf{I}\Delta_0)$-numerations.

In this section, we have two main goals. Firstly, we are going to see that to prove a version of G2 one does not necessarily need the theory T to satisfy "provability implies provable provability"; secondly, we are going to guarantee the existence of a numeration τ such that Pr_τ defines T-provability in $\mathbb{N}$ and $T \vdash \mathsf{Pr}_\tau(x) \to \mathsf{Pr}_\tau(\ulcorner \mathsf{Pr}_\tau(\dot{x}) \urcorner)$; this result holds for very weak T, not necessarily including S^1_2. These are not competing goals: the first one shows that G2 does not need the mentioned condition, the second one presents a particular case for which the condition holds. The idea used in the construction of the numeration for Theorem 5.3.1 will be used throughout.

5.4.1 Numeral Completeness

We show that we can avoid to use the derivability condition C3 of Fact 5.2.1, the internal Σ^b_1-completeness, in the proof of the Numeral Completeness of S^1_2.

Theorem 5.4.1. *The Numeral Completeness of* S^1_2 *can be proved without using* C3.

Proof. Let us follow the notation of the proof of the Numeral Completeness of S^1_2 (Theorem 5.3.1), but consider instead the numeration

$$\tau(v) := \psi(v) \vee \neg(\vec{Q}\,\vec{x}.\mathsf{Pr}_\psi(\ulcorner \varphi(\dot{\vec{x}}) \urcorner)).$$

By assumption, $\mathbb{N} \models \vec{Q}\vec{x}.\varphi(\vec{x})$, and so, as S^1_2 is Σ^b_1-complete, $\mathbb{N} \models \vec{Q}\vec{x}.\mathsf{Pr}_\psi(\ulcorner\varphi(\dot{\vec{x}})\urcorner)$. Repeating the initial reasoning made in the proof of Theorem 5.3.1, we can claim that τ is a numeration of T in S^1_2 and that the numeration ψ is included in τ (i.e., $\mathsf{S}^1_2 \vdash \psi \to \tau$). Thus, the conditions of Fact 5.2.1(A) hold and, in particular, by C5 we have (a) $\mathsf{S}^1_2 \vdash \mathsf{Pr}_\psi \to \mathsf{Pr}_\tau$.

Repeating item (i) from the proof of Theorem 5.3.1, we have (b) $\mathsf{S}^1_2 \vdash \neg\mathsf{Pr}_\tau(\ulcorner\bot\urcorner) \to \vec{Q}\vec{x}.\mathsf{Pr}_\psi(\ulcorner\varphi(\dot{\vec{x}})\urcorner)$. This uses only C4.

Repeating item (ii) from the proof of Theorem 5.3.1, we have that $\mathsf{S}^1_2 \vdash \mathsf{Pr}_\tau(\ulcorner\bot\urcorner) \to \mathsf{Pr}_\tau(\ulcorner\varphi(\dot{\vec{x}})\urcorner)$, and thus (c) $\mathsf{S}^1_2 \vdash \mathsf{Pr}_\tau(\ulcorner\bot\urcorner) \to \vec{Q}\vec{x}.\mathsf{Pr}_\tau(\ulcorner\varphi(\dot{\vec{x}})\urcorner)$ holds. This uses only C1 and C2.

From (a) and (b) we obtain $\mathsf{S}^1_2 \vdash \neg\mathsf{Pr}_\tau(\ulcorner\bot\urcorner) \to \vec{Q}\vec{x}.\mathsf{Pr}_\tau(\ulcorner\varphi(\dot{\vec{x}})\urcorner)$. Finally, this together with (c) leads to $\mathsf{S}^1_2 \vdash \vec{Q}\vec{x}.\mathsf{Pr}_\tau(\ulcorner\varphi(\dot{\vec{x}})\urcorner)$. The derivability condition C3 was not used. ⊣

Although the previous numeration does not require the use of C3 by using the much weaker condition C5, this comes with a cost: this numeration has a bigger arithmetical complexity than the original one. The arithmetical complexity of the numeration is of interest by its own right, but we will also see in Section 6 that it plays an important role to extract functional information.

5.4.2 Gödel's Second Incompleteness Theorem for Weak Theories

In the rest of this section, and only here, T is a theory that includes Robinson's Q (and not necessarily S^1_2) and ψ is any fixed $\Sigma_1(T)$-numeration of T in T. Such theory T is potentially very weak, as we do not demand the derivability condition C3, i.e. the internal Σ_1-completeness or the internal Σ^b_1-completeness. For simplicity, here we write 'numeration of T' instead of 'numeration of T in T'.

Consider the derivability conditions of Section 5.2.2, but for T. For C1, C2, and C4 we consider the following non-uniform[6] versions:

C1'. If $T \vdash \varphi$, then $T \vdash \mathsf{Pr}_\tau(\ulcorner\varphi\urcorner)$;

C2'. $T \vdash \mathsf{Pr}_\tau(\ulcorner\varphi \to \psi\urcorner) \to (\mathsf{Pr}_\tau(\ulcorner\varphi\urcorner) \to \mathsf{Pr}_\tau(\ulcorner\psi\urcorner))$;

C4'. $T \vdash \tau(\ulcorner\varphi\urcorner) \to \mathsf{Pr}_\tau(\ulcorner\varphi\urcorner)$.

The following result is in the same spirit as Hilbert-Bernays Completeness Theorem from [91, p. 860], but here it is stated for T, a potentially very weak theory.

Theorem 5.4.2 (Consistency Completeness of T)**.** *Let T be a sound theory that satisfies* C4' *for any numeration of T that includes ψ. Consider φ any true sentence. There is a numeration τ of T such that Pr_τ is a provability predicate for T and $T + \neg\mathsf{Pr}_\tau(\ulcorner\bot\urcorner) \vdash \varphi$.*

[6]This term is commonly used in reference to the fact that these conditions are not stated using numeral abstraction of the free-variables of the considered formulas.

Proof. Consider ψ the mentioned $\Sigma_1(T)$-numeration of T. Take[7]

$$\tau(v) := \psi(v) \vee (\neg\varphi \wedge v = \ulcorner\bot\urcorner).$$

From a previously reasoning, τ is a numeration of T (this works because T includes Q, and so is Σ_1-complete; we also require the soundness of T). Furthermore, by C4', we obtain $T + \neg\mathsf{Pr}_\tau(\ulcorner\bot\urcorner) \vdash \neg\tau(\ulcorner\bot\urcorner)$, consequently $T + \neg\mathsf{Pr}_\tau(\ulcorner\bot\urcorner) \vdash \varphi$. The fact that Pr_τ defines T-provability in $\mathbb{N}$ follows from the fact that τ defines the T-axioms in $\mathbb{N}$ and the fact that T is sound. ⊣

Similarly to what is done in Proposition 5.3.1, we revisit the previous result, dropping the soundness of T and the assumption that φ is true and substitute them by $T \nvdash \neg\varphi$, but we can no longer guarantee that Pr_τ defines T-provability in $\mathbb{N}$. This means that, in this context, T might have some false principles, since we do not require soundness.

Proposition 5.4.1 (Revisited Consistency Completeness of T)**.** *Let T be a theory that satisfies* C4' *for any numeration of T that includes ψ and such that $T \vdash \forall x.\forall y.x \leq y \vee y < x$. Let φ be any sentence (not necessarily true) with $T \nvdash \neg\varphi$. Then, there is a numeration τ of T such that $T + \neg\mathsf{Pr}_\tau(\ulcorner\bot\urcorner) \vdash \varphi$.*

Proof. Without loss of generality, we may assume that $\psi(v)$ is of the form $\exists x.\delta(v,x)$, for some Δ_0-formula $\delta(v,x)$. By the Diagonalization Lemma, we can construct a formula $\tau(v)$ such that, for any $n \in \mathbb{N}$,

$$T \vdash \tau(\overline{n}) \leftrightarrow \Big(\exists x.\delta(\overline{n},x) \wedge \forall y < x.\neg\mathsf{Prf}_\psi(\ulcorner\tau(\overline{n})\urcorner, y)\Big) \vee \neg\varphi.$$

Let us confirm that τ is a numeration of T. It suffices to establish $T \vdash \psi(\overline{n}) \iff T \vdash \tau(\overline{n})$.

- Suppose, in view of obtaining a contradiction, that $T \vdash \psi(\overline{n})$ and $T \nvdash \tau(\overline{n})$. Since $\exists x.(\delta(\overline{n},x) \wedge \forall y < x.\neg\mathsf{Prf}_\psi(\ulcorner\tau(\overline{n})\urcorner, y))$ is true, thus, from the Σ_1-completeness of T, $T \vdash \exists x.(\delta(\overline{n},x) \wedge \forall y < x.\neg\mathsf{Prf}_\psi(\ulcorner\tau(\overline{n})\urcorner, y))$; and so $T \vdash \tau(\overline{n})$, a contradiction.

- Suppose, towards a contradiction, that $T \vdash \tau(\overline{n})$ and $T \nvdash \psi(\overline{n})$. So, $\exists y.\mathsf{Prf}_\psi(\ulcorner\tau(\overline{n})\urcorner, y) \wedge \forall x.\neg\delta(\overline{n},x)$ is true and so, from Σ_1-completeness,

  $$T \vdash \exists y.(\mathsf{Prf}_\psi(\ulcorner\tau(\overline{n})\urcorner, y) \wedge \forall x \leq y.\neg\delta(\overline{n},x)). \tag{I}$$

  Reason inside the theory T and assume[8], aiming at a contradiction, that $\exists x_0.(\delta(\overline{n},x_0) \wedge \forall y < x_0.\neg\mathsf{Prf}_\psi(\ulcorner\tau(\overline{n})\urcorner, y))$. Consider such an x_0 and consider y_0 satisfying the existentially quantified (I). This means that

  a) $\mathsf{Prf}_\psi(\ulcorner\tau(\overline{n})\urcorner, y_0) \wedge \forall x \leq y_0.\neg\delta(\overline{n},x)$, and

[7]The proof clearly works for $\tau(v) := \psi(v) \vee \neg\varphi$, but the numeration we are considering is going to be useful for the proof of Theorem 5.4.4.

[8]This is a reasoning similar to the one used to establish Rosser's version of the First Incompleteness Theorem.

b) $\delta(\overline{n}, x_0) \wedge \forall y < x_0.\neg\texttt{Prf}_\psi(\ulcorner\tau(\overline{n})\urcorner, y)$.

We know that $x_0 \leq y_0 \vee y_0 < x_0$. Let us analyze both cases:

$x_0 \leq y_0$: In this case, from b), $\delta(\overline{n}, x_0)$. From a), as $x_0 \leq y_0$, we also get $\neg\delta(\overline{n}, x_0)$, a contradiction.

$y_0 < x_0$: From a), $\texttt{Prf}_\psi(\ulcorner\tau(\overline{n})\urcorner, y_0)$, and, as $y_0 < x_0$, from b) we get $\neg\texttt{Prf}_\psi(\ulcorner\tau(\overline{n})\urcorner, y_0)$, a contradiction.

As in both cases we obtained a contradiction, we can therefore conclude $\neg\exists x_0.(\delta(\overline{n}, x_0) \wedge \forall y < x_0.\neg\texttt{Prf}_\psi(\ulcorner\tau(\overline{n})\urcorner, y))$. Going outside T, we established $T \vdash \neg\exists x_0.(\delta(\overline{n}, x_0) \wedge \forall y < x_0.\neg\texttt{Prf}_\psi(\ulcorner\tau(\overline{n})\urcorner, y))$. Then, $T \vdash \tau(\overline{n}) \leftrightarrow \neg\varphi$; so $T \vdash \neg\varphi$, a contradiction.

We can establish $T + \neg\texttt{Pr}_\tau(\ulcorner\bot\urcorner) \vdash \varphi$ in a similar way to the one followed in the proof of Theorem 5.4.2. ⊣

Emission[9] is one of the standard properties for provability predicates. It states, for provability predicates P_0 and P_1, that, for any formula φ, $T \vdash P_0(\ulcorner\varphi\urcorner) \to P_0(\ulcorner P_1(\ulcorner\varphi\urcorner)\urcorner)$. The next result presents a weak form of a property similar to emission, namely $T \vdash P_0(\ulcorner\varphi\urcorner) \to P_1(\ulcorner P_0(\ulcorner\varphi\urcorner)\urcorner)$. Such a property is going to be very useful to study "provability implies provable provability", see Theorem 5.4.4.

Proposition 5.4.2 (Local Form of Emission and Σ_1-Completeness for T)**.** *Let T be a $\Sigma_2(T)$-sound theory that satisfies* C1', C2', *and* C4' *for generic numerations of T that include ψ, and let σ be any Σ_1-formula. Then, there is a numeration τ of T such that $\texttt{Pr}_\tau$ is a provability predicate for T and $T \vdash \sigma \to \texttt{Pr}_\tau(\ulcorner\sigma\urcorner)$. In particular, for any formula φ, there is τ such that $T \vdash \texttt{Pr}_\psi(\ulcorner\varphi\urcorner) \to \texttt{Pr}_\tau(\ulcorner\texttt{Pr}_\psi(\ulcorner\varphi\urcorner)\urcorner)$.*

Proof. Clearly, $\sigma \to \texttt{Pr}_\psi(\ulcorner\sigma\urcorner)$ is true. Take τ as defined in the proof of the Theorem 5.4.2 applied to the mentioned true sentence (this uses C4')[10]. Then, $T + \neg\texttt{Pr}_\tau(\ulcorner\bot\urcorner) \vdash \sigma \to \texttt{Pr}_\psi(\ulcorner\sigma\urcorner)$, and so $T + \neg\texttt{Pr}_\tau(\ulcorner\bot\urcorner) \vdash \sigma \to \texttt{Pr}_\tau(\ulcorner\sigma\urcorner)$. Furthermore, it is straightforward to see, using C1' and C2', that we have $T + \texttt{Pr}_\tau(\ulcorner\bot\urcorner) \vdash \sigma \to \texttt{Pr}_\tau(\ulcorner\sigma\urcorner)$, and so the result follows. (The fact that $\texttt{Pr}_\tau$ is a provability predicate follows from the fact that τ defines the T-axioms in $\mathbb{N}$, from the fact that $\texttt{Pr}_\tau$ is $\Sigma_2(T)$, and the fact that T is $\Sigma_2(T)$-sound.) ⊣

We now show a uniform version of the Local Form of Emission for T.

Proposition 5.4.3 (Uniform Form of Emission for T)**.** *Let T be a $\Sigma_2(T)$-sound theory that satisfies* C1, C2, *and* C4', *for generic numerations of T that include ψ. Then, there is a numeration τ of T such that $\texttt{Pr}_\tau$ is a provability predicate for T and $T \vdash \forall x.\texttt{Pr}_\psi(x) \to \texttt{Pr}_\tau(\ulcorner\texttt{Pr}_\psi(\dot{x})\urcorner)$.*

[9] For a study of emission and absorption, with a special focus on the latter, see [102].

[10] For this result, $\Sigma_2(T)$-soundness is enough, since we only need soundness to guarantee that τ is a numeration, where $\tau(v) := \psi(v) \vee (\neg\varphi \wedge v = \ulcorner\bot\urcorner)$, with $\varphi := \sigma \to \texttt{Pr}_\psi(\ulcorner\sigma\urcorner)$ a $\Sigma_2(T)$-sentence.

Proof. It is clear that $\forall x.\mathsf{Pr}_\psi(x) \to \mathsf{Pr}_\psi(\ulcorner \mathsf{Pr}_\psi(\dot{x})\urcorner)$ is a true sentence. Let τ be the numeration guaranteed for that true sentence by the Consistency Completeness of T (this step uses C4' and $\Sigma_2(T)$-soundness). Then, $T + \neg\mathsf{Pr}_\tau(\ulcorner\bot\urcorner) \vdash \forall x.\mathsf{Pr}_\psi(x) \to \mathsf{Pr}_\psi(\ulcorner \mathsf{Pr}_\psi(\dot{x})\urcorner)$, and so $T + \neg\mathsf{Pr}_\tau(\ulcorner\bot\urcorner) \vdash \forall x.\mathsf{Pr}_\psi(x) \to \mathsf{Pr}_\tau(\ulcorner \mathsf{Pr}_\psi(\dot{x})\urcorner)$. It is easy to conclude, using C1 and C2, that $T + \mathsf{Pr}_\tau(\ulcorner\bot\urcorner) \vdash \forall x.\mathsf{Pr}_\psi(x) \to \mathsf{Pr}_\tau(\ulcorner \mathsf{Pr}_\psi(\dot{x})\urcorner)$, which proves the result. ⊣

The next result is a form of G2 for T, in the sense that it claims the existence of a numeration τ such that the corresponding consistency statement, namely $\neg\mathsf{Pr}_\tau(\ulcorner\bot\urcorner)$, is not provable in T.

Theorem 5.4.3 (G2 for Weak Theories)**.** *Let T be a $\Sigma_1(T)$-sound theory that satisfies* C1', C2', *and* C4' *for any numeration of T that includes ψ. Then, there is a numeration τ of T such that Pr_τ is a provability predicate for T and $T \vdash \neg\mathsf{Pr}_\tau(\ulcorner\bot\urcorner) \leftrightarrow \neg\mathsf{Pr}_\psi(\ulcorner\neg\mathsf{Pr}_\tau(\ulcorner\bot\urcorner)\urcorner)$. In particular:*

1. $T \nvdash \neg\mathsf{Pr}_\tau(\ulcorner\bot\urcorner)$*;*

2. *ψ verifies* G2 *for τ, i.e. $T \vdash \neg\mathsf{Pr}_\tau(\ulcorner\bot\urcorner) \to \neg\mathsf{Pr}_\psi(\ulcorner\neg\mathsf{Pr}_\tau(\ulcorner\bot\urcorner)\urcorner)$.*

Proof. As T includes Q, by the Diagonalization Lemma, we ensure the existence of a sentence $\mathcal{G}$ such that $T \vdash \mathcal{G} \leftrightarrow \neg\mathsf{Pr}_\psi(\ulcorner\mathcal{G}\urcorner)$. Take τ as defined in the Consistency Completeness of T for the true sentence[11] $\mathcal{G}$. As $T \vdash \psi \to \tau$, it follows from logic that $T \vdash \mathsf{Pr}_\psi \to \mathsf{Pr}_\tau$, therefore $T \vdash \neg\mathsf{Pr}_\tau(\ulcorner\bot\urcorner) \to \neg\mathsf{Pr}_\psi(\ulcorner\bot\urcorner)$.

As $T + \mathcal{G} \vdash \tau \leftrightarrow \psi$, it follows that $T + \mathcal{G} \vdash \mathsf{Pr}_\tau \leftrightarrow \mathsf{Pr}_\psi$, and so we can claim that

$$T + \mathcal{G} \vdash \neg\mathsf{Pr}_\psi(\ulcorner\bot\urcorner) \to \neg\mathsf{Pr}_\tau(\ulcorner\bot\urcorner). \tag{I}$$

Clearly, from C1' and C2', $T \vdash \mathsf{Pr}_\psi(\ulcorner\bot\urcorner) \to \mathsf{Pr}_\psi(\ulcorner\mathcal{G}\urcorner)$, so $T \vdash \neg\mathsf{Pr}_\psi(\ulcorner\mathcal{G}\urcorner) \to \neg\mathsf{Pr}_\psi(\ulcorner\bot\urcorner)$, i.e. $T \vdash \mathcal{G} \to \neg\mathsf{Pr}_\psi(\ulcorner\bot\urcorner)$. From (I) we obtain that $T + \mathcal{G} \vdash \neg\mathsf{Pr}_\tau(\ulcorner\bot\urcorner)$, and consequently $T \vdash \mathcal{G} \to \neg\mathsf{Pr}_\tau(\ulcorner\bot\urcorner)$. By the Consistency Completeness of T (that uses C4') we get $T \vdash \neg\mathsf{Pr}_\tau(\ulcorner\bot\urcorner) \to \mathcal{G}$, and so $T \vdash \neg\mathsf{Pr}_\tau(\ulcorner\bot\urcorner) \leftrightarrow \mathcal{G}$. So, $T \vdash \neg\mathsf{Pr}_\tau(\ulcorner\bot\urcorner) \leftrightarrow \neg\mathsf{Pr}_\psi(\ulcorner\neg\mathsf{Pr}_\tau(\ulcorner\bot\urcorner)\urcorner)$. ⊣

5.4.3 'Provability Implies Provable Provability' for Weak Theories

The next result holds for $T = \mathsf{I}\Delta_0$.

Theorem 5.4.4 ('Provability Implies Provable Provability' for Weak Theories)**.** *Let T be a $\Sigma_2(T)$-sound theory with a numeration ψ of the axioms of T satisfying the conditions of Proposition 5.4.3. Then, there is a $\Sigma_2(T)$-numeration τ of the axioms of T such that Pr_τ is a provability predicate for T and $T \vdash \mathsf{Pr}_\tau(x) \to \mathsf{Pr}_\tau(\ulcorner\mathsf{Pr}_\tau(\dot{x})\urcorner)$.*

[11] To apply this result, we just need $\Sigma_1(T)$-soundness to guarantee that $\tau(v) := \psi(v) \vee (\neg\mathcal{G} \wedge v = \ulcorner\bot\urcorner)$, a $\Sigma_1(T)$-sentence, is a numeration.

Proof. Let τ be the numeration that is guaranteed to exist from Proposition 5.4.3 such that[12]

$$T \vdash \mathsf{Pr}_{\psi}(x) \rightarrow \mathsf{Pr}_{\tau}(\ulcorner \mathsf{Pr}_{\psi}(\dot{x})\urcorner). \tag{I}$$

By construction, $T \vdash \psi \rightarrow \tau$, and so $T \vdash \mathsf{Pr}_{\psi} \rightarrow \mathsf{Pr}_{\tau}$, and thus

$$T \vdash \mathsf{Pr}_{\tau}(\ulcorner \mathsf{Pr}_{\psi}(\dot{x})\urcorner) \rightarrow \mathsf{Pr}_{\tau}(\ulcorner \mathsf{Pr}_{\tau}(\dot{x})\urcorner). \tag{II}$$

Reason inside T. Let us assume $\mathsf{Pr}_{\tau}(x)$ and prove $\mathsf{Pr}_{\tau}(\ulcorner \mathsf{Pr}_{\tau}(\dot{x})\urcorner)$. So, $\exists y.\mathsf{Prf}_{\tau}(x,y)$, namely

$$\begin{aligned} &\exists y.\mathsf{Sq}(y) \wedge \neg \mathsf{L}(y) = 0 \wedge (\forall u < \mathsf{L}(y).\mathsf{Fm}((y)_u) \wedge \\ &(\tau((y)_u) \vee \exists v < u.\exists w < u.(y)_v = (y)_w \dot{\rightarrow} (y)_u)) \\ &\wedge x = (y)_{\mathsf{L}(y) \dot{-} 1}. \end{aligned} \tag{III}$$

From logic, either **1:** $\forall u < \mathsf{L}(y).\mathsf{Fm}((y)_u) \wedge (\psi((y)_u) \vee \exists v < u.\exists w < u.(y)_v = (y)_w \dot{\rightarrow} (y)_u)$, or **2:** $\exists u < \mathsf{L}(y).\neg \mathsf{Fm}((y)_u) \vee (\neg \psi((y)_u) \wedge \neg \exists v < u.\exists w < u.(y)_v = (y)_w \dot{\rightarrow} (y)_u)$. Let us analyze the two possibilities:

1: In this case, it follows that $\mathsf{Pr}_{\psi}(x)$, so, using (I), it follows $\mathsf{Pr}_{\tau}(\ulcorner \mathsf{Pr}_{\psi}(\dot{x})\urcorner)$; using (II) we obtain the desired $\mathsf{Pr}_{\tau}(\ulcorner \mathsf{Pr}_{\tau}(\dot{x})\urcorner)$.

2: From (III), we know that $\forall u < \mathsf{L}(y).\mathsf{Fm}((y)_u)$, so $\exists u < \mathsf{L}(y).\neg \psi((y)_u) \wedge \neg \exists v < u.\exists w < u.(y)_v = (y)_w \dot{\rightarrow} (y)_u$. Let $u_0 < \mathsf{L}(y)$ satisfy the previous existentially quantified formula. Then, $\neg \psi((y)_{u_0}) \wedge \neg \exists v < u_0.\exists w < u_0.(y)_v = (y)_w \dot{\rightarrow} (y)_{u_0}$, in particular $\neg \exists v < u_0.\exists w < u_0.(y)_v = (y)_w \dot{\rightarrow} (y)_{u_0}$. From (III), $\tau((y)_{u_0})$. Using again the construction of τ from Proposition 5.4.3, $\psi((y)_{u_0}) \vee (\neg \varphi \wedge (y)_{u_0} = \ulcorner \bot \urcorner)$, where $\varphi := \forall x.\mathsf{Pr}_{\psi}(x) \rightarrow \mathsf{Pr}_{\psi}(\ulcorner \mathsf{Pr}_{\psi}(\dot{x})\urcorner)$. As we concluded $\neg \psi((y)_{u_0})$, we have $(y)_{u_0} = \ulcorner \bot \urcorner$; this entails $\tau(\ulcorner \bot \urcorner)$, *ergo* $\mathsf{Pr}_{\tau}(\ulcorner \bot \urcorner)$. Using the fact that[13] $\mathsf{Pr}_{\tau}(\ulcorner \bot \urcorner) \rightarrow \mathsf{Pr}_{\tau}(\ulcorner \mathsf{Pr}_{\tau}(\dot{x})\urcorner)$, we conclude $\mathsf{Pr}_{\tau}(\ulcorner \mathsf{Pr}_{\tau}(\dot{x})\urcorner)$.

As both in **1** and in **2** we were able to conclude $\mathsf{Pr}_{\tau}(\ulcorner \mathsf{Pr}_{\tau}(\dot{x})\urcorner)$, it follows that $\mathsf{Pr}_{\tau}(\ulcorner \mathsf{Pr}_{\tau}(\dot{x})\urcorner)$ must, in fact, hold. Canceling the initial assumption, we obtain $\mathsf{Pr}_{\tau}(x) \rightarrow \mathsf{Pr}_{\tau}(\ulcorner \mathsf{Pr}_{\tau}(\dot{x})\urcorner)$; stepping outside T it entails the desired fact: $T \vdash \mathsf{Pr}_{\tau}(x) \rightarrow \mathsf{Pr}_{\tau}(\ulcorner \mathsf{Pr}_{\tau}(\dot{x})\urcorner)$. ⊣

5.5 Characterization of the Provability Predicates that Satisfy Numeral Completeness and Consistency

In this section, we give a characterization of the provability predicates that satisfy the completeness and the consistency results from before, for theories including EA[14]. Thus,

[12] It is straightforward that τ is $\Sigma_2(T)$.

[13] From logic, $T \vdash \bot \rightarrow \mathsf{Pr}_{\tau}(x)$, and so $T \vdash \mathsf{Pr}_{\tau}(\ulcorner \bot \rightarrow \mathsf{Pr}_{\tau}(\dot{x})\urcorner)$, thus $T \vdash \mathsf{Pr}_{\tau}(\ulcorner \bot \urcorner) \rightarrow \mathsf{Pr}_{\tau}(\ulcorner \mathsf{Pr}_{\tau}(\dot{x})\urcorner)$.

[14] An analogous approach could be followed for S^1_2.

along this section, S and T are theories (not necessarily sound) extending EA. In this context, for P a provability predicate for T over a sub-theory S of T we consider the following derivability conditions:

P1. If $T \vdash \varphi(\vec{x})$, then $S \vdash P(\ulcorner \varphi(\dot{\vec{x}}) \urcorner)$;

P2. $S \vdash P(\ulcorner \varphi(\dot{\vec{x}}) \to \psi(\dot{\vec{x}}) \urcorner) \to (P(\ulcorner \varphi(\dot{\vec{x}}) \urcorner) \to P(\ulcorner \psi(\dot{\vec{x}}) \urcorner))$;

P3. For all $\Sigma_1(S)$-formula $\varphi(\vec{x})$, $S \vdash \varphi(\vec{x}) \to P(\ulcorner \varphi(\dot{\vec{x}}) \urcorner)$.

Lemma 5.5.1. *Let P be a provability predicate for T over a sub-theory S of T satisfying* P1–P3. *If $\varphi(\vec{x})$ is a $\Sigma_1(S)$-formula such that $S \vdash \neg P(\ulcorner \bot \urcorner) \to \vec{Q}\,\vec{x}.\varphi(\vec{x})$, then $S \vdash \vec{Q}\,\vec{x}.P(\ulcorner \varphi(\dot{\vec{x}}) \urcorner)$.*

Proof. Clearly, $S \vdash \bot \to \varphi(\vec{x})$, which implies, by P1, $S \vdash P(\ulcorner \bot \to \varphi(\dot{\vec{x}}) \urcorner)$, and consequently by P2, $S \vdash P(\ulcorner \bot \urcorner) \to P(\ulcorner \varphi(\dot{\vec{x}}) \urcorner)$. Therefore, $S \vdash \neg \vec{Q}\,\vec{x}.\varphi(\vec{x}) \to P(\ulcorner \varphi(\dot{\vec{x}}) \urcorner)$, so $S \vdash \neg \vec{Q}\,\vec{x}.\varphi(\vec{x}) \to \vec{Q}\,\vec{x}.P(\ulcorner \varphi(\dot{\vec{x}}) \urcorner)$. Moreover, by P3, $S \vdash \varphi(\vec{x}) \to P(\ulcorner \varphi(\dot{\vec{x}}) \urcorner)$, so $S \vdash \vec{Q}\,\vec{x}.\varphi(\vec{x}) \to \vec{Q}\,\vec{x}.P(\ulcorner \varphi(\dot{\vec{x}}) \urcorner)$. In sum, $S \vdash \vec{Q}\,\vec{x}.P(\ulcorner \varphi(\dot{\vec{x}}) \urcorner)$. ⊣

Lemma 5.5.2. *Let P be a provability predicate for T over a sub-theory S of T satisfying* P1–P3. *If $\varphi(\vec{x})$ is a $\Pi_1(S)$-formula satisfying $S \vdash \vec{Q}\,\vec{x}.P(\ulcorner \varphi(\dot{\vec{x}}) \urcorner)$, then $S \vdash \neg P(\ulcorner \bot \urcorner) \to \vec{Q}\,\vec{x}.\varphi(\vec{x})$.*

Proof. Assume the antecedent of the implication we want to prove. It is clear, using P1 and P2, that $S + P(\ulcorner \varphi(\dot{\vec{x}}) \urcorner) \wedge P(\ulcorner \neg\varphi(\dot{\vec{x}}) \urcorner) \vdash P(\ulcorner \bot \urcorner)$, so

$$S + \neg P(\ulcorner \bot \urcorner) \vdash P(\ulcorner \varphi(\dot{\vec{x}}) \urcorner) \to \neg P(\ulcorner \neg\varphi(\dot{\vec{x}}) \urcorner). \tag{I}$$

Since $S \vdash \neg\varphi(\vec{x}) \to P(\ulcorner \neg\varphi(\dot{\vec{x}}) \urcorner)$, we get $S \vdash \neg P(\ulcorner \neg\varphi(\dot{\vec{x}}) \urcorner) \to \varphi(\vec{x})$. By (I), we conclude $S + \neg P(\ulcorner \bot \urcorner) \vdash P(\ulcorner \varphi(\dot{\vec{x}}) \urcorner) \to \varphi(\vec{x})$. Then, $S + \neg P(\ulcorner \bot \urcorner) \vdash \vec{Q}\,\vec{x}.P(\ulcorner \varphi(\dot{\vec{x}}) \urcorner) \to \vec{Q}\,\vec{x}.\varphi(\vec{x})$. Consequently, $S \vdash \neg P(\ulcorner \bot \urcorner) \to \vec{Q}\,\vec{x}.\varphi(\vec{x})$. ⊣

Using the previous results, we are in condition of stating the characterization concerning Numeral Completeness and Numeral Consistency.

Theorem 5.5.1 (Characterization of Numeral Completeness)**.** *Let P be a provability predicate for T over a sub-theory S of T satisfying* P1–P3. *If $\vec{Q}\,\vec{x}.\varphi(\vec{x})$ is a sentence, with $\varphi(\vec{x})$ a $\Delta_1(S)$-formula, then the following statements are equivalent:*

I. $S \vdash \neg P(\ulcorner \bot \urcorner) \to \vec{Q}\,\vec{x}.\varphi(\vec{x})$;

II. $S \vdash \vec{Q}\,\vec{x}.P(\ulcorner \varphi(\dot{\vec{x}}) \urcorner)$.

Proof. Follows from the two previous Lemmata. ⊣

Theorem 5.5.2 (Characterization of Numeral Consistency)**.** *Let P be a provability predicate in the conditions of the previous result and Prf a $\Delta_1(S)$ proof predicate for a consistent theory. The following statements are equivalent:*

I. $S \vdash \exists x.Prf(\ulcorner \bot \urcorner, x) \to P(\ulcorner \bot \urcorner)$;

II. $S \vdash \forall x.P(\ulcorner \neg Prf(\ulcorner \bot \urcorner, \dot{x}) \urcorner)$.

Proof. Follows from the characterization of the Numeral Completeness. ⊣

We concluded that the provability predicates P which satisfy Numeral Completeness, for a sentence $\vec{Q}\,\vec{x}.\varphi(\vec{x})$, are exactly the ones for which it holds $S \vdash \neg P(\ulcorner \bot \urcorner) \to \vec{Q}\,\vec{x}.\varphi(\vec{x})$. The characterization of the consistency result was obtained from the previous characterization; more precisely, the provability predicates P that satisfy the mentioned result are the ones that satisfy $S \vdash \exists x.Prf(\ulcorner \bot \urcorner, x) \to P(\ulcorner \bot \urcorner)$.

5.6 Finitist Consistency

In this section, we give practical consequences of Numeral Completeness to the study of Finitist Consistency. We start with a simple observation[15].

Observation 5.6.1. Let us assume that $\varphi(x)$ is a $\Delta_1^b(\mathsf{S}_2^1)$-formula and τ a $\Delta_1^b(\mathsf{S}_2^1)$-numeration of S_2^1 in S_2^1. Then, $\mathsf{S}_2^1 \vdash \varphi(x) \to \exists y.\mathtt{Prf}_\tau(\ulcorner \varphi(\dot{x}) \urcorner, y)$, more concretely $\mathsf{S}_2^1 \vdash \forall x.\exists y.(\neg\varphi(x) \vee \mathtt{Prf}_\tau(\ulcorner \varphi(\dot{x}) \urcorner, y))$. As the formula $\neg\varphi(x) \vee \mathtt{Prf}_\tau(\ulcorner \varphi(\dot{x}) \urcorner, y)$ is $\Sigma_1^b(\mathsf{S}_2^1)$, we can conclude the existence of a polytime function f such that, for all $n \in \mathbb{N}$, $\mathsf{S}_2^1 \vdash \varphi(\overline{n}) \to \mathtt{Prf}_\tau(\ulcorner \varphi(\overline{n}) \urcorner, \overline{f(n)})$.

Proposition 5.6.1 (Polytime Finitist Consistency of S_2^1)**.** *Let Prf be a $\Delta_1^b(\mathsf{S}_2^1)$ proof predicate for a theory T. Then, there is a polytime function f such that, for each $n \in \mathbb{N}$, $f(n)$ is (the code of) a proof of $\neg Prf(\ulcorner \bot \urcorner, \overline{n})$ in S_2^1.*

Proof. Follows from the previous Observation. ⊣

Definition 5.6.1. We define the set $\mathtt{Tot}(T)$ to be the set of functions[16] whose graph can be given by a Σ_1-formula $\varphi(x, y)$ satisfying $T \vdash \forall x.\exists! y.\varphi(x, y)$.

We now consider *Kalmar elementary functions*. The next result confirms that the definable functions of EA are exactly the Kalmar elementary functions; for this reason, for the rest of this section, T is a theory including EA.

Fact 5.6.1 ([20, Fact 9][17])**.** *If* EA $\vdash \forall x.\exists! y.\psi(x, y)$ *with* $\psi(x, y)$ *a* Σ_1*-formula, then* $\psi(x, y)$ *is the graph of a Kalmar elementary function. Conversely, any Kalmar elementary function can be representable in* EA *by a* Σ_1*-formula. Thus,* $\mathtt{Tot}(\mathsf{EA})$ *is the set of Kalmar elementary functions.*

Definition 5.6.2. We define EA^* to be the theory EA extended by a function-symbol defining each Kalmar elementary function.

[15] This was kindly observed by Emil Jeřábek from a much more complex argument that we initially had.
[16] It is easy to generalize this definition for functions of different arities.
[17] See also [3].

With Numeral Completeness we can improve Observation 5.6.1 by internalizing the statement. We state it for EA^*, but the result can be adapted for other theories, like S^1_2.[18]

Theorem 5.6.1 ($\forall\Delta_1(\mathsf{EA})$-Witnessing)**.** *Let T be a $\Delta_1(\mathsf{EA})$-definable theory. If $\forall\vec{x}.\varphi(\vec{x})$ is true, with $\varphi(\vec{x})$ a $\Delta_1(\mathsf{EA})$-formula, then there are a numeration τ of T in* EA *and a function-symbol* f *such that* $\mathtt{Prf}_\tau$ *defines T-proofs in $\mathbb{N}$ and* $\mathsf{EA}^* \vdash \forall\vec{x}.\mathtt{Prf}_\tau(\ulcorner\varphi(\dot{\vec{x}})\urcorner, \mathsf{f}(\vec{x}))$.

Proof. From[19] Theorem 5.3.2, we know that $\mathsf{EA}^* \vdash \forall\vec{x}.\mathtt{Pr}_\tau(\ulcorner\varphi(\dot{\vec{x}})\urcorner)$, *videlicet* $\mathsf{EA}^* \vdash \forall\vec{x}.\exists y.\mathtt{Prf}_\tau(\ulcorner\varphi(\dot{\vec{x}})\urcorner, y)$. As Herbrand Analysis of [88] can be applied to EA^*, there is a function-symbol f such that $\mathsf{EA}^* \vdash \forall\vec{x}.\mathtt{Prf}_\tau(\ulcorner\varphi(\dot{\vec{x}})\urcorner, \mathsf{f}(\vec{x}))$. ⊣

Corollary 5.6.1 (Internal Finitist Consistency of EA)**.** *Let Prf be a $\Delta_1(\mathsf{EA})$ proof predicate for a theory T. Then there are a numeration τ of T in* EA *and a function-symbol* f *such that* $\mathtt{Prf}_\tau$ *defines T-proofs in $\mathbb{N}$ and* $\mathsf{EA}^* \vdash \forall x.\mathtt{Prf}_\tau(\ulcorner\neg Prf(\ulcorner\bot\urcorner, \dot{x})\urcorner, \mathsf{f}(x))$.

Proof. Immediate from the previous Theorem. ⊣

The next result shows that if Observation 5.6.1 would hold for $\Sigma^b_1(\mathsf{S}^1_2)$-formulas in general, then $\mathsf{P} = \mathsf{NP}$ (see Theorem 3.2.7 and Theorem 4.2.1 of [79] for results of a similar nature).

Theorem 5.6.2. *If for all $\Sigma^b_1(\mathsf{S}^1_2)$-formulas $\varphi(x)$ there is a polytime function f such that, for all $n \in \mathbb{N}$, $\mathbb{N} \models \varphi(\overline{n}) \to \mathtt{Prf}_\tau(\ulcorner\varphi(\overline{n})\urcorner, \overline{f(n)})$; then* $\mathsf{P} = \mathsf{NP}$.

Proof. Consider X a NP-complete set. From [10, p. 20], we know that there is a Σ^b_1-formula φ defining X, i.e. $n \in X \iff \mathbb{N} \models \varphi(\overline{n})$. Consider

$$g(n) := \begin{cases} 1, & f(n) \text{ is (the code of) a proof } \varphi(\overline{n}) \text{ in } \mathsf{S}^1_2; \\ 0, & \text{otherwise.} \end{cases}$$

As f is a polytime function and as we can decide in polytime if "x is a proof of y in S^1_2", we conclude that g is a polytime function.

Clearly, if $g(n) = 1$, then $f(n)$ is (the code of) a proof of $\varphi(\overline{n})$ in S^1_2, so $\mathsf{S}^1_2 \vdash \varphi(\overline{n})$, and hence $\mathbb{N} \models \varphi(\overline{n})$ (this step requires soundness). Furthermore, if $\mathbb{N} \models \varphi(\overline{n})$, then, as by assumption $\mathbb{N} \models \varphi(\overline{n}) \to \mathtt{Prf}_\tau(\ulcorner\varphi(\overline{n})\urcorner, \overline{f(n)})$, it follows $\mathbb{N} \models \mathtt{Prf}_\tau(\ulcorner\varphi(\overline{n})\urcorner, \overline{f(n)})$, i.e. $g(n) = 1$. In sum, $g(n) = 1 \iff \mathbb{N} \models \varphi(\overline{n}) \iff n \in X$. So, g is polytime and

$$g(n) = \begin{cases} 1, n \in X; \\ 0, n \notin X. \end{cases}$$

This entails $X \in \mathsf{P}$. ⊣

[18] Any theory allowing Sieg's Herbrand Analyses of [88] would work; in the case of S^1_2, we would need to use either some form of a collection principle or witnessing functions.

[19] This result uses the soundness of EA^*.

We would like to generalize the result for arbitrary extensions T of EA, instead of EA^*, unfortunately, we have no guarantee that T has function-symbols for each definable function. So we are able to state only a result similar to the *Polytime Finitist Consistency of* S^1_2 (Proposition 5.6.1).

As in [2], let $\mathsf{B}\Gamma$ be the principle

$$\forall x \leq a.\exists y.\psi(x,y) \to \exists z.\forall x \leq a.\exists y \leq z.\psi(x,y),$$

where $\psi(x,y)$ is a Γ-formula, possibly containing other parameters.

Fact 5.6.2 ([2, Corollary 4.1])**.** $\mathsf{EA}+\mathsf{B}\Sigma_1$ *is* Π_2*-conservative over* EA.

Proposition 5.6.2. *If* $\mathsf{EA}+\mathsf{B}\Sigma_1 \vdash \forall x.\exists! y.\psi(x,y)$ *with* $\psi(x,y)$ *a* Σ_1*-formula, then* $\psi(x,y)$ *is the graph of a Kalmar elementary function. Conversely, any Kalmar elementary function can be representable in* $\mathsf{EA}+\mathsf{B}\Sigma_1$ *by a* Σ_1*-formula. Thus,* $\mathtt{Tot}(\mathsf{EA}+\mathsf{B}\Sigma_1)$ *is the set of Kalmar elementary functions.*

Proof. Follows from the fact that $\forall x.\exists! y.\psi(x,y)$ is $\Pi_2(\mathsf{EA})$ and from the two previous facts. ⊣

Let us now state a useful technical result that uses a standard procedure in metamathematics.

Lemma 5.6.1. *Given a theory* T*, suppose that, for some* Σ_1*-formula* $\varphi(x,y)$*,* $T \vdash \forall x.\exists y.\varphi(x,y)$*. Then, there is a function* $f \in \mathtt{Tot}(T)$ *such that, for all* $n \in \mathbb{N}$*,* $T \vdash \varphi(\overline{n}, \overline{f(n)})$*.*

Proof. Consider $\varphi(x,y) := \exists z_1.\cdots\exists z_m.\psi(x,y,z_1,\ldots,z_m)$, with ψ a Δ_0-formula. Take

$$\begin{aligned}\varphi'(x,y) := \exists u.(y = (u)_0 \wedge \psi(x,(u)_0,\ldots,(u)_m) \wedge \\ \forall t < u.\neg\psi(x,(t)_0,\ldots,(t)_m)).\end{aligned}$$

The rest of the proof is a straightforward application of the definition to $\varphi'(x,y)$ exploiting the fact that EA has sequences functions and projections. ⊣

Corollary 5.6.2. *If* $\mathsf{EA}+\mathsf{B}\Sigma_1 \vdash \forall x.\exists y.\varphi(x,y)$*, with* $\varphi(x,y)$ *a* $\Sigma_1(\mathsf{EA})$*-formula, then there is a Kalmar elementary function* f *such that, for all* $n \in \mathbb{N}$*,* $\mathsf{EA} \vdash \varphi(\overline{n}, \overline{f(n)})$*.*

Proof. Follows immediately from the previous Lemma and from Proposition 5.6.2. ⊣

Theorem 5.6.4 will be a generalization of the *Polytime Finitist Consistency of* S^1_2 (Proposition 5.6.1). It will be corollary of the following theorem.

Theorem 5.6.3 (Proofs Witnessing $\Delta_1(T)$-truth)**.** *Let* $\varphi(x)$ *be a* $\Delta_1(T)$*-formula. There is a function* $f \in \mathtt{Tot}(T)$ *such that, for each* $n \in \mathbb{N}$*, if* $\mathbb{N} \models \varphi(\overline{n})$*, then* $f(n)$ *is* (*the code of*) *a* T*-proof of* $\varphi(\overline{n})$*.*

Proof. With Observation 5.6.1, adapted to the internal Σ_1-completeness of T, we have that, for τ a $\Sigma_1(T)$-numeration of T,

$$T \vdash \forall x.\exists y.(\neg\varphi(x) \vee \mathtt{Prf}_\tau(\ulcorner \varphi(\dot{x}) \urcorner, y)).$$

By Lemma 5.6.1, there is a function $f \in \mathtt{Tot}(T)$ such that, for all $n \in \mathbb{N}$,

$$T \vdash \varphi(\overline{n}) \to \mathtt{Prf}_\tau(\ulcorner \varphi(\overline{n}) \urcorner, \overline{f(n)}).$$

It follows that, if $\mathbb{N} \models \varphi(\overline{n})$, then $\mathbb{N} \models \mathtt{Prf}_\tau(\ulcorner \varphi(\overline{n}) \urcorner, \overline{f(n)})$. ⊣

Theorem 5.6.4 (General Finitist Consistency for T)**.** *Let Prf be a $\Delta_1(T)$ proof predicate of a (consistent) theory T. Then, there is a function $f \in \mathtt{Tot}(T)$ such that, for each $n \in \mathbb{N}$, $f(n)$ is (the code of) a T-proof of the sentence $\neg Prf(\ulcorner \bot \urcorner, \overline{f(n)})$.*

Proof. Follows immediately from the previous result. ⊣

The next result states that the construction of the "Proofs Witnessing $\Delta_1(T)$-truth" cannot be carried out inside T, in the sense that it cannot be codified by a single function $f(\#\varphi, n) \in \mathtt{Tot}(T)$.

Theorem 5.6.5. *There is* no *function $f \in \mathtt{Tot}(T)$ such that, for any Σ_1-formula $\varphi(x)$, $\mathbb{N} \models \varphi(\overline{n}) \implies \mathbb{N} \models \mathtt{Prf}_\tau(\ulcorner \varphi(\overline{n}) \urcorner, \overline{f(\#\varphi, n)})$.*

Proof. Suppose, aiming a contradiction, that there is such a function $f(\#\varphi, n) \in \mathtt{Tot}(T)$. Let R_f be a $\Sigma_1(T)$-formula representing f in T. By the Diagonalization Lemma, there is $\varphi(x)$ such that

$$T \vdash \varphi(x) \leftrightarrow \exists y.(\neg\mathtt{Prf}_\tau(\ulcorner \varphi(\dot{x}) \urcorner, y) \wedge R_f(\ulcorner \varphi \urcorner, x, y)).$$

So, $T \vdash \neg\varphi(\overline{n}) \to (R_f(\ulcorner \varphi \urcorner, \overline{n}, \overline{f(\#\varphi, n)}) \to \mathtt{Prf}_\tau(\ulcorner \varphi(\overline{n}) \urcorner, \overline{f(\#\varphi, n)}))$. As $T \vdash R_f(\ulcorner \varphi \urcorner, \overline{n}, \overline{f(\#\varphi, n)})$, $T \vdash \neg\varphi(\overline{n}) \to \mathtt{Prf}_\tau(\ulcorner \varphi(\overline{n}) \urcorner, \overline{f(\#\varphi, n)})$. If $\varphi(\overline{n})$ was false, it would follow $\mathtt{Prf}_\tau(\ulcorner \varphi(\overline{n}) \urcorner, \overline{f(\#\varphi, n)})$, and thus $T \vdash \varphi(\overline{n})$, a contradiction; consequently $\mathbb{N} \models \varphi(\overline{n})$. As φ is a $\Sigma_1(T)$-formula and T is Σ_1-complete, we conclude $T \vdash \varphi(\overline{n})$, hence $T \vdash \exists y.(\neg\mathtt{Prf}_\tau(\ulcorner \varphi(\overline{n}) \urcorner, y) \wedge R_f(\ulcorner \varphi \urcorner, \overline{n}, y))$. This entails $T \vdash \neg\mathtt{Prf}_\tau(\ulcorner \varphi(\overline{n}) \urcorner, \overline{f(\#\varphi, n)})$, which contradicts the assumption. ⊣

5.7 A general Negative Bound on Finitist Consistency Proofs

For a fixed Hilbert-style system, we use the notation $T \overset{n \text{ steps}}{\vdash} \varphi$ to express that there is a T-proof of φ whose number of lines is at most n; and the notation $T \overset{n}{\vdash} \varphi$ to express that there is a T-proof of φ whose size (length of the binary representation of the Gödel-number of the proof) is bounded by n.

In this section, we are going to study *bounded notions of provability*, i.e. formulas $B(x, y)$ such that $T \vdash \varphi \iff \exists n \in \mathbb{N}.T \vdash B(\ulcorner \varphi \urcorner, \overline{n})$. Clearly, $T \overset{k \text{ steps}}{\vdash}$ and $T \overset{n}{\vdash}$ are bounded notions of provability, after a suitable representation is made inside T; moreover, every proof predicate is trivially a bounded notion of provability[20]. Given B a bounded notion of provability, we define $T \mid\!\frac{n}{B}\, \varphi$ by $T \vdash B(\ulcorner\varphi\urcorner, \overline{n})$; and $\mathsf{Con}_B(x) := \neg B(\ulcorner\bot\urcorner, x)$.

We are going to generalize the reasoning of Theorem 4 from [12] (see also Theorem 3.1 of [76]), that was stated for $T \mid\!\frac{n}{}$, to any decidable (i.e., $\Delta_1(T)$) bounded notion of provability. Although this result is not a consequence of Numeral Completeness, it is related to Corollary 5.6.1. The latter is concerned to finitist consistency and was proved directly using Numeral Completeness. In a sense, this section is nothing more than an additional observation to what we have stated so far.

Theorem 5.7.1 (Negative Bound on Proofs of Consistency). *Suppose that the formula $B(x,y)$ is $\Delta_1(T)$ and that, for a fixed $\Sigma_1(T)$-numeration τ of T in T:*

E1. $T \vdash B(\ulcorner\varphi(\dot{x})\urcorner, y) \to \mathsf{Pr}_\tau(\ulcorner\varphi(\dot{x})\urcorner)$;

E2. *If $T \vdash \varphi$, then there is $c \in \mathbb{N}$ such that $T \vdash B(\ulcorner\varphi\urcorner, \overline{c})$;*

E3. *There is a function k, $\Sigma_1(T)$-representable in T by the term* k, *such that*

$$T \vdash B(\ulcorner\varphi(x)\urcorner, y) \to B(\ulcorner\varphi(\dot{x})\urcorner, \mathsf{k}(y,x));$$

E4. *For each $\Delta_1(T)$-formula $\varphi(x)$ there is a function $f_{\#\varphi}$, $\Delta_1(T)$-definable in T by a term $\mathsf{f}_{\#\varphi}$, such that $T \vdash \varphi(x) \to B(\ulcorner\varphi(\dot{x})\urcorner, \mathsf{f}_{\#\varphi}(x))$;*

E5. *There is a function g, $\Sigma_1(T)$-representable in T by the term* g, *such that*

$$T \vdash B(\ulcorner\varphi(\dot{x})\urcorner, a) \wedge B(\ulcorner\varphi(\dot{x}) \to \psi(\dot{x})\urcorner, b) \to \\ B(\ulcorner\psi(\dot{x})\urcorner, a + b + \mathsf{g}(\ulcorner\varphi\urcorner, \ulcorner\psi\urcorner, x));$$

E6. $T \vdash x \leq y \wedge B(z,x) \to B(z,y)$.

Suppose that j is any function, $\Sigma_1(T)$-representable in T by the term j, *such that that there is a term* t *in T satisfying the following property: for all fixed $n_0, n_1, n_2 \in \mathbb{N}$, for n sufficiently big*

$$T \vdash \mathsf{j}(\overline{n}) + \mathsf{k}(\overline{n_0}, \overline{n}) + \mathsf{g}(\overline{n_1}, \overline{n_2}, \overline{n}) \leq \mathsf{t}(\overline{n}).$$

Then, it is not *the case that*

$$T \vdash B(\ulcorner\mathsf{Con}_B(\mathsf{h}(\overline{n}))\urcorner, \mathsf{j}(\overline{n})), \quad n \in \mathbb{N},$$

i.e. it is not *the case that*

$$T \mid\!\frac{j(n)}{B}\, \mathsf{Con}_B(\mathsf{h}(\overline{n})), \quad n \in \mathbb{N},$$

with $\mathsf{h}(x) := \mathsf{t}(x) + \mathsf{f}_{n_0}(x) + \mathsf{g}(\overline{n_1}, \overline{n_2}, x)$, for certain fixed n_0, n_1, and n_2.

[20] Here the term 'bounded' is used in reference to the previously stated notions on the number of steps and symbols; further conditions on B could have been made, but we decided to state our results in a very general form.

Conditions E1 and E2 express the fact that B is a bounded notion of provability. E3 affirms that a feasible form of universal instantiation holds for B. Likewise, E4 expresses a form of $\Delta_1(T)$-completeness, and E5 a form of *modus ponens*. (In fact, a sufficient condition to satisfy E4 is, for $\Delta_1(T)$-formulas $\varphi(x)$, the $\Delta_1(T)$-completeness of the form $T \vdash \varphi(x) \to \exists y.B(\ulcorner\varphi(\dot{x})\urcorner, y)$.) The last condition, guarantees monotonicity in the sense of E6.

Proof of Theorem 5.7.1. Take $\varphi(x)$ given by the Diagonalization Lemma such that

$$T \vdash \varphi(x) \leftrightarrow \neg B(\ulcorner\varphi(\dot{x})\urcorner, \mathsf{t}(x)). \tag{I}$$

So, $T \vdash \neg\varphi(x) \to B(\ulcorner\varphi(\dot{x})\urcorner, \mathsf{t}(x))$; moreover, as $\neg\varphi(x)$ is a $\Delta_1(T)$-formula, we have by E4 that $T \vdash \neg\varphi(x) \to B(\ulcorner\neg\varphi(\dot{x})\urcorner, \mathsf{f}_{\#\neg\varphi}(x))$. From E5 it follows (assuming $\neg\varphi := \varphi \to \bot$)

$$T \vdash \neg\varphi(x) \to B(\ulcorner\bot\urcorner, \mathsf{t}(x) + \mathsf{f}_{\#\neg\varphi}(x) + \mathsf{g}(\ulcorner\varphi\urcorner, \ulcorner\bot\urcorner, x)).$$

Therefore,

$$T \vdash \mathsf{Con}_B(\mathsf{t}(x) + \mathsf{f}_{\#\neg\varphi}(x) + \mathsf{g}(\ulcorner\varphi\urcorner, \ulcorner\bot\urcorner, x)) \to \varphi(x).$$

Take $\mathsf{h}(x) := \mathsf{t}(x) + \mathsf{f}_{\#\neg\varphi}(x) + \mathsf{g}(\ulcorner\varphi\urcorner, \ulcorner\bot\urcorner, x)$. Suppose, aiming a contradiction, that, for each $n \in \mathbb{N}$, $T \vdash B(\ulcorner\mathsf{Con}_B(\mathsf{h}(\overline{n}))\urcorner, \mathsf{j}(\overline{n}))$. From before and E2, there is $c \in \mathbb{N}$ such that $T \vdash B(\ulcorner\mathsf{Con}_B(\mathsf{h}(x)) \to \varphi(x)\urcorner, \overline{c})$, so, by E3, it follows $T \vdash B(\ulcorner\mathsf{Con}_B(\mathsf{h}(\dot{x})) \to \varphi(\dot{x})\urcorner, \mathsf{k}(\overline{c}, x))$. Thus, again by E5, for $n \in \mathbb{N}$,

$$T \vdash B(\ulcorner\varphi(\overline{n})\urcorner, \mathsf{j}(\overline{n}) + \mathsf{k}(\overline{c}, \overline{n}) + \mathsf{g}(\ulcorner\mathsf{Con}_B(\mathsf{h}(x))\urcorner, \ulcorner\varphi(x)\urcorner, \overline{n})).$$

As, for n big enough,

$$T \vdash \mathsf{j}(\overline{n}) + \mathsf{k}(\overline{c}, \overline{n}) + \mathsf{g}(\ulcorner\mathsf{Con}_B(\mathsf{h}(x))\urcorner, \ulcorner\varphi(x)\urcorner, \overline{n}) \leq \mathsf{t}(\overline{n}),$$

we conclude $T \vdash B(\ulcorner\varphi(\overline{n})\urcorner, \mathsf{t}(\overline{n}))$. From E1, $T \vdash \varphi(\overline{n})$; but by (I), $T \vdash \neg\varphi(\overline{n})$, which is a contradiction. ⊣

Consider $\mathsf{Step}_y(x)$ as the formalization of $T \overset{k \text{ steps}}{\vdash}$. The next result contribute to the study of k-provability (complementing results which can be found, for instance, in [12]).

Theorem 5.7.2 (Negative Bound on the Number of Steps of Consistency)**.** *Let* j *represent any* $\Sigma_1(T)$*-function* j *and assume that* E4 *holds for* $\mathsf{Step}_y(x)$ *in* T *and that this formula is* $\Delta_1(T)$*. Then, there are* $n_0, n_1 \in \mathbb{N}$ *such that it is* not *the case that*

$$T \overset{j(n) \text{ steps}}{\vdash} \mathsf{Con}_{steps}(\mathsf{f}_{n_0}(\overline{n}) + \mathsf{j}(\overline{n}) + \overline{n_1}), \quad n \in \mathbb{N}$$

where $\mathsf{Con}_{steps}(x) := \neg\mathsf{Step}(\ulcorner\bot\urcorner, x)$.

Proof. Follows from the previous result by considering $\mathsf{g}(x, y, z) = \overline{1}$ (since an application of *modus ponens* simply corresponds to adding 1 to the total number of steps); $\mathsf{k}(x, y) = \overline{c}$, with $c \in \mathbb{N}$, (since an application of the universal instantiation increases the number of steps in a finite fixed way); and $\mathsf{t}(x) := \mathsf{j}(x) + \overline{1} + \overline{c}$. ⊣

PART B: Philosophical Analysis

5.8 Provable Consistency and the Uniformity of Incompleteness

In this first section of Part B, we show that provable consistency is, in fact, a nice feature to have; moreover, we confirm that the provability predicates we obtained in the proof of Numeral Completeness are intensionally sound. The statement of G1 has a uniform nature, since one can formulate it as follows:

> For every provability predicate P of a consistent extension of R, there is a formula φ such that $T \nvdash P(\ulcorner\varphi\urcorner)$ and $T \nvdash P(\ulcorner\neg\varphi\urcorner)$.

Naturally, one might wonder if an analogous formulation holds for G2, namely:

> Does it hold for every provability predicate P in the previous conditions, $T \nvdash \mathsf{Con}_P$?

The next result answers this question negatively.

Theorem 5.8.1 ([91, p. 841])**.** *Let P be any provability predicate for T, then $T \vdash \mathsf{Con}_{P^R}$ and $T \vdash \mathsf{Con}_{P^M}$.*

This means that, unlike G1, G2 does not have a uniform formulation for *every* provability predicate. Despite this fact, there are several works in the literature that aim to find general versions of G2, see [98] as an example of excellent work in that direction. Here one could trace the previous unexpected fact to a technical "trick" on the provability predicate, but the real question is: is it always a "trick" or is it always possible to prove some form of consistency; furthermore, does this form of consistency represent the "*actual*" one? One might be tempted to exclude provability predicates of the form P^R, since we have the following fact.

Theorem 5.8.2. *For all provability predicates P, there is* no *numeration τ of the axioms of T such that $T \vdash P^R(x) \leftrightarrow \mathsf{Pr}_\tau(x)$.*

Proof. We know that, for all τ, if φ is a Σ_1-formula, then $T \vdash \varphi \to \mathsf{Pr}_\tau(\ulcorner\varphi\urcorner)$; a property that does not hold for Rosser provability predicates (see Proposition 3.4 from [64]). ⊣

In contrast, we have the following fact established by Feferman (see 5.9 Theorem of [33]).

Theorem 5.8.3. *There is a $\Pi_1(T)$-numeration α of the axioms of T such that $T \vdash \mathsf{Con}_\alpha$.*

This result entails that the provability of consistency is not a particular feature of Rosser-like provability predicates. As Feferman observed in [33], one might be tempted, due to nice technical properties, to consider only $\Sigma_1(T)$-numerations of the axioms, but other numerations turned out to be very relevant for other uses:

> There is nothing "wrong" with the use of arbitrary formulas α; rather, the guiding consideration should be to investigate how different restrictions on the choice of α affect the results of arithmetization.

It is in this Fefermanian spirit that we proceed in our investigations.

We will exhibit now several examples of uniformities and desirable technical properties that are satisfied by provability predicates that prove a form of consistency, but that usually do not hold for other kinds of provability predicates. This serves to emphasize that provable consistency is, by no means, an undesirable technical feature.

Here is the first pertinent result. Under a very simple assumption, T proves that all provability predicates P that admit a form of consistency ($P(\ulcorner\neg\varphi\urcorner) \to \neg P(\ulcorner\varphi\urcorner)$), determine the same $\Delta_1(T)$-sentences in T.

Theorem 5.8.4. *Let P_0 and P_1 be provability predicates that satisfy*

- Σ_1-COMPLETENESS. *For each $\Sigma_1(T)$-sentence φ, $T \vdash \varphi \to P_i(\ulcorner\varphi\urcorner)$, for $i \in \{0,1\}$, and also*
- CONSISTENCY. *For all φ and $i \in \{0,1\}$, $T \vdash P_i(\ulcorner\neg\varphi\urcorner) \to \neg P_i(\ulcorner\varphi\urcorner)$.*

Then, for all $\Delta_1(T)$-sentences φ, $T \vdash P_0(\ulcorner\varphi\urcorner) \leftrightarrow P_1(\ulcorner\varphi\urcorner)$.

Proof. As φ is a $\Sigma_1(T)$-sentence, $T+\neg P_1(\ulcorner\varphi\urcorner) \vdash \neg\varphi$, and as $\neg\varphi$ is a $\Sigma_1(T)$-sentence, we have $T + \neg\varphi \vdash P_0(\ulcorner\neg\varphi\urcorner)$, hence $T + \neg P_1(\ulcorner\varphi\urcorner) \vdash P_0(\ulcorner\neg\varphi\urcorner)$. Using the hypothesis on consistency, $T + \neg P_1(\ulcorner\varphi\urcorner) \vdash \neg P_0(\ulcorner\varphi\urcorner)$, so $T \vdash P_0(\ulcorner\varphi\urcorner) \to P_1(\ulcorner\varphi\urcorner)$. One can prove the other implication in a similar way. ⊣

Observe that Pr_α from Theorem 5.8.3 satisfies the conditions of the previous result. We also have that, under the provable consistency assumption, a form of ω-consistency (see [91, p. 852] for more details on the definition) is also provable:

Theorem 5.8.5. *Let P be a provability predicate that satisfies*

- Σ_1-COMPLETENESS. *For every $\Sigma_1(T)$-formula $\varphi(x)$, $T \vdash \varphi(x) \to P(\ulcorner\varphi(\dot{x})\urcorner)$, and*
- CONSISTENCY. *For all $\varphi(x)$, $T \vdash P(\ulcorner\neg\varphi(\dot{x})\urcorner) \to \neg P(\ulcorner\varphi(\dot{x})\urcorner)$.*

Then, if $\varphi(x)$ is a $\Sigma_1(T)$-formula such that $\exists x.\varphi(x)$ is a $\Pi_1(T)$-sentence[21],

$$T \vdash \forall x.P(\ulcorner\neg\varphi(\dot{x})\urcorner) \to \neg P(\ulcorner\exists x.\varphi(x)\urcorner).$$

Proof. Clearly, $T+\forall x.P(\ulcorner\neg\varphi(\dot{x})\urcorner) \vdash \forall x.\neg P(\ulcorner\varphi(\dot{x})\urcorner)$. As $\varphi(x)$ is a $\Sigma_1(T)$-formula, $T \vdash \varphi(x) \to P(\ulcorner\varphi(\dot{x})\urcorner)$, so $T \vdash \neg P(\ulcorner\varphi(\dot{x})\urcorner) \to \neg\varphi(x)$. From before, $T + \forall x.P(\ulcorner\neg\varphi(\dot{x})\urcorner) \vdash \forall x.\neg\varphi(x)$, i.e. $T + \forall x.P(\ulcorner\neg\varphi(\dot{x})\urcorner) \vdash \neg\exists x.\varphi(x)$. As $\neg\exists x.\varphi(x)$ is a $\Sigma_1(T)$-sentence, we have $T \vdash \neg\exists x.\varphi(x) \to P(\ulcorner\neg\exists x.\varphi(x)\urcorner)$, and consequently $T + \forall x.P(\ulcorner\neg\varphi(\dot{x})\urcorner) \vdash P(\ulcorner\neg\exists x.\varphi(x)\urcorner)$, therefore we get $T + \forall x.P(\ulcorner\neg\varphi(\dot{x})\urcorner) \vdash \neg P(\ulcorner\exists x.\varphi(x)\urcorner)$, and thus $T \vdash \forall x.P(\ulcorner\neg\varphi(\dot{x})\urcorner) \to \neg P(\ulcorner\exists x.\varphi(x)\urcorner)$. ⊣

[21] This means that $\exists x.\varphi(x)$ is a $\Delta_1(T)$-sentence. We used this formulation to emphasize the $\varphi(x)$-part.

Provable consistency also gives rise to other kinds of uniformities: all provability predicates that prove their consistency statement are Mostowski provability predicates.

Theorem 5.8.6. *Let P be a provability predicate such that $T \vdash \mathsf{Con}_P$. Then, P is (provably equivalent to) a Mostowski provability predicate.*

Proof. It is a straightforward. ⊣

Moreover, all provability predicates that prove a form of a consistency statement—to wit $T \vdash P(\dot{\neg}x) \to \neg P(x)$—are also Rosser provability predicates.

Theorem 5.8.7. *Let $P(x) := \exists y.Prf(x,y)$ be a provability predicate that satisfies $T \vdash P(\dot{\neg}x) \to \neg P(x)$. Then P is (provably equivalent to) a Rosser provability predicate.*

Proof. It is not hard to see that

$$T \vdash Prf(x,y) \to (Prf(x,y) \wedge \forall z \leq y.\neg Prf(\dot{\neg}x,z)).$$

The result follows immediately from the previous equivalence. ⊣

Given any numeration ξ, the next Theorem guarantees that one can construct a numeration τ that proves its own consistency statement and such that Pr_τ entails, in T, the provability predicate Pr_ξ. Here we assume that PA is contained in T.

Theorem 5.8.8. *For every numeration ξ of the axioms of T including* PA, *there is a numeration τ such that $T \vdash \mathsf{Pr}_\tau(x) \to \mathsf{Pr}_\xi(x)$ and $T \vdash \neg P_\tau(\ulcorner \bot \urcorner)$.*

Proof. It is known that if T includes PA, then for every numeration θ, $T \vdash \mathsf{Pr}_{\mathsf{Pr}_\theta} \leftrightarrow \mathsf{Pr}_\theta$ (see [33]). Consider $\tau(v) := \alpha(v) \wedge \xi(v)$, with $\alpha(v)$ any numeration such that $T \vdash \neg\mathsf{Pr}_\alpha(\ulcorner \bot \urcorner)$ (use, for instance, the one by Feferman). Clearly, $T \vdash \tau(x) \to \xi(x)$, and so, $T \vdash \tau(x) \to \mathsf{Pr}_\xi(x)$. Thus, $T \vdash \mathsf{Pr}_\tau(x) \to \mathsf{Pr}_{\mathsf{Pr}_\xi}(x)$, and so by what was initially observed, $T \vdash \mathsf{Pr}_\tau(x) \to \mathsf{Pr}_\xi(x)$. Furthermore, $T \vdash \tau(x) \to \alpha(x)$, so $T \vdash \mathsf{Pr}_\tau(x) \to \mathsf{Pr}_\alpha(x)$. As $T \vdash \neg\mathsf{Pr}_\alpha(\ulcorner \bot \urcorner)$, it follows that $T \vdash \neg P_\tau(\ulcorner \bot \urcorner)$. ⊣

Finally, we give a proof that if one considers provability predicates that prove the consistency statement $P(\ulcorner \neg\varphi \urcorner) \to \neg P(\ulcorner \varphi \urcorner)$, for each formula φ, one can prove a form of G1 without the need for further assumptions (like ω-consistency).

Theorem 5.8.9. *Let P be a provability predicate that satisfies*

- *Consistency. For all formulas φ, $T \vdash P(\ulcorner \neg\varphi \urcorner) \to \neg P(\ulcorner \varphi \urcorner)$.*

Then it does not *hold that*

- *Completeness. For all formulas φ, $T \vdash \varphi \to P(\ulcorner \varphi \urcorner)$.*

Proof. Suppose, aiming at a contradiction, that the mentioned condition holds. Consider $\mathcal{G}_T$ given by the Diagonalization Lemma such that $T \vdash \mathcal{G}_T \leftrightarrow \neg P(\ulcorner \mathcal{G}_T \urcorner)$. Then, assuming completeness, $T \vdash \mathcal{G}_T \rightarrow \neg P(\ulcorner \mathcal{G}_T \urcorner) \wedge P(\ulcorner \mathcal{G}_T \urcorner)$, and thus $T \vdash \neg \mathcal{G}_T$. Hence, $T \vdash P(\ulcorner \neg \mathcal{G}_T \urcorner)$; from the assumption, it follows that $T \vdash \neg P(\ulcorner \mathcal{G}_T \urcorner)$. By construction of the Diagonalization Lemma, $T \vdash P(\ulcorner \mathcal{G}_T \urcorner)$. Thus, $T \vdash \bot$; which contradicts the assumption we made on T regarding its consistency (because T needs to be a true theory of arithmetic). $\dashv$

Let us end this section with a remark on intensionality. One possible reaction to what we have established so far would be to say that the provability predicates we are considering are not *intensional*, i.e. they do not fully express the *meaning* of provability in T or even the meaning of the consistency statements. In [70] one can find a profound discussion of intensionality in metamathematics. We do not wish to enter into this discussion in detail since the goal of this chapter is to present metamathematical results in line with a form of HP (Hilbert's Program), not to enter a meaning/intensional discussion. Nevertheless, we would like to draw the reader's attention to several facts.

One might agree that P^R has very strange technical features in order to prove incompleteness and one might even question to what extent it is really "the" provability predicate. However, one cannot ignore the fact that one really needs it to prove G1 without the use of ω-consistency (this is a strong indication of the relevance of P^R)—see [91, p. 841]. More generally, one cannot ignore the technical uniformities and desirable results that arise from the use of provability predicates that prove some form of consistency. Of course, Rosser provability predicates are, indeed, provability predicates, one cannot simply deny it. Despite this fact, when one uses them in a negative way, i.e., $\neg P^R$, one has no guarantee that the meaning is preserved; the reader should keep in mind that $\mathsf{EA} \vdash \neg P^R(\ulcorner \bot \urcorner)$ might be provable even for the case where P^R is a provability predicate for an inconsistent theory—one just need to satisfy that $\neg \bot$ has a proof that is smaller than the smallest proof of $\bot$. The same happens for the Mostowski predicate. From this example, we conclude that there is, in general, no problem when provability predicates are used in a positive way—there they simply fulfil their representation-job—problems might arise when they are used negatively.

Our approach is not problematic from an intensional perspective because we do not use provability predicates in a negative way. In our Numeral Completeness result, $\mathsf{EA} \vdash \vec{Q}\vec{x}.\mathsf{Pr}_\tau(\ulcorner \varphi(\dot{\vec{x}}) \urcorner)$ (when $\vec{Q}\vec{x}.\varphi(\vec{x})$ is true in $\mathbb{N}$), the provability predicate is used in a positive way in the same way as in our Numeral Consistency result, that is $\mathsf{EA} \vdash \forall x.\mathsf{Pr}_\tau(\ulcorner \neg Prf(\ulcorner \bot \urcorner, \dot{x}) \urcorner)$. Furthermore, the provability predicate used in the consistency result is a $\Sigma_1(\mathsf{EA})$-provability predicate, i.e., it is a standard provability predicate. These predicates are commonly accepted as intensionally sound even for more skeptical authors; in fact, according to [70], a provability predicate needs to be Σ_1 in order to be considered intensional. All this fits with Feferman's quote on the use of different numerations, not necessarily Σ_1, for metamathematical studies.

5.9 Hilbert's Programs and Finitist Consistency

Hilbert's various forms of a consistency program expressed one of the fundamental approaches to the foundations of mathematics in the 20$^{\text{th}}$ century. In 1922, Hilbert formulated his famous proposal to solve the consistency problem for mathematics by focusing on its formalization and restricting the metamathematical methods to finitist ones. The underlying notions of formal proof and syntactic consistency had been developed in the Hilbert School in [48] (a published version can be found in [30, pp. 31–270]). With respect to the axiomatic method that did not yet appeal to a rigorous formal language and a logical calculus, Hilbert had observed already in 1902 [40, p. 540],

> Every science takes its starting point from a sufficiently coherent body of facts as given. It takes form, however, only by *organizing* this body of facts. This organization takes place through the *axiomatic method*, i.e., one constructs a *logical structure of concepts* so that the relationships between the concepts correspond to relationships between the facts to be organized.
> There is arbitrariness in the construction of such a structure of concepts; we, however, *demand* of it: 1) completeness, 2) independence, 3) consistency.

There is no reason to believe that Hilbert intended at this point completeness in the syntactic way we understand it today; indeed, lacking a formal language, that modern concept could not even be properly formulated. Sieg observed in [89, p. 87] about Hilbert's ideas in 1900,

> [T]he axiomatic method has to confront two fundamental problems that are formulated in [Hilbert's Paris Lecture of 1900] at first for geometry and then also for arithmetic:
>
> > The necessary task then arises of showing the consistency and the completeness of these axioms, i.e., it must be proved that the application of the given axioms can never lead to contradictions, and, further, that the system of axioms suffices to prove all geometric propositions.
>
> It is not clear, whether completeness of the respective axioms requires the proof of all true geometric or arithmetic statements, or whether provability of those that are part of the established corpora is sufficient; the latter would be a quasi-empirical notion of completeness.

These fundamental ideas can be traced back to Dedekind (see [89, p. 95]). They are formulated most explicitly in a famous letter of Dedekind to Keferstein [41, p. 101].

Throughout the rest of the chapter we will refer by HP to the *finitist consistency program*. For more details on HP we refer to [60], [89], and [105]. There is consensus that the finitist methods include the mathematical principles of PRA, see [87]. From the finitist proof of the consistency of PRA sketched by Hilbert in [47] (the reader can find an English translation in [29, pp. 1136–1148]), we know that finitist reasoning is strictly stronger than PRA. Nevertheless, we use EA as the basic theory for our considerations. It is a sub-theory of PRA.

Our results state a form of completeness that partially fulfils Hilbert's goals from late 1930 and mid-1931 in [46] and [45]. In the following remark from [89, pp. 160–161] regarding HP after the incompleteness results, $\mathfrak{A}$ is a quantifier-free, finitistically meaningful formula:

> Hilbert's rule (HR) is viewed as finitist and allows the introduction of universally quantified formulae $(x)\mathfrak{A}(x)$ as initial ones, just in case the numeric instances $\mathfrak{A}(\mathfrak{z})$ have been established finitistically as correct for arbitrary numerals $\mathfrak{z}$.

This idea was formulated in [46] and then generalized by Hilbert in [45], where $\mathfrak{A}$ is no longer required to be quantifier-free, but can be a formula of arbitrary syntactic complexity.

> If the statement $\mathfrak{A}(\mathfrak{z})$ is correct as soon as $\mathfrak{z}$ is a numeral, then the statement $(x)\mathfrak{A}(x)$ holds; in this case $(x)\mathfrak{A}(x)$ is called correct.

For this version of HP, it suffices to confirm all the numerical instances in a broader methodological frame in order to conclude the correctness of the universally quantified statement.[22]

In our discussion we introduce an informal term, '*codification of a formula* φ' or simply '*codification of* φ'. We are using this term with the broad meaning of "expressing the considered sentence φ". Intuitively speaking, a formula φ^* *codifies* φ if φ and φ^* express the same content instance-wise. For example, $\forall x.\psi(x)$ and $\forall x.\mathsf{Pr}_\tau(\ulcorner\psi(\dot{x})\urcorner)$, with $\psi(x)$ a Σ_1-formula, codify each-other when τ is a numeration of a Σ_1-complete and Σ_1-sound theory. In semantic terms, one statement cannot hold instance-wise without the other one doing so as well. As the *extreme trivial case*, any formula codifies itself. The general notion of codification is an intuitive one; we emphasize that the results of this chapter do not depend on it, as we considered only the particular codifications $\vec{Q}\vec{x}.\mathsf{Pr}_\tau(\ulcorner\psi(\dot{\vec{x}})\urcorner)$.

With this concept of "codification" one can articulate a more general form of HP:

1. If φ is a true sentence and φ^* is a codification of φ, then φ^* is provable in a fixed theory of arithmetic.

[22] For the immediate historical context, see [89, pp. 174–175]. There is, first of all, the recognition formulated by Bernays in a letter to Hilbert that was written on October 11, 1931; namely, that this *is* an extension of the finitist standpoint. Then there is, secondly, the fact that Gentzen intended to complete Hilbert's consideration in late 1931 and early 1932 in order to obtain a consistency proof of full arithmetic (in his "Urdissertation"). HR should not be confused with the ω-rule in semi-formal systems. For a historical discussion of the relation of HR and the ω-rule see [53]. Thirdly, in Herbrand's last paper [43], the consistency of a theory is established that includes the restricted form of HR.

2. If Con is a standard consistency statement for a consistent theory and Con^* is a codification of that statement, then Con^* is provable in a fixed theory of arithmetic.

The Incompleteness Theorems show that the goals 1 and 2 cannot be fulfilled, as they present witnesses of the falsity of the stated universal claims. The First Incompleteness Theorem witnesses the falsity of 1: there is a sentence such that its trivial codification is not provable in the considered theory of arithmetic. The Second Incompleteness Theorem falsifies 2 with the trivial codification of the standard consistency statement. But what about the existential statements corresponding to 1 and 2? Are they also necessarily false on account of the Incompleteness phenomenon? More precisely, are the following intuitive goals achievable:

1'. For every true sentence φ, there is a codification φ^* provable in a fixed theory of arithmetic.

2'. There is a formula Con^* that codifies the consistency statement for a given consistent theory and is provable in a fixed theory of arithmetic.

We proved that these goals can be achieved. Hilbert's work has been interpreted and adapted to today's logical framework in the very strong form 1 and 2, even after Gödel's Incompleteness Theorems had been found. If one interprets HP as 1' and 2', then it is, in a sense that we will make precise, achievable. Thus, our analysis complements the Incompleteness results by a dual counterpart: while the Incompleteness Theorems state the falsity of the universal statements of 1 and 2, we established the existential statements 1' and 2'. These facts do not contradict but rather complement each other. A more detailed account of the notion 'codification' falls under the scope of intensionality. That is not the focus of our chapter and is not essential for understanding our results and their connection to HP.

We proved that, for every formula $\vec{Q}\vec{x}.\varphi(\vec{x})$ that is true in $\mathbb{N}$, there is a numeration τ of the axioms of EA such that $\mathsf{EA} \vdash \vec{Q}\vec{x}.\mathsf{Pr}_\tau(\ulcorner\varphi(\dot{\vec{x}})\urcorner)$. This constitutes a (weak) form of syntactic completeness because we created a syntactic interpretation for each true formula: from the given true formula $\vec{Q}\vec{x}.\varphi(\vec{x})$, one obtains directly its codification $\vec{Q}\vec{x}.\mathsf{Pr}_\tau(\ulcorner\varphi(\dot{\vec{x}})\urcorner)$ which states that under the scope of $\vec{Q}\vec{x}$, EA can internally and numeral-wise prove $\varphi(\vec{x})$. These ideas make use of the following definition; namely, a function f is *truth-preserving* if whenever φ is a true formula in $\mathbb{N}$, so is the formula whose code is $f(\#\varphi)$.

Theorem 5.9.1. *Given a sound theory T and a numeration τ of T:*

1. *The following is* false: *for every truth-preserving function f definable in T, if φ is a true formula, then $T \vdash \mathsf{Pr}_\tau(\overline{f(\#\varphi)})$.*

2. *There is a truth-preserving function f definable in T such that if φ is a true formula, then $T \vdash \mathsf{Pr}_\tau(\overline{f(\#\varphi)})$.*

Proof. 1 follows from G1 by considering f as the identity function, and 2 follows from our Numeral Completeness result, since the function f such that

$$f(\#[\vec{Q}\vec{x}.\varphi(\vec{x})]) := \#[\vec{Q}\vec{x}.\mathsf{Pr}_\tau(\ulcorner\psi(\dot{\vec{x}})\urcorner)],$$

with τ from the mentioned result, is definable in T and is truth-preserving. ⊣

We used the formula $\vec{Q}\vec{x}.\mathsf{Pr}_\tau(\ulcorner\varphi(\dot{\vec{x}})\urcorner)$ to provably codify the fact that the formula $\vec{Q}\vec{x}.\varphi(\vec{x})$ is true. That is interesting in its own right, but is especially relevant since we proved completeness for this codification. In sum, the First Incompleteness Theorem guarantees that some formula $\vec{Q}\vec{x}.\varphi(\vec{x})$ is not provable, whereas our result ensures its codification $\vec{Q}\vec{x}.\mathsf{Pr}_\tau(\ulcorner\varphi(\dot{\vec{x}})\urcorner)$ is. Notice that this fact holds for a suitable numeration τ, and that τ depends on the initial formula: for each formula such a numeration is guaranteed to exist, but there is no single numeration for all formulas.

We formulated in an analogous way a codification of consistency that we called Numeral Consistency: for every $\Delta_1(\mathsf{EA})$-proof predicate *Prf* for a consistent theory T, there is a $\Sigma_1(\mathsf{EA})$-numeration τ of the axioms of EA such that $\mathsf{EA} \vdash \forall x.\mathsf{Pr}_\tau(\ulcorner\neg Prf(\ulcorner\bot\urcorner, \dot{x})\urcorner)$. The latter claim asserts, EA can prove, numeral-wise and under a Σ_1-provability predicate, that no natural number is the code of a proof of $\bot$ according to the proof predicate *Prf*. Keep in mind that establishing a property using Feferman's dot notation is enough to guarantee that it holds for all natural numbers. These two facts—Numeral Completeness and Numeral Consistency—yield the fulfilment of a form of HP when articulated by 1' and 2'.

The connection of the Completeness result with HP has yet another facet, namely, one can show that a codified version of HR is provable. Let us analyze this carefully. Assume that $\mathfrak{A}(x)$ is a quantifier-free formula that has been established finitistically for each numeral $\mathfrak{z}$. From the Completeness result, it follows, for a suitable numeration τ, $\mathsf{EA} \vdash \forall x.\mathsf{Pr}_\tau(\ulcorner\mathfrak{A}(\dot{x})\urcorner)$. Thus, EA—a finitist theory—can, from the premises of HR, confirm that $\forall x.\mathfrak{A}(x)$ is a formula that must hold for the natural numbers[23]. Our form of HR can be summarized in the following way: given a quantifier-free formula $\mathfrak{A}$,

$$\frac{\mathfrak{A}(\mathfrak{z}),\ \text{for all numerals } \mathfrak{z}}{\exists\!\exists \tau.\mathsf{EA} \vdash \forall x.\mathsf{Pr}_\tau(\ulcorner\mathfrak{A}(\dot{x})\urcorner)}$$

where $\exists\!\exists$ represents the existential quantifier in the meta-language[24]—these considerations can also be carried through for the unrestricted form of HR.

[23]Here one clearly makes an additional assumption, namely that the premises of HR are established as correct in a finitist theory T (and that τ numerates T).

[24]The numeration τ may be different for different formulas $\mathfrak{A}$.

Chapter 6

Provability Implies Provable Provability

6.1 Introduction

The derivability condition 'provability implies provable provability', namely $\mathsf{Pr}_\xi(x) \rightarrow \mathsf{Pr}_\xi(\ulcorner \mathsf{Pr}_\xi(\dot{x}) \urcorner)$, is very sensitive to the basis theory one is considering. In this paper, we study it using the computational complexity class FLINSPACE. Let us now make some introductory remarks.

$\mathcal{E}^2$ is a class of functions over $\mathbb{N}$ defined by the closure of a set of initial functions under composition and bounded recursion. It belongs to a hierarchy of classes $\mathcal{E}^n$ introduced by Grzegorczyk in [38] and that characterizes the primitive recursive functions [19, pp. 458, 459]; Sieg in [88] calls it the *G*-hierarchy. Grzegorczyk's class $\mathcal{M}^2$ has a similar definition, but for it one substitutes bounded recursion for closure under *bounded minimization* ($\mu_{\leq}$) (see [17, Definition 3.47] and [23]); like $\mathcal{E}^2$, the class $\mathcal{M}^2$ belongs to a hierarchy of functions $\mathcal{M}^n$. For $n \geq 3$, $\mathcal{M}^n = \mathcal{E}^n$ [19, Theorem 6.3.20], but for $n = 2$ this is still an open problem (*conferatur* [19, p. 460], [81, p. 110]). $\mathcal{E}^2$ and $\mathcal{M}^2$ correspond to FLINSPACE (linear space) and FLTH (linear time hierarchy), respectively. It is also known that the definable functions of the theory $\mathsf{I}\Delta_0$ are exactly the functions in $\mathcal{M}^2$ [23].

We develop a theory of arithmetic for FLINSPACE (the same as the class $\mathcal{E}^2$) that we call G_2 (the number '2' in 'G_2' stands for the '2' in '$\mathcal{E}^2$', and the letter 'G' stands for 'Grzegorczyk'), clearly not to be confused with Gödel's Second Incompleteness Theorem; 'G' is also used in reference to the name *G*-hierarchy from Sieg's [88]. G_2 is similar to the theory $\mathsf{I}\mathcal{E}^2_*$ of Woods from [103] (see also [26] and [5]), but the basis functions are not the same (we have the 0 and 1 functions, as well as projections in the basis); moreover, we account for subtleties in the definition of the bounded recursion schema, namely in giving a name to the created functions and in the way we state the schema using the `min` function—usually, to define $\mathsf{I}\mathcal{E}^2_*$, one adds a function-symbol to each function in $\mathcal{E}^2$ and the respective definition axiom, but in our context, we explicitly develop a particular way of doing so that is going to be useful to establish, for instance, Theorem 6.3.3 and that to guarantee that G_2 is recursively enumerable. In a sense, G_2 is a specific way to construct a version of $\mathsf{I}\mathscr{E}^2_*$. The latter, and consequently G_2, is a very useful theory to study, in a weak arithmetic setting, several results from Number Theory; the key feature is that $\mathsf{I}\mathscr{E}^2_*$ allows a form of counting, see [24] for an example of such an approach, where Cornaros studies the quadratic reciprocity law using that theory.

The fact that we develop a theory for FLINSPACE that includes $\mathsf{I}\Delta_0$, like $\mathsf{I}\mathscr{E}^2_*$, is going to be useful. This feature does not hold for some of the theories one can find in the literature, for instance the theory for FLINSPACE from [96]; this feature, and the way we name the $\mathscr{E}^2$-functions, is especially useful to establish sufficient conditions for relations between $\mathscr{M}^2$ and $\mathscr{E}^2$ (see Theorem 6.3.3). As it is unknown whether $\mathscr{M}^2 = \mathscr{E}^2$, it is *a forteriori* an open problem whether G_2 is a conservative extension of $\mathsf{I}\Delta_0$.

We follow closely some of the constructions made in [34] for the theory PTCA; we, in addition, use some definitions and results from [23].

After a quick bootstrapping[1] of G_2, we study the main focus of our chapter, namely "provability implies provable provability" in G_2: $\mathtt{Pr}_\xi(x) \to \mathtt{Pr}_\xi(\ulcorner \mathtt{Pr}_\xi(\dot{x}) \urcorner)$. This conditions is very sensitive in terms of the basis theory, especially if it is weak; for example, it is in general an open problem for $\mathsf{I}\Delta_0$. Our approach is relevant because G_2 is an example of such a weak theory that might even be a conservative extension of $\mathsf{I}\Delta_0$ (this is still an open problem also, since it entails $\mathscr{M}^2 = \mathscr{E}^2$). In a sense, in this chapter, we use the complexity class FLINSPACE to study the mentioned derivability condition.

We study forms of internal Σ_1-completeness for G_2. We prove that if G_2 can verify its axioms, in the sense that, for a suitable G_2-function `verifier`, $\mathsf{G}_2 \vdash \xi(x) \to \mathtt{Prf}_\xi(\ulcorner \mathtt{Pr}_\xi(\dot{x}) \urcorner, \mathtt{verifier}(x))$, then $\mathsf{G}_2 \vdash \mathtt{Pr}_\xi(x) \to \mathtt{Pr}_\xi(\ulcorner \mathtt{Pr}_\xi(\dot{x}) \urcorner)$. Moreover, we confirm that the `verifier` condition can be dropped if one focuses on just a finite set of axioms (here we include also a finite number of logical axioms): if $\mathtt{Pr}^{\mathcal{S}}(x)$ is a provability predicate for the finite set $\mathcal{S}$, then $\mathsf{G}_2 \vdash \mathtt{Pr}^{\mathcal{S}}(x) \to \mathtt{Pr}_\xi(\ulcorner \mathtt{Pr}_\xi(\dot{x}) \urcorner)$.

Finally, we present conditions for a numeration ξ_0 of a finitely axiomatizable theory to satisfy $\mathsf{G}_2 \vdash \mathtt{Pr}_{\xi_0}(x) \to \mathtt{Pr}_\xi(\ulcorner \mathtt{Pr}_{\xi_0}(\dot{x}) \urcorner)$.

6.2 The Grzegorczyk class $\mathscr{E}^2$

We give two alternative definitions of the Grzegorczyk class $\mathscr{E}^2$.

Definition 6.2.1. (1) $\mathscr{E}^2$ is the class of functions over $\mathbb{N}$ that includes $0, 1, +, \times$,

[1]We use this term in reference to [10, p. 37].

$\mathtt{min}, \mathtt{max}, \pi^k_j$, that is closed for composition, and for *bounded recursion*: $f(\vec{m}, n)$ is defined by bounded recursion from g, h, k if $f(\vec{m}, n) \leq k(\vec{m}, n)$ and:

$$f(\vec{m}, 0) := g(\vec{m}); \qquad \text{(BR)}$$
$$f(\vec{m}, n+1) := h(\vec{m}, n, f(\vec{m}, n)).$$

(2) Alternatively, $\mathcal{E}^2$ coincides with the class that includes the functions $0, 1, +, \times, \mathtt{min}, \mathtt{max}, \pi^k_j$, that is closed for composition, and that is closed under the following form of bounded recursion: given f, g, and k already in the class, then the following function also is in the class

$$f(\vec{m}, 0) := \mathtt{min}(g(\vec{m}), k(\vec{m}, 0)); \qquad \text{(BRmin)}$$
$$f(\vec{m}, n+1) := \mathtt{min}(h(\vec{m}, n, f(\vec{m}, n)), k(\vec{m}, n+1)).$$

Item (1) corresponds to the "classic" definition of $\mathcal{E}^2$, see [25], and (2) is, essentially, the caractherization of $\mathcal{E}^2$ presented in [5]. Clearly, we can omit the min function in the first equation of BRmin, we could have written $f(\vec{m}, 0) := g(\vec{m})$; we decided to still have it because it immediately gives the bound k for f. The item (2) of the Definition above plays a main role in our work because it yields a recursively enumerable theory (G_2) and because it allows to establish facts that are similar to the ones from [34]. Each part of the definition of our theory G_2 corresponds to a part of this characterization of $\mathcal{E}^2$. The next result is of major importance.

Fact 6.2.1. $\mathcal{E}^2 = \mathsf{FLINSPACE}$.

Proof. See [23] (we recommend [19, pp. 461, 469] for further details). ⊣

FLINSPACE is the class of functions computable in linear space by deterministic Turing Machines. We consider Turing Machines computing functions over $0-1$ words, while $\mathcal{E}^2$ is a class of functions over $\mathbb{N}$. So, for a function $f \in \mathsf{FLINSPACE}$ we write $f \in \mathcal{E}^2$ (and vice-versa) under the standard bijection from $0-1$ words and $\mathbb{N}$. This is used extensively along the chapter, without further notice.

Theorem 6.2.1. *For each function $f \in \mathcal{E}^2$, we can construct a polynomial p_f such that $f \leq p_f$ (this statement is also provable in the theory G_2 that we are going to develop).*

Proof. Immediate by induction using the definition of $\mathcal{E}^2$. ⊣

The following observation confirms that the function that sums inputs is also in $\mathcal{E}^2$.

Observation 6.2.1. Suppose that f is in $\mathcal{E}^2$. Let us see that the function $\mathtt{sum}_f$ defined by

$$\begin{cases} \mathtt{sum}_f(\vec{x}, 0) := f(\vec{x}, 0); \\ \mathtt{sum}_f(\vec{x}, y+1) := \mathtt{sum}_f(\vec{x}, y) + f(\vec{x}, y+1) \end{cases}$$

is also in $\mathscr{E}^2$. $\mathtt{sum}_f$ can be defined by recursion, the challenge is to ensure the bound. From Theorem 6.2.1, we know that $f \leq p_f$, so

$$\mathtt{sum}_f(\vec{x},y) = \sum_{i=0}^{y} f(\vec{x},i) \leq \sum_{i=0}^{y} p_f(\vec{x},i).$$

Let us write p_f in the following way (this is always possible to perform)

$$p_f(\vec{x},i) = \sum_{m=0}^{\ell} p_m(\vec{x}) \times i^{k_m},$$

where p_m are polynomials on $\vec{x}$, and k_m are suitable constants. So,

$$\begin{aligned}\sum_{i=0}^{y} f(\vec{x},i) &\leq \sum_{i=0}^{y}\sum_{m=0}^{\ell} p_m(\vec{x}) \times i^{k_m} \leq \sum_{m=0}^{\ell}\sum_{i=0}^{y} p_m(\vec{x}) \times i^{k_m} \\ &\leq \sum_{m=0}^{\ell}\left(p_m(\vec{x}) \times \sum_{i=0}^{y} i^{k_m}\right).\end{aligned}$$

From [16], we know that the inner polynomials are Faulhaber polynomials on $\mathbb{Q}$, more specifically

$$\sum_{i=0}^{y} i^{k_m} = \frac{1}{k_m+1}\sum_{j=0}^{k_m+1}\binom{k_m+1}{j}(-1)^{k_m+1-j} \times B_{k_m+1-j} \times y^j,$$

where B_j denotes the j-th Bernoulli number. So, the previous polynomial can be bounded by a polynomial $q_m(y)$ with coefficients on $\mathbb{N}$, yielding $\sum_{i=0}^{y} i^{k_m} \leq q_m(y)$, and consequently

$$\mathtt{sum}_f(\vec{x},y) = \sum_{i=0}^{y} f(\vec{x},i) \leq \sum_{m=0}^{\ell} p_m(\vec{x}) \times q_m(y),$$

which implies that $\mathtt{sum}_f$ is in $\mathscr{E}^2$.

6.3 A theory for FLINSPACE

The theory that we are going to develop here is a theory in first-order classical logic over a language $\mathscr{L}$ which extends $\{0, \mathsf{S}, +, \times, =, \leq\}$ by allowing further function-symbols. Let us settle some well-known definitions.

Definition 6.3.1. 1.($\Delta_0(\mathscr{L})$-formulas) $\Delta_0(\mathscr{L})$-formulas, also called *bounded formulas*, are defined recursively. If t_0 and t_1 are terms of $\mathscr{L}$, then $t_0 = t_1$ and $t_0 \leq t_1$ are bounded formulas; and if A and B are are bounded formulas and t is a term, then so are: $\neg A$, $A \wedge B$, $\forall z \leq t.A$, and $\exists z \leq t.A$ (A can have free variables besides z, and it can also be a sentence).[2]

[2]We are assuming, without loss of generality, $\{\neg, \wedge, \forall, \exists\}$ as the logical basis and that $\vee$, $\rightarrow$ are defined accordingly.

2.($\Sigma_1(\mathscr{L})$-formulas) A formula A is a $\Sigma_1(\mathscr{L})$-formula if it is of the form $\exists\vec{y}.B$, with B a bounded formula (possibly with several free variables).[3]

3.($\Gamma(T)$-formula) A formula φ is a $\Gamma(T)$-formula if there is a Γ-formula φ_0 that is T-equivalent to φ (we tacitly consider the language $\mathscr{L}$ of the theory T).

We omit the $\mathscr{L}$ whenever there is no danger of ambiguity.

For the next definition, we are following [23].

Definition 6.3.2. Let T be a Σ_1-sound theory in $\mathscr{L}$. A function $f:\mathbb{N}^n \to \mathbb{N}$ is *definable in* T if there is a $\Sigma_1(\mathscr{L})$-formula $\varphi(\vec{x},y)$ of T such that:

R1: $T \vdash \forall\vec{x}.\exists!\, y.\varphi(\vec{x},y)$;

R2: For all $a_1,\ldots,a_k,b \in \mathbb{N}$, $f(\vec{a}) = b \iff \mathbb{N} \models \varphi(\overline{\vec{a}},\overline{b})$.

We denote the class of functions definable in T by $\mathscr{R}(T)$ ('$\mathscr{R}$' stands for 'recursive').

Now we define our theory G_2 for FLINSPACE.

6.3.1 The theory G_2

We describe simultaneously the theory G_2 and its language $\mathscr{L}$.

Definition 6.3.3. G_2 is formulated in a language $\mathscr{L}$ containing a constant 0, a relation-symbol $\leq$ (besides =), and function-symbols S, $+$, $\times$, $\mathtt{min}$, $\mathtt{max}$, π^k_j with non-logical axioms:

Arithmetic:

A1. $\mathsf{S}(x) \neq 0$;

A2. $\mathsf{S}(x) = \mathsf{S}(y) \to x = y$;

A3. $x + 0 = x$;

A4. $x + \mathsf{S}(y) = \mathsf{S}(x+y)$;

A5. $x \times 0 = 0$;

A6. $x \times \mathsf{S}(y) = (x \times y) + x$;

A7. $x \leq y \leftrightarrow \exists z.x + z = y$;

A8. $x = 0 \vee \exists y.x = \mathsf{S}(y)$;

A9. $\mathtt{min}(x,y) = z \leftrightarrow .(x \leq y \wedge z = x) \vee (y \leq x \wedge z = y)$;

A10. $\mathtt{max}(x,y) = z \leftrightarrow .(x \leq y \wedge z = y) \vee (y \leq x \wedge z = x)$;

A11. $\pi^k_j(x_0,\ldots,x_k) = x_j$.

[3]A different definition could have been considered at this point by including the $\Delta_0(\mathscr{L})$-formulas in the $\Sigma_1(\mathscr{L})$-formulas. Both definitions work in our context.

Composition: Given function-symbols $f, g_0, \ldots, g_k$, we consider a new function-symbol $\mathsf{COMP}[f; g_0, \ldots, g_k]$ in $\mathscr{L}$ and the following axiom in G_2

$$\mathsf{COMP}[f; g_0, \ldots, g_k](\vec{x}) = f(g_0(\vec{x}), \ldots, g_k(\vec{x})).$$

Bounded recursion: For each terms t_0, t_1, and t_2 we add a function-symbol $\mathsf{BRmin}[t_0, t_1; t_2]$ to $\mathscr{L}$ and the following axioms in G_2

$$\mathsf{BRmin}[t_0, t_1; t_2](\vec{x}, 0) = \mathtt{min}(t_0(\vec{x}), t_2(\vec{x}, 0));$$
$$\mathsf{BRmin}[t_0, t_1; t_2](\vec{x}, \mathsf{S}(y)) = \mathtt{min}(t_1(\vec{x}, y, \mathsf{BRmin}[t_0, t_1; t_2](\vec{x}, y)), t_2(\vec{x}, \mathsf{S}(y))).$$

Δ_0-Induction: If B is a $\Delta_0(\mathscr{L})$-formula[4], then the following formula is also an axiom

$$B(0) \wedge \forall y.(B(y) \to B(\mathsf{S}(y))). \to \forall x.B(x).$$

It is immediate that G_2 includes $\mathsf{I}\Delta_0$ and extends its language (yet, it is not known if it is a conservative extension, since it is unknown whether $\mathscr{M}^2 = \mathscr{E}^2$). All functions in $\mathscr{E}^2$ are directly definable in G_2.

In the following, to improve readability, while defining function(-symbol)s by BRmin, whenever the existence of the bound t_2 is obvious, we omit it and we neglect the occurrence of `min` in BRmin.

Proposition 6.3.1. *For each bounded formula B, there is a function-symbol χ_B such that*[5] $\mathsf{G}_2 \vdash (B(\vec{x}) \to \chi_B(\vec{x}) = 1) \wedge (\neg B(\vec{x}) \to \chi_B(\vec{x}) = 0)$.

Proof. Clearly, by definition of G_2, for every term t, there is a function-symbol f such that $\mathsf{G}_2 \vdash f(\vec{x}) = t(\vec{x})$ (this follows from the recursive definition of the terms and the composition axiom). Let us prove the result by induction on the complexity of B. As the function $\chi_=$ is in $\mathscr{E}^2$, we can represent it by a function-symbol (let us use the same symbol to denote the function-symbol)[6]; we can define in a similar way $\chi_{\neq}$. It is also straightforward to define $\chi_{\leq}$.

Define $\chi_{\neg B}(\vec{x}) := \chi_=(\chi_B(\vec{x}), 0)$, and $\chi_{B \wedge C}(\vec{x}) := \chi_B(\vec{x}) \times \chi_C(\vec{x})$, whenever χ_B is already defined. Suppose $B(\vec{x}, y)$ is $\forall z \leq y.C(\vec{x}, z)$ and that χ_C is already defined. Consider f defined by BRmin (omitting t_2 and `min`)

$$\begin{cases} f(\vec{x}, 0) := \chi_C(\vec{x}, 0); \\ f(\vec{x}, \mathsf{S}(z)) := f(\vec{x}, z) \times \chi_C(\vec{x}, \mathsf{S}(z)). \end{cases}$$

[4] $\mathscr{L}$ denotes the constructed language for G_2 with all the added function-symbols.

[5] For simplicity, we are considering $1 := \mathsf{S}(0)$; of course in the context of numerals (that appears in the following sections), we also consider $\overline{1} := \mathsf{S}(0)$.

[6] We use the 1 for the affirmative case, i.e.

$$\chi_=(x, y) := \begin{cases} 1, & x = y \\ 0, & x \neq y \end{cases}$$

Clearly, the function f is in $\mathcal{E}^2$, and so we are allowed to use the symbol f as the function-symbol denoting the mentioned function. Define $\chi_B(\vec{x}, y) := \chi_=(f(\vec{x}, y), 1)$. If $B'(\vec{x}, y)$ is $\exists z \leq y.C'(\vec{x}, z)$ and $\chi_{C'}$ is already defined, then we set $\chi_{B'}(\vec{x}, y) := \chi_=(\chi_{\forall z \leq y.\neg C'(\vec{x},z)}(\vec{x}, y), 0)$. ⊣

Proposition 6.3.2. *Given bounded formulas $A_0(\vec{x}), \ldots, A_n(\vec{x})$ and function-symbols $f_0(\vec{x}), \ldots, f_{n+1}(\vec{x})$, there is a function-symbol $f(\vec{x})$ such that*

$$\begin{aligned} \mathsf{G}_2 \vdash & (A_0(\vec{x}) \wedge f(\vec{x}) = f_0(\vec{x})) \vee (\neg A_0(\vec{x}) \wedge A_1(\vec{x}) \wedge f(\vec{x}) = f_1(\vec{x})) \vee \cdots \vee \\ & (\neg A_0(\vec{x}) \wedge \cdots \wedge \neg A_n(\vec{x}) \wedge f(\vec{x}) = f_{n+1}(\vec{x})). \end{aligned}$$

Proof. The general result follows by induction on the construction of the result for $n = 1$; so, let us prove the result for $n = 1$. Define

$$f(\vec{x}) := (\chi_{A_0}(\vec{x}) \times f_0(\vec{x})) + (\chi_{\neg A_0 \wedge A_1}(\vec{x}) \times f_1(\vec{x})) + (\chi_{\neg A_0 \wedge \neg A_1}(\vec{x}) \times f_2(\vec{x})).$$

⊣

Proposition 6.3.3. *For each bounded formula $B(\vec{x}, y)$, there is a function-symbol f such that $\mathsf{G}_2 \vdash \exists y \leq z.B(\vec{x}, y) \to f(\vec{x}, z) \leq z \wedge B(\vec{x}, f(\vec{x}, z))$.*

Proof. Define f by BRmin (again, omitting t_2 and min)

$$\begin{cases} f(\vec{x}, 0) := \chi_{\neg B}(\vec{x}, 0); \\ f(\vec{x}, \mathsf{S}(y)) := \begin{cases} f(\vec{x}, y), & f(\vec{x}, y) < \mathsf{S}(y) \wedge B(\vec{x}, f(\vec{x}, y)) \\ \mathsf{S}(y), & \neg f(\vec{x}, y) < \mathsf{S}(y) \wedge B(\vec{x}, \mathsf{S}(y)) \\ \mathsf{S}(\mathsf{S}(y)), & \text{otherwise.} \end{cases} \end{cases}$$

⊣

Proposition 6.3.4. *If $\mathfrak{M} \models \mathsf{G}_2$ and $\mathfrak{N}$ is a substructure of $\mathfrak{M}$, then $\mathfrak{N} \models \mathsf{G}_2$.*

Proof. We prove, by induction on the complexity of B, that bounded formulas are absolute. The only hard step is to argue when $B(\vec{x}, y)$ is $\forall z \leq y.C(\vec{x}, z)$. Consider $\vec{a}$ and b in $\mathfrak{N}$. It is clear, by induction hypothesis, that if $\mathfrak{M} \models B(\vec{a}, b)$, then $\mathfrak{N} \models B(\vec{a}, b)$. Let us argue for the reciprocal. Suppose $\mathfrak{M} \not\models B(\vec{a}, b)$. Then, by the previous result, there is a function-symbol f such that $f(\vec{a}, b) \leq b$ and $\mathfrak{M} \not\models C(\vec{a}, f(\vec{a}, b))$. Using the induction hypothesis and the fact that $f(\vec{a}, b)$ is defined in $\mathfrak{N}$, we conclude that $\mathfrak{N} \not\models B(\vec{a}, b)$.

The absoluteness of the bounded formulas implies the induction axiom in $\mathfrak{N}$: for a in $\mathfrak{N}$, and any bounded formula B,

$$\mathfrak{M} \models B(0) \wedge \forall y \leq a.(B(y) \to B(\mathsf{S}(y))). \to B(a);$$

by absoluteness this holds in $\mathfrak{N}$, and from logic we obtain the induction axiom. ⊣

Theorem 6.3.1. *If $B(\vec{x}, y)$ is a bounded formula and $\mathsf{G}_2 \vdash \forall \vec{x}.\exists y.B(\vec{x}, y)$, then there is a function-symbol f such that $\mathsf{G}_2 \vdash \forall \vec{x}.B(\vec{x}, f(\vec{x}))$.*

Proof. From the previous Proposition and Loś-Tarski Theorem [49, p. 143], we know that G_2 is a universal theory. The result follows from Herbrand's Theorem [14, Theorem 1] and Proposition 6.3.2. ⊣

Theorem 6.3.2. *If $B(\vec{x}, y, \vec{z})$ is a bounded formula and $\mathsf{G}_2 \vdash \forall \vec{x}.\exists y.\exists \vec{z}.B(\vec{x}, y, \vec{z})$, then there is a function-symbol f such that $\mathsf{G}_2 \vdash \forall \vec{x}.\exists \vec{z}.B(\vec{x}, f(\vec{x}), \vec{z})$.*

Proof. This result is a straightforward consequence of the previous result, since there is a pairing function $\langle x, y\rangle$ in $\mathscr{E}^2$, such that the respective projections are are in $\mathscr{E}^2$, see [25]. ⊣

Corollary 6.3.1. $\mathscr{R}(\mathsf{G}_2) = \mathscr{E}^2$.

Proof. It is clear that each function in $\mathscr{E}^2$ is definable in G_2. The previous result ensures the converse direction. ⊣

6.3.2 A relation between $\mathscr{E}^2$ and $\mathscr{M}^2$ using G_2

The next observation presents a sufficient condition to the setting of Theorem 6.3.3 in this section.

Observation 6.3.1. Assume, in this observation, that there is a $\Delta_0(\mathsf{I}\Delta_0)$-definable function β in $\mathscr{E}^2$ such that, for each $m_0, \ldots, m_{n-1}$, there is c such that $\beta(c, n, i) = m_i$. Assume also the existence of functions seq_n in $\mathscr{E}^2$, of a concatenation function $*$, and a polynomial p satisfying:

1. $\beta(\mathsf{seq}_n(m_0, \ldots, m_{n-1}), n, i) = m_i$;
2. $\mathsf{seq}_n(m_0, \ldots, m_{n-1}) * m_n = \mathsf{seq}_{n+1}(m_0, \ldots, m_n)$;
3. $\mathsf{seq}_n(m_0, \ldots, m_{n-1}) \leq p\left(\sum_{i=0}^{n-1} m_i, n\right)$.

Given a function f in $\mathscr{E}^2$, define $\mathtt{computation}_f$ by

$$\begin{cases} \mathtt{computation}_f(\vec{x}, 0) := \mathsf{seq}_1(f(\vec{x}, 0)); \\ \mathtt{computation}_f(\vec{x}, y+1) := \mathtt{computation}_f(\vec{x}, y) * f(\vec{x}, y+1). \end{cases}$$

Observe that

$$\begin{aligned} \mathtt{computation}_f(\vec{x}, n) = \mathsf{seq}_{n+1}(f(\vec{x}, 0), \ldots, f(\vec{x}, n)) &\leq p\left(\sum_{i=0}^{n} f(\vec{x}, i), n+1\right) \\ &\leq p(\mathsf{sum}_f(\vec{x}, n), n+1). \end{aligned}$$

From Observation 6.2.1, sum_f is in $\mathscr{E}^2$ when f is, so $\mathtt{computation}_f$ is also in $\mathscr{E}^2$ when f is.

The next result show that if $\mathsf{I}\Delta_0$ has a definable β-function, then $\mathcal{M}^2 = \mathcal{E}^2$; to establish it, we use the theory G_2—this result relies on the fact that G_2 extends $\mathsf{I}\Delta_0$, so other theories for $\mathcal{E}^2$, like the one from [96], cannot be used in this context.

Theorem 6.3.3. *Suppose there is a function-symbol β, a $\Delta_0(\mathsf{I}\Delta_0)$-formula B, and, for each function-symbol f of G_2, a function-symbol $\mathtt{computation}_f$ satisfying:*

B1. $\mathsf{G}_2 \vdash \mathsf{B}(x,y,z,v) \leftrightarrow \beta(x,y,z) = v$;

B2. $\mathsf{G}_2 \vdash u \leq y \to \beta(\mathtt{computation}_f(\vec{x},y), \mathsf{S}(y), u) = f(\vec{x},u)$.

Then, $\mathcal{M}^2 = \mathcal{E}^2$.

Proof. Let us define an auxiliary theory G_2^0, that is a version of G_2, but without extending the language of $\mathsf{I}\Delta_0$. The theory G_2^0 extends the axioms of $\mathsf{I}\Delta_0$ and has the following extra axioms stating the totally of bounded formulas F_f, for each f in $\mathcal{E}^2$:

Basis: It is easy to construct bounded formulas (of $\mathsf{I}\Delta_0$) F_f representing the basis functions of $\mathcal{E}^2$, i.e. for $f \in \{0, 1, +, \times, \mathtt{min}, \mathtt{max}, \pi^k_j\}$; add accordingly the respective axioms $\forall \vec{x}.\exists! y.F_f(\vec{x},y)$.

Composition: Given $g, h_0, \ldots, h_k$ in $\mathcal{E}^2$ such that the bounded formulas F_g, $F_{h_0}, \ldots, F_{h_k}$ were already constructed and the respective axioms added; define[7]

$$F_{\mathsf{COMP}[g;h_0,\ldots,h_k]}(\vec{x},y) :=$$
$$\exists z_0 \leq p_0(\vec{x}). \cdots \exists z_k \leq p_k(\vec{x}). \left(\bigwedge_{i=0}^{k} F_{h_i}(\vec{x},z_i) \right) \wedge F_g(z_0,\ldots,z_k,y),$$

where p_i is a polynomial bounding h_i (see Theorem 6.2.1), and add the axiom $\forall \vec{x}.\exists! y.F_{\mathsf{COMP}[g;h_0,\ldots,h_k]}(\vec{x},y)$.

Bounded Recursion: If f is defined by bounded recursion from g and h with bound k, we define

$$\begin{aligned}
&F_f(\vec{x},y,z) := \exists c \leq P(\vec{x},y).(\exists r, r' \leq p_g(\vec{x}).\exists r'' \leq p_k(\vec{x},0).\mathsf{B}(c,\mathsf{S}(y),0,r') \\
&\wedge F_g(\vec{x},r) \wedge F_k(\vec{x},0,r'') \wedge r' = \mathtt{min}(r,r'')) \wedge \forall u \leq y.(u \neq y \to \\
&\exists r \leq p_k(\vec{x},\mathsf{S}(u)).\exists w \leq p_f(\vec{x},u).\exists v \leq p_h(\vec{x},u,w).\exists v' \leq p_f(\vec{x},\mathsf{S}(u)). \\
&\mathsf{B}(c,\mathsf{S}(y),u,w) \wedge \mathsf{B}(c,\mathsf{S}(y),\mathsf{S}(u),v') \wedge F_h(\vec{x},u,w,v) \wedge F_k(\vec{x},\mathsf{S}(u),r) \\
&\wedge v' = \mathtt{min}(v,r)) \wedge \mathsf{B}(c,\mathsf{S}(y),y,z),
\end{aligned}$$

where P is a polynomial bounding $\mathtt{computation}_f$ from Theorem 6.2.1, p_g is a polynomial bounding g, p_f is a polynomial bounding f, and p_k bounding k; and we add the axiom $\forall \vec{x}.\forall y.\exists! z.F_f(\vec{x},y,z)$.

[7] The use of '$\mathtt{min}$' can be substituted by the suitable formula in order to have the same language as $\mathsf{I}\Delta_0$.

It is straightforward that $\mathscr{E}^2 \subseteq \mathscr{R}(\mathsf{G}_2^0)$.

Let us argue that, for each function f in $\mathscr{E}^2$, $\mathsf{G}_2 \vdash \forall \vec{x}.\exists! y.F_f(\vec{x},y)$. It is immediate that $\mathsf{G}_2 \vdash \mathsf{I}\Delta_0 \vdash$ **Basis**. If $\mathsf{G}_2 \vdash \forall \vec{x}.\exists! z_i.F_{h_i}(\vec{x},z_i)$, for $0 \le i \le k$, and $\mathsf{G}_2 \vdash \forall \vec{x}.\exists! y.F_g(\vec{x},y)$, then, from Theorem 6.3.1, there are function-symbols $g, h_0, \dots, h_k$ (representing, respectively, the considered functions) such that

$$\mathsf{G}_2 \vdash F_g(\vec{x},y) \leftrightarrow g(\vec{x}) = y;$$
$$\mathsf{G}_2 \vdash F_{h_i}(\vec{x},y) \leftrightarrow h_i(\vec{x}) = y, \quad 0 \le i \le k.$$

From the previous facts and Theorem 6.2.1, we get $\mathsf{G}_2 \vdash \forall \vec{x}.\exists! y.F_{\mathsf{COMP}[g;h_0,\dots,h_k]}(\vec{x},y)$. So, $\mathsf{G}_2 \vdash$ **Composition**. Now, assume that $\mathsf{G}_2 \vdash \forall \vec{x}.\exists! y.F_g(\vec{x},y)$ and also $\mathsf{G}_2 \vdash \forall \vec{x}.\forall y.\exists! z.F_h(\vec{x},y,z)$, and that f is in $\mathscr{E}^2$ and can be defined using bounded recursion from g and h, with bound k. From Theorem 6.3.1, we can guarantee function-symbols g, h, and k such that

$$\mathsf{G}_2 \vdash F_g(\vec{x},y) \leftrightarrow g(\vec{x}) = y; \qquad \mathsf{G}_2 \vdash F_h(\vec{x},y,z,v) \leftrightarrow h(\vec{x},y,z) = v;$$
$$\mathsf{G}_2 \vdash F_k(\vec{x},y,z) \leftrightarrow k(\vec{x},y) = z.$$

Let f denote $\mathsf{BR}[g,h;k]$. We can guarantee the existence of a function-symbol $\mathtt{computation}_f$ satisfying B1 and B2.

Reason inside G_2. Consider $\vec{x}$ and y arbitrarily fixed and take c to be $\mathtt{computation}_f(\vec{x},y)$. Clearly, $\mathtt{computation}_f(\vec{x},y) \le P(\vec{x},y)$, because P is a polynomial bounding the function $\mathtt{computation}_f$. Moreover, it is straightforward that $g(\vec{x}) \le p_g(\vec{x})$ and $\beta(c,\mathsf{S}(y),0) = f(\vec{x},0) = \mathtt{min}(g(\vec{x}),k(\vec{x},0))$, hence, by considering $r := g(\vec{x})$, $r'' := k(\vec{x},0)$, and $r' := f(\vec{x},0)$, we have

$$\exists r,r' \le p_g(\vec{x}).\exists r'' \le p_k(\vec{x},0).\mathsf{B}(c,\mathsf{S}(y),0,r') \wedge F_g(\vec{x},r)$$
$$\wedge F_k(\vec{x},0,r'') \wedge r' = \mathtt{min}(r,r'').$$

Assume $u \le y$ such that $u \ne y$. Take $w := f(\vec{x},u)$, $v := h(\vec{x},u,f(\vec{x},u))$, $v' := f(\vec{x},\mathsf{S}(u))$, and $r := k(\vec{x},\mathsf{S}(u))$. Clearly, $w \le p_f(\vec{x},u)$, $v \le p_h(\vec{x},u,w)$, $v' \le p_f(\vec{x},\mathsf{S}(u))$, and $r \le p_k(\vec{x},\mathsf{S}(u))$. Also, $\beta(c,\mathsf{S}(y),u) = f(\vec{x},u) = w$ and $\beta(c,\mathsf{S}(y),\mathsf{S}(u)) = f(\vec{x},\mathsf{S}(u)) = v'$. Moreover,

$$v' = f(\vec{x},\mathsf{S}(u)) = \mathtt{min}(h(\vec{x},u,f(\vec{x},u)),k(\vec{x},\mathsf{S}(u))) = \mathtt{min}(v,r).$$

Thus, $\mathsf{B}(c,\mathsf{S}(y),u,w) \wedge \mathsf{B}(c,\mathsf{S}(y),\mathsf{S}(u),v') \wedge F_h(\vec{x},u,w,v) \wedge F_k(\vec{x},\mathsf{S}(u),r) \wedge v' = \mathtt{min}(v,r)$. Moreover, $\beta(c,\mathsf{S}(y),y) = f(\vec{x},y)$. So, taking z as being $f(\vec{x},y)$,

$$\forall u \le y.(u \ne y \to \exists r \le p_k(\vec{x},\mathsf{S}(u)).\exists w \le p_f(\vec{x},u).\exists v \le p_h(\vec{x},u,w).$$
$$\exists v' \le p_f(\vec{x},\mathsf{S}(u)).\mathsf{B}(c,\mathsf{S}(y),u,w) \wedge \mathsf{B}(c,\mathsf{S}(y),\mathsf{S}(u),v') \wedge F_h(\vec{x},u,w,v)$$
$$\wedge F_k(\vec{x},\mathsf{S}(u),r) \wedge v' = \mathtt{min}(v,r)) \wedge \mathsf{B}(c,\mathsf{S}(y),y,z).$$

Stepping outside G_2, we established $\mathsf{G}_2 \vdash \forall \vec{x}.\forall y.\exists z.F_f(\vec{x},y,z)$, the unicity can also be easily established; this confirms $\mathsf{G}_2 \vdash$ **Bounded Recursion**. All this entails $\mathsf{G}_2 \vdash \mathsf{G}_2^0$. In all,

$$\mathscr{E}^2 \subseteq \mathscr{R}(\mathsf{G}_2^0) \subseteq \mathscr{R}(\mathsf{G}_2) \subseteq \mathscr{E}^2;$$

consequently $\mathscr{R}(\mathsf{G}_2^0) = \mathscr{E}^2$ (the previous result can also be obtained by adapting the reasoning from [34, §4]). From Corollary 2 of [23], we know that if there is a theory T extending $\mathsf{I}\Delta_0$ such that $\mathscr{R}(T) = \mathscr{E}^2$, then $\mathscr{M}^2 = \mathscr{E}^2$. As G_2^0 is an extension of $\mathsf{I}\Delta_0$ with the same language as $\mathsf{I}\Delta_0$ satisfying $\mathscr{R}(\mathsf{G}_2^0) = \mathscr{E}^2$, it follows that $\mathscr{M}^2 = \mathscr{E}^2$. ⊣

The conditions of Observation 6.3.1 are sufficient to guarantee the antecedent of the previous result, namely B1 and B2.

Corollary 6.3.2. *Suppose the conditions of the previous Theorem. Then,* $\mathsf{FLTH} = \mathsf{FLINSPACE}$.

Proof. Follows from the previous result and the two following facts: $\mathscr{M}^2 = \mathsf{FLTH}$ and $\mathscr{E}^2 = \mathsf{FLINSPACE}$. ⊣

6.4 Bootstrapping G_2

Throughout this section we use the fact that $\mathscr{E}^2$ coincides with $\mathsf{FLINSPACE}$, and the fact that each function in $\mathscr{E}^2$ can be represented by a function-symbol in G_2, and vice-versa. We use the term 'bootstrapping'[8] in reference to [10, p. 37]. We confirm that the bootstrapping procedure also works for G_2: we assume that the reader is aware of [39] and [10]. We assume a fixed sequence-function and its respective β-function, several options could have been made here, but we are going to consider a fix one.

Definition 6.4.1. We consider the sequence function that assigns, assuming binary notation, '1'↦ '11', '0'↦'00', and ','↦'01'; so, for example[9]

$$\langle 11_2, 1_2, 101_2\rangle = 1111011101110011_2.$$

It is easy to see that this function, its respective concatenation $*$, and its respective β-function are in $\mathsf{FLINSPACE}$ (by a β-function, in this context, we simply mean projection-functions $(\cdot)_i$ such that $(\langle x_0, \ldots, x_k\rangle)_i = x_i$), and consequently can be represented in G_2 by function-symbols.[10] We can define terms and formulas in the usual way using this sequence-function (see [39, pp. 312, 313]), having $\Delta_0(\mathsf{G}_2)$-formulas $\texttt{Term}(x)$ and $\texttt{Formula}(x)$ respectively identifying them (see also [10, pp. 126–135] for more details).[11] We assume that a formula is being codified in the order it is written, for example

$$\ulcorner \exists x.(\mathsf{S}(x) + y = z)\urcorner = \langle \ulcorner\exists\urcorner, \ulcorner x\urcorner, \ulcorner .\urcorner, \ulcorner(\urcorner, \ulcorner \mathsf{S}\urcorner, \ulcorner(\urcorner, \ulcorner x\urcorner, \ulcorner)\urcorner, \ulcorner+\urcorner, \\ \ulcorner y\urcorner, \ulcorner=\urcorner, \ulcorner z\urcorner, \ulcorner)\urcorner\rangle.$$

[8] This term is used in computer science to describe the operations needed to start a computer, commonly a small amount of software.

[9] As usual, $\cdot_2$ is used to represent binary notation; for example, $101_2 = 5$.

[10] This does not contradict Theorem 6.3.3, there the β-function needed to be definable in $\mathsf{I}\Delta_0$; these projections are definable in G_2.

[11] This work because G_2 includes $\mathsf{I}\Delta_0$, the theory considered in [39, pp. 312, 313]; all the definitions from [39] work in our context, we just need to adapt them for the sequence function we decided to consider to simplify our proofs.

Definition 6.4.2. As usual, we can define a function numeral inside G_2 (see [10, p. 29] for further details) such that:

$$\mathtt{numeral}(0) := \langle \ulcorner 0 \urcorner \rangle;$$
$$\mathtt{numeral}(\overline{2} \times x + \overline{1}) := \mathtt{numeral}(\overline{2} \times x) * \langle \ulcorner + \urcorner, \ulcorner (\urcorner, \ulcorner \mathsf{S} \urcorner, \ulcorner (\urcorner, \ulcorner 0 \urcorner, \ulcorner) \urcorner, \ulcorner) \urcorner \rangle;$$
$$\mathtt{numeral}(\overline{2} \times (x + \overline{1})) := \langle \ulcorner (\urcorner, \ulcorner \mathsf{S} \urcorner, \ulcorner (\urcorner, \ulcorner \mathsf{S} \urcorner, \ulcorner (\urcorner, \ulcorner 0 \urcorner, \ulcorner) \urcorner, \ulcorner) \urcorner, \ulcorner) \urcorner, \ulcorner \times \urcorner, \ulcorner (\urcorner \rangle$$
$$* \mathtt{numeral}(x + \overline{1}) * \langle \ulcorner) \urcorner \rangle.$$

The fact that numeral is in G_2 follows from the fact that the efficient numerals[12] have size proportional to the length of the binary representation of the number they denote [10, p. 29]: the idea is that the suitable Turing Machine that computes numeral will use an amount of space proportional to the size of the input, it is therefore easy to argue that numeral is in FLINSPACE, i.e. definable in G_2.

The next result serves to illustrate that the standard results of metamathematics are provable in this context.

Proposition 6.4.1. $\mathsf{G}_2 \vdash \forall x.\mathsf{Term}(\mathtt{numeral}(x))$.

Proof. Let us see that when $\varphi(x)$ is a $\Delta_0(\mathsf{G}_2)$-formula, if

$$\mathsf{G}_2 \vdash \varphi(0) \wedge \forall x.(\varphi(\overline{2} \times x) \rightarrow \varphi(\overline{2} \times x + \overline{1})) \wedge$$
$$\forall x.(\varphi(x + \overline{1}) \rightarrow \varphi(\overline{2} \times (x + \overline{1}))),$$

then $\mathsf{G}_2 \vdash \forall x.\varphi(x)$. Assume $\mathsf{G}_2 \vdash \varphi(0) \wedge \forall x.(\varphi(\overline{2} \times x) \rightarrow \varphi(\overline{2} \times x + \overline{1})) \wedge \forall x.(\varphi(x + \overline{1}) \rightarrow \varphi(\overline{2} \times (x + \overline{1})))$ and consider $\Phi(x) := \forall y \leq x.(\varphi(y) \wedge \varphi(y) \rightarrow \varphi(y + \overline{1}))$ (this is a $\Delta_0(\mathsf{G}_2)$-formula). It is not hard to see that $\mathsf{G}_2 \vdash \forall x.\exists y.x = \overline{2} \times y \vee x = \overline{2} \times y + \overline{1}$.

Reason in G_2. Clearly, $\Phi(0)$. Suppose, by induction hypothesis, that $\Phi(x)$. Then, in particular, $\varphi(x) \wedge \varphi(x) \rightarrow \varphi(x + \overline{1})$, and so $\varphi(x + \overline{1})$. Let us confirm that $\varphi(x + \overline{1}) \rightarrow \varphi((x + \overline{1}) + \overline{1})$. Suppose $\varphi(x + \overline{1})$. We know that $\exists y.x + \overline{1} = \overline{2} \times y \vee x + \overline{1} = \overline{2} \times y + \overline{1}$. If $x + \overline{1} = \overline{2} \times y$, as $\varphi(\overline{2} \times y) \rightarrow \varphi(\overline{2} \times y + \overline{1})$ and $\varphi(x + \overline{1})$, it follows $\varphi((x + \overline{1}) + \overline{1})$. Suppose $x + \overline{1} = \overline{2} \times y + \overline{1}$. If $x = 0$, then we are done; so assume also $x > 0$. Then, $y + \overline{1} \leq x$, and so, by induction hypothesis, $\varphi(y + \overline{1})$. By assumption, $\varphi(y + \overline{1}) \rightarrow \varphi(\overline{2} \times (y + \overline{1}))$, and so $\varphi(\overline{2} \times (y + \overline{1}))$, i.e. $\varphi((x + \overline{1}) + \overline{1})$. In sum, $\Phi(x + \overline{1})$. All this means that $\Phi(x) \rightarrow \Phi(x + \overline{1})$. So, $\forall x.\Phi(x)$, and consequently $\forall x.\varphi(x)$. Going outside G_2, we established $\mathsf{G}_2 \vdash \forall x.\varphi(x)$.

It is not hard to see that

$$\mathsf{G}_2 \vdash \mathsf{Term}(\mathtt{numeral}(0));$$
$$\mathsf{G}_2 \vdash \forall x.\mathsf{Term}(\mathtt{numeral}(\overline{2} \times x)) \rightarrow \mathsf{Term}(\mathtt{numeral}(\overline{2} \times x + \overline{1}));$$
$$\mathsf{G}_2 \vdash \forall x.\mathsf{Term}(\mathtt{numeral}(x + \overline{1})) \rightarrow \mathsf{Term}(\mathtt{numeral}(\overline{2} \times (x + \overline{1}))).$$

The result follows from what we have just observed. ⊣

[12]The efficient numerals are defined by (see [10, p. 29]):

$$I_0 := 0; \qquad I_{2k+1} := (I_{2k}) + (\mathsf{S}(0)); \qquad I_{2(k+1)} := (\mathsf{S}(\mathsf{S}(0))) \times (I_{k+1}).$$

We use the notation $\overline{n} := I_n$.

As we mentioned, the usual metamathematical results from [10] and [39] concerning the internalized terms and formulas can be obtained in this setting, just like the previous result.

Definition 6.4.3. We define Feferman's dot notation using a function-symbol sub of G_2 such that, if $\varphi(x)$ is a formula with x being its only free-variable and t a term, $\mathsf{G}_2 \vdash \mathsf{sub}(\ulcorner\varphi\urcorner,\ulcorner t\urcorner) = \ulcorner\varphi(t)\urcorner$, and by defining

$$\ulcorner\varphi(\overset{\bullet}{x})\urcorner := \mathsf{sub}(\ulcorner\varphi\urcorner, \mathtt{numeral}(x)),$$

we will also write $\ulcorner\varphi(\overbrace{\quad x \quad}^{\bullet})\urcorner$ for more complex values of x (we can define in a similar way the previous notations for several variables). We will assume function-symbols $\overset{\bullet}{=}$, $\overset{\bullet}{\leq}$, $\overset{\bullet}{\to}$, and $\overbrace{\forall x.}^{\bullet}$ of G_2 satisfying:

$$\mathsf{G}_2 \vdash \overset{\bullet}{=} (\overline{n},\overline{m}) = \ulcorner \overline{n} = \overline{m}\urcorner; \qquad \mathsf{G}_2 \vdash \overset{\bullet}{\leq} (\overline{n},\overline{m}) = \ulcorner \overline{n} \leq \overline{m}\urcorner;$$

$$\mathsf{G}_2 \vdash \overset{\bullet}{\to} (\ulcorner\varphi\urcorner,\ulcorner\psi\urcorner) = \ulcorner\varphi \to \psi\urcorner; \qquad \mathsf{G}_2 \vdash \overbrace{\forall \vec{x}.}^{\bullet}\ (\ulcorner\varphi(\vec{x})\urcorner) = \ulcorner\forall\vec{x}.\varphi(\vec{x})\urcorner.$$

As usual, we write $x \overset{\bullet}{=} y$ for $\overset{\bullet}{=} (x,y)$, $x \overset{\bullet}{\leq} y$ for $\overset{\bullet}{\leq} (x,y)$, $x \overset{\bullet}{\to} y$ for $\overset{\bullet}{\to} (x,y)$, and $\overbrace{\forall \vec{x}.}^{\bullet}\ y$ for $\overbrace{\forall \vec{x}.}^{\bullet}\ (y)$.

For example, if $\varphi(x) := \exists y.(x + y = \mathsf{S}(0))$, then

$$\begin{aligned}\mathsf{G}_2 \vdash \ulcorner\varphi(\overset{\bullet}{x})\urcorner =& \langle\ulcorner\exists\urcorner,\ulcorner y\urcorner,\ulcorner.\urcorner,\ulcorner(\urcorner\rangle * \mathtt{numeral}(x)\\ &* \langle\ulcorner+\urcorner,\ulcorner y\urcorner,\ulcorner=\urcorner,\ulcorner\mathsf{S}\urcorner,\ulcorner(\urcorner,\ulcorner 0\urcorner,\ulcorner)\urcorner,\ulcorner)\urcorner\rangle.\end{aligned}$$

Definition 6.4.4. We say that ξ is a *numeration* of the axioms of a theory T if the set $\{n \in \mathbb{N} | \mathsf{G}_2 \vdash \xi(\overline{n})\}$ coincides with the set of the codes of the axioms of T. Throughout the rest of this chapter, we assume that ξ denotes a generic $\Delta_0(\mathsf{G}_2)$-numeration of the axioms of T that includes G_2. For a numeration ξ of the axioms of T, we will denote by $\mathtt{Prf}_\xi$ the *standard proof predicate* for ξ as constructed by Feferman in [33]:

$$\begin{aligned}\mathtt{Prf}_\xi(x,y) :=& \mathsf{Sq}(y) \wedge \neg\mathsf{L}(y) = 0 \wedge (\forall u < \mathsf{L}(y).\mathtt{Formula}((y)_u) \wedge\\ &(\xi((y)_u) \vee \exists v < u.\exists w < u.(y)_v = (y)_w \overset{\bullet}{\to} (y)_u))\\ &\wedge x = (y)_{\mathsf{L}(y) \dot{-} 1},\end{aligned}$$

where L is defined in such a way that $\mathsf{G}_2 \vdash \mathsf{L}(\langle x_0, \ldots, x_k\rangle) = \overline{k+1}$ and Sq $\Delta_0(\mathsf{G}_2)$-identifies all sequences inside G_2. Feferman emphasized the logical axioms using another formula, we do not follow that approach here; we assume that even the logical axioms are codified in ξ. We also denote by $\mathtt{Pr}_\xi(x) := \exists y.\mathtt{Prf}_\xi(x,y)$ the *standard provability predicate* for ξ. It is clear that if ξ is $\Delta_0(\mathsf{G}_2)$, so is $\mathtt{Prf}_\xi$.

We have the following derivability conditions (see, for instance, [33] for further details):

C1: $T \vdash \varphi \implies \mathsf{G}_2 \vdash \mathtt{Pr}_\xi(\ulcorner \varphi \urcorner)$, this condition is immediate from the Σ_1-completeness of G_2;

C2: $\mathsf{G}_2 \vdash \xi \to \mathtt{Pr}_\xi$, this condition[13] follows from the definition of $\mathtt{Pr}_\xi$ and from the fact that the function $n \mapsto \langle n \rangle$ is in $\mathscr{E}^2$;

C3: $\mathsf{G}_2 \vdash \mathtt{Pr}_\xi(x \mathbin{\dot{\to}} y) \to (\mathtt{Pr}_\xi(x) \to \mathtt{Pr}_\xi(y))$; this condition holds because, provably in G_2, if π_0 is a proof of $x \mathbin{\dot{\to}} y$ according to ξ and π_1 is a proof of x according to ξ, then $\pi_0 * \pi_1 * \langle y \rangle$ is a proof of y according to ξ;

C4: $\mathsf{G}_2 \vdash \mathtt{Pr}_\xi(\ulcorner \forall \vec{x}.\varphi(\vec{x}) \urcorner) \to \forall \vec{x}.\mathtt{Pr}_\xi(\ulcorner \varphi(\dot{\vec{x}}) \urcorner)$; this condition holds whenever we have $\mathsf{G}_2 \vdash$ $\forall \mathtt{Formula}(y).\forall \mathtt{Term}(z).\xi(\dot{\overbrace{\forall x.}}\ y \mathbin{\dot{\to}} \mathtt{sub}(y,z))$ (as usual, the formula $\forall \mathtt{Formula}(x).\Phi(x,y)$ abbreviates $\forall x.\mathtt{Formula}(x) \to \Phi(x,y)$).

Condition **C3** holds for generic formulas ξ, that are not necessarily numerations of any specific set of axioms[14]; for this reason, we will assume that G_2 includes **C3** as its axiom schema for generic formulas ξ (this is constitutes a conservative extension of the original set of axioms of G_2). This means that, for any one-free-variable formula φ, we will assume for Theorem 6.5.2

$$\mathsf{G}_2 \vdash \forall \mathtt{Formula}(x).\forall \mathtt{Formula}(y).\xi(\ulcorner \dot{\overbrace{\mathtt{Pr}_\varphi}}(x \mathbin{\dot{\to}} y) \to (\mathtt{Pr}_\varphi(\dot{x}) \to \mathtt{Pr}_\varphi(\dot{y}))\urcorner). \quad \text{(Assumption 1)}$$

This assumption is not strictly necessary to require, but it will allow some proofs to be less complex; see Footnote 22. In practice, Assumption 1 might constitute a very subtle expansion of G_2.

6.5 'Provability implies provable provability' in G_2

We start this section with a useful result.

Proposition 6.5.1. *Assume that ξ is such that $\mathsf{G}_2 \vdash \forall x.\mathtt{Term}(x) \to \xi(x \mathbin{\dot{=}} x)$. Then, $\mathsf{G}_2 \vdash \forall x.\forall y.x = y \to \mathtt{Pr}_\xi(\ulcorner \dot{x} = \dot{y} \urcorner)$.*

Proof. Reason in G_2 and suppose $x = y$. From Proposition 6.4.1, $\mathtt{Term}(\mathtt{numeral}(x))$, so $\xi(\mathtt{numeral}(x) \mathbin{\dot{=}} \mathtt{numeral}(x))$. As $x = y$, then we have $\mathtt{numeral}(x) = \mathtt{numeral}(y)$, and consequently $\xi(\mathtt{numeral}(x) \mathbin{\dot{=}} \mathtt{numeral}(y))$; so, as $\mathsf{G}_2 \vdash \ulcorner \dot{x} = \dot{y} \urcorner = (\mathtt{numeral}(x) \mathbin{\dot{=}} \mathtt{numeral}(y))$, then $\xi(\ulcorner \dot{x} = \dot{y} \urcorner)$. The result follows by **C2**. ⊣

[13]By $\xi \to \mathtt{Pr}_\xi$ we obviously mean $\forall x.\xi(x) \to \mathtt{Pr}_\xi(x)$; we will use this kind of abbreviation throughout this chapter.

[14]The justification we gave for **C3** suffcies to conclude this fact for generic formulas when one has in mind that $\mathsf{G}_2 \vdash \mathtt{Formula}(x) \wedge \mathtt{Formula}(y) \leftrightarrow \mathtt{Formula}(x \mathbin{\dot{\to}} y)$.

Throughout the rest of this section, we are going to assume that ξ is a $\Delta_0(\mathsf{G}_2)$-numeration of a theory T satisfying the conditions of the previous result. The next result is a sufficient condition for G_2 to prove the internal Δ_0-completeness for ξ, and hence the internal Σ_1-completeness (Corollary 6.5.1).

Theorem 6.5.1. *Suppose* G_2 *satisfies* **C4** *for* ξ. *If, for every function-symbol* f, *we have* $\mathsf{G}_2 \vdash \mathsf{Pr}_\xi(\ulcorner \overset{\bullet}{\overbrace{f(\vec{x})}} = \overline{1}\urcorner) \to \mathsf{Pr}_\xi(\ulcorner f(\dot{\vec{x}}) = \overline{1}\urcorner)$, *then, for every* Δ_0*-formula* $A(\vec{x})$ *of* G_2, $\mathsf{G}_2 \vdash A(\vec{x}) \to \mathsf{Pr}_\xi(\ulcorner A(\dot{\vec{x}})\urcorner)$.

Proof. Let $A(\vec{x})$ be a Δ_0-formula. From Proposition 6.3.1 we know that

$$\mathsf{G}_2 \vdash \forall \vec{x}.(A(\vec{x}) \leftrightarrow \chi_A(\vec{x}) = \overline{1}).$$

So, from **C1**, $\mathsf{G}_2 \vdash \mathsf{Pr}_\xi(\ulcorner \forall \vec{x}.(A(\vec{x}) \leftrightarrow \chi_A(\vec{x}) = \overline{1})\urcorner)$. From **C4**, it follows $\mathsf{G}_2 \vdash \forall \vec{x}.\mathsf{Pr}_\xi(\ulcorner A(\dot{\vec{x}}) \leftrightarrow \chi_A(\dot{\vec{x}}) = \overline{1}\urcorner)$.

Now reason in G_2. Assume $A(\vec{x})$. Then, $\chi_A(\vec{x}) = \overline{1}$, and so, from Proposition 6.5.1, $\mathsf{Pr}_\xi(\ulcorner \overset{\bullet}{\overbrace{\chi_A(\vec{x})}} = \overline{1}\urcorner)$. By assumption, $\mathsf{Pr}_\xi(\ulcorner \chi_A(\dot{\vec{x}}) = \overline{1}\urcorner)$, and thus, using **C3**, $\mathsf{Pr}_\xi(\ulcorner A(\dot{\vec{x}})\urcorner)$. ⊣

Corollary 6.5.1. *Suppose* G_2 *satisfies* **C4** *for* ξ *and that, for every function-symbol* f, $\mathsf{G}_2 \vdash \mathsf{Pr}_\xi(\ulcorner \overset{\bullet}{\overbrace{f(\vec{x})}} = \overline{1}\urcorner) \to \mathsf{Pr}_\xi(\ulcorner f(\dot{\vec{x}}) = \overline{1}\urcorner)$. *If* $A(\vec{x})$ *is a* $\Sigma_1(\mathsf{G}_2)$*-formula, then* $\mathsf{G}_2 \vdash A(\vec{x}) \to \mathsf{Pr}_\xi(\ulcorner A(\dot{\vec{x}})\urcorner)$.

Proof. Take $B(\vec{x},\vec{y})$ a bounded formula such that $\mathsf{G}_2 \vdash A(\vec{x}) \leftrightarrow \exists \vec{y}.B(\vec{x},\vec{y})$. From the previous result, $\mathsf{G}_2 \vdash B(\vec{x},\vec{y}) \to \mathsf{Pr}_\xi(\ulcorner B(\dot{\vec{x}},\dot{\vec{y}})\urcorner)$, so $\mathsf{G}_2 \vdash \exists \vec{y}.B(\vec{x},\vec{y}) \to \exists \vec{y}.\mathsf{Pr}_\xi(\ulcorner B(\dot{\vec{x}},\dot{\vec{y}})\urcorner)$. We know that $\mathsf{G}_2 \vdash B(\vec{x},\vec{y}) \to \exists \vec{y}.B(\vec{x},\vec{y})$, so $\mathsf{G}_2 \vdash \mathsf{Pr}_\xi(\ulcorner B(\dot{\vec{x}},\dot{\vec{y}})\urcorner) \to \mathsf{Pr}_\xi(\ulcorner \exists \vec{y}.B(\dot{\vec{x}},\vec{y})\urcorner)$, and so the result follows. ⊣

Corollary 6.5.2. *Suppose* G_2 *satisfies* **C4** *for* ξ *and that, for every function-symbol* f, $\mathsf{G}_2 \vdash \mathsf{Pr}_\xi(\ulcorner \overset{\bullet}{\overbrace{f(\vec{x})}} = \overline{1}\urcorner) \to \mathsf{Pr}_\xi(\ulcorner f(\dot{\vec{x}}) = \overline{1}\urcorner)$. *Then,*

$$\mathsf{G}_2 \vdash \mathsf{Pr}_\xi(x) \to \mathsf{Pr}_\xi(\ulcorner \mathsf{Pr}_\xi(\dot{x})\urcorner).$$

Proof. Follows from the fact that, as ξ is $\Delta_0(\mathsf{G}_2)$, $\mathsf{Pr}_\xi(x)$ is a $\Sigma_1(\mathsf{G}_2)$-formula. ⊣

Now we present one of our main results.

Theorem 6.5.2. (*Under Assumption 1*) *If there is a function-symbol* `verifier` *in* G_2 *such that it satisfies* $\mathsf{G}_2 \vdash \xi(x) \to \mathsf{Prf}_\xi(\ulcorner \mathsf{Pr}_\xi(\dot{x})\urcorner, \mathtt{verifier}(x))$, *then*

$$\mathsf{G}_2 \vdash \mathsf{Pr}_\xi(x) \to \mathsf{Pr}_\xi(\ulcorner \mathsf{Pr}_\xi(\dot{x})\urcorner).$$

Proof. Consider a function $\mathtt{transform}_0$ that, on input $y = \langle y_0, \ldots, y_k \rangle$ (the inputs that are not sequences do not matter), goes entry-by-entry of the sequence y testing if $\xi(y_i)$ and, in the positive case, and just for that case, substitutes the content[15] of the entry by the content of $\mathtt{verifier}(y_i)$. Let us confirm that $\mathtt{transform}_0(y)$ can be computed using space linearly bounded by the size of the whole sequence y (and thus $\mathtt{transform}_0$ is in $\mathsf{FLINSPACE}$). Consider $y = \langle y_0, \ldots, y_k \rangle$ and $\mathtt{double}$ the $\mathsf{FLINSPACE}$-function that, on input x, duplicates the occurrences of each digit of the binary representation of x (for example, $\mathtt{double}(101_2) = 110011_2$). Then, in terms of strings representing binary notation,

$$y = \mathtt{double}(y_0)01 \cdots 01\mathtt{double}(y_k).$$

For the position i, the Turing machine that computes $\mathtt{transform}_0$ starts by reducing $\mathtt{double}(y_i)$, in a new tape, to just y_i (this needs no extra space, of course besides writing down y_i). After that, it computes—using space bounded by $k_0 \cdot |y_i|$, with k_0 a fixed constant—$\chi_\xi(y_i)$; this function is in $\mathsf{FLINSPACE}$ because it is definable by the following Δ_0-formula in G_2: $(\xi(x) \wedge z = 1) \vee (\neg\xi(x) \wedge z = 0)$, see Proposition 6.3.1. In the positive case, the machine substitutes in the original y the position $\mathtt{double}(y_i)$ for $\mathtt{verifier}(y_i)$, using space bounded by $k_1 \cdot |y_i|$, with k_1 a constant bounding the space used to compute the function $\mathtt{verifier}$. The whole process yields a computation space bounded by

$$\sum_{i=0}^{k} (k_0 + k_1) \cdot |y_i| \leq \sum_{i=0}^{k} (k_0 + k_1) \cdot |\mathtt{double}(y_i)| \leq (k_0 + k_1) \cdot |y|.$$

This confirms that $\mathtt{transform}_0$ is in $\mathsf{FLINSPACE}$, as desired.

Let us describe the Turing Machine $\mathtt{TMsimulate}$. Consider the input y. If y is not a sequence, then $\mathtt{TMsimulate}$ outputs 0. If $y = \langle y_0, \ldots, y_k \rangle$, then $\mathtt{TMsimulate}$ applies to y the function $\mathtt{transform}_0$ and outputs the result to the second tape. Then, starting from the first entry of the first tape, $\mathtt{TMsimulate}$ goes entry-by-entry testing whether there are $j, m < i$ such that $y_m = y_j \mathbin{\dot{\to}} y_i$ (*videlicet* to test if *modus ponens* was used in that position of y); in the positive case, $\mathtt{TMsimulate}$ substitutes, in the second tape, the content of the position in the second tape corresponding to position m in the first tape by the content of $\langle \ulcorner \mathtt{Pr}_\xi(\dot{y}_m) \urcorner, \ulcorner \mathtt{Pr}_\xi(\dot{y}_m) \to (\mathtt{Pr}_\xi(\dot{y}_j) \to \mathtt{Pr}_\xi(\dot{y}_i)) \urcorner, \ulcorner \mathtt{Pr}_\xi(\dot{y}_j) \to \mathtt{Pr}_\xi(\dot{y}_i) \urcorner \rangle$, and the content of the position corresponding to position i of the first tape by the content of $\langle \ulcorner \mathtt{Pr}_\xi(\dot{y}_i) \urcorner \rangle$; in the negative case, the machine just proceeds to the next entry. By corresponding position in the second tape we mean the position that, before the described changes were applied, was the considered initial one; this means that, in the previously described situation where $y_m = y_j \mathbin{\dot{\to}} y_i$, if we are in the case $\xi(y_m)$, then $\mathtt{verifier}$ was already applied in the second tape (it terminates in $\ulcorner \mathtt{Pr}_\xi(\dot{y}_m) \urcorner$) and now one adds to it the content of $\langle \ulcorner \mathtt{Pr}_\xi(\dot{y}_m) \to (\mathtt{Pr}_\xi(\dot{y}_j) \to \mathtt{Pr}_\xi(\dot{y}_i)) \urcorner, \ulcorner \mathtt{Pr}_\xi(\dot{y}_j) \to \mathtt{Pr}_\xi(\dot{y}_i) \urcorner \rangle$.

Let $\mathtt{simulate}$ be the function computed by $\mathtt{TMsimulate}$. As we saw before, the application of $\mathtt{verifier}$ to y has computational space linearly bounded by $|y|$; moreover, using

[15] By the *content* of a given sequence $\langle z_0, \ldots, z_n \rangle$, we mean the array of elements $z_0, \ldots, z_n$ not under the sequence function. For example, substituting y_0 for the content of $\langle z_0, z_1, z_2 \rangle$ in $\langle y_0, y_1 \rangle$ yields $\langle z_0, z_1, z_2, y_1 \rangle$.

a similar argumentation, the second part of the action of `simulate` also does, since the computation space needed to compute $\langle \ulcorner \mathtt{Pr}_\xi(\dot{y}_m)\urcorner, \ulcorner \mathtt{Pr}_\xi(\dot{y}_m) \to (\mathtt{Pr}_\xi(\dot{y}_j) \to \mathtt{Pr}_\xi(\dot{y}_i))\urcorner, \ulcorner \mathtt{Pr}_\xi(\dot{y}_j) \to \mathtt{Pr}_\xi(\dot{y}_i)\urcorner\rangle$ is linearly bounded on the size of y_i, y_j, y_m; thus, that computation space is bounded by $k \cdot (|y_i|+|y_j|+|y_m|)$, and so the total computational space of the second part of the action of `simulate` is bounded by $K \cdot |y|$, for a suitable constant K.[16] All this means that `simulate` is in $\mathscr{E}^2$.[17]

Let `upto` be a function in $\mathscr{E}^2$ such that

$$G_2 \vdash \mathtt{upto}(\langle y_0, \ldots, y_k\rangle, j) = \langle y_0, \ldots, y_j\rangle,$$

let `fromto` be a function in $\mathscr{E}^2$ such that

$$G_2 \vdash \mathtt{fromto}(\langle y_0, \ldots, y_k\rangle, j, m) = \langle y_j, \ldots, y_m\rangle,$$

and `position` a function that, on inputs y and i, gives, whenever y is a sequence, the position in $\mathtt{simulate}(y)$ that corresponds to the same position in i (now under the scope of $\ulcorner \mathtt{Pr}_\xi(\cdot)\urcorner$). The function `position` is also in $\mathscr{E}^2$; to see this, we just need to create a suitable TM that follows along the computation of `TMsimulate` and keeps track of the position corresponding to the initial one.[18]

Let us reason inside G_2. Assume $\mathtt{Prf}_\xi(x, y)$, with $y = \langle y_0, \ldots, y_{\mathsf{L}(y) \dot{-} 1}\rangle$. Now construct $\mathtt{simulate}(y)$. It is not hard to establish that

$$G_2 \vdash (\mathtt{simulate}(y))_{\mathtt{position}(y,i)} = \ulcorner \mathtt{Pr}_\xi(\dot{y}_i)\urcorner.$$

Let us see that $\mathtt{Prf}_\xi(\ulcorner \mathtt{Pr}_\xi(\dot{x})\urcorner, \mathtt{simulate}(y))$. It is easy to see that the last entry of $\mathtt{simulate}(y)$

[16]The idea is that to create $\langle \ulcorner \mathtt{Pr}_\xi(\cdot)\urcorner, \ulcorner \mathtt{Pr}_\xi(\cdot) \to (\mathtt{Pr}_\xi(\cdot) \to \mathtt{Pr}_\xi(\cdot))\urcorner, \ulcorner \mathtt{Pr}_\xi(\cdot) \to \mathtt{Pr}_\xi(\cdot)\urcorner\rangle$, with the empty slot indicated by $\cdot$, we just need a fixed amount of space; then we substitute the empty slot $\cdot$ by $\mathtt{double}(\mathtt{numeral}(y_m))$, $\mathtt{double}(\mathtt{numeral}(y_i))$, and $\mathtt{double}(\mathtt{numeral}(y_j))$ accordingly; the `numeral` gives the Gödel-number of the numeral of y_m, and the `double` function accounts for the sequence formation. Clearly, the space needed to perform these actions is proportional to $|y_m|+|y_i|+|y_j|$, *id est* proportional to y_k.

[17]Let us give a toy example here. If we are given $y = \langle y_0, y_0 \dot{\to} y_1, y_1\rangle$, with $\xi(y_0)$ and $\xi(y_0 \dot{\to} y_1)$, then firstly we use the function $\mathtt{transform}_0$ to write in the second tape $\langle \underbrace{\cdots, \ulcorner \mathtt{Pr}_\xi(\dot{y}_0)\urcorner}_{\text{proof of } \mathtt{Pr}_\xi(y_0)}, \underbrace{\cdots, \ulcorner \mathtt{Pr}_\xi(\overbrace{y_0 \dot{\to} y_1}^{\bullet})\urcorner}_{\text{proof of } \mathtt{Pr}_\xi(y_0 \dot{\to} y_1)}, y_1\rangle$. After that, we use the second part of the proceedure of `simulate` to obtain $\langle \cdots, \ulcorner \mathtt{Pr}_\xi(\dot{y}_0)\urcorner, \cdots, \ulcorner \mathtt{Pr}_\xi(\overbrace{y_0 \dot{\to} y_1}^{\bullet})\urcorner, \ulcorner \mathtt{Pr}_\xi(\overbrace{y_0 \dot{\to} y_1}^{\bullet}) \to (\mathtt{Pr}_\xi(\dot{y}_0) \to \mathtt{Pr}_\xi(\dot{y}_1))\urcorner, \ulcorner \mathtt{Pr}_\xi(\dot{y}_0) \to \mathtt{Pr}_\xi(\dot{y}_1)\urcorner, \ulcorner \mathtt{Pr}_\xi(\dot{y}_1)\urcorner\rangle$. So, from a proof of y_1 we obtained a proof of $\mathtt{Pr}_\xi(y_1)$.

[18]In the example of Footnote 17, the position corresponding to y_0 is the one with $\ulcorner \mathtt{Pr}_\xi(\dot{y}_0)\urcorner$, the position corresponding to $y_0 \dot{\to} y_1$ is the one with $\ulcorner \mathtt{Pr}_\xi(\overbrace{y_0 \dot{\to} y_1}^{\bullet})\urcorner$, etc.

is, by construction, $\ulcorner \mathsf{Pr}_\xi(\dot{x})\urcorner$. Let us argue by induction[19] on $i < \mathsf{L}(y)$ that

$$\forall i < \mathsf{L}(y).\mathsf{Prf}_\xi((\mathsf{simulate}(y))_{\mathsf{position}(y,i)}, \mathsf{upto}(\mathsf{simulate}(y), \mathsf{position}(y,i));$$

this is possible since the needed formula is $\Delta_0(\mathsf{G}_2)$. Clearly, y_0 needs to be an axiom according to ξ, and so the corresponding position in $\mathsf{simulate}(y)$ is $\mathsf{verifier}(y_0)$, a proof of $\ulcorner \mathsf{Pr}_\xi(\dot{y_0})\urcorner$. Suppose that the result holds for all $j \le i < \mathsf{L}(y)$ and that $\mathsf{S}(i) < \mathsf{L}(y)$. By induction hypothesis,

$$\mathsf{Prf}_\xi((\mathsf{simulate}(y))_{\mathsf{position}(y,i)}, \mathsf{upto}(\mathsf{simulate}(y), \mathsf{position}(y,i)).$$

Let us see that

$$\mathsf{Prf}_\xi((\mathsf{simulate}(y))_{\mathsf{position}(y,\mathsf{S}(i))}, \mathsf{upto}(\mathsf{simulate}(y), \mathsf{position}(y,\mathsf{S}(i))).$$

By construction, the last element of $\mathsf{upto}(\mathsf{simulate}(y), \mathsf{position}(y,\mathsf{S}(i)))$ is exactly $(\mathsf{simulate}(y))_{\mathsf{position}(y,\mathsf{S}(i))}$, so it remains to confirm that, in fact, $\mathsf{upto}(\mathsf{simulate}(y),$ $\mathsf{position}(y,\mathsf{S}(i)))$ is a proof.

As $\mathsf{S}(i) < \mathsf{L}(y)$, then $\xi(y_{\mathsf{S}(i)})$ or there are $j, m < \mathsf{S}(i)$ such that $y_m = y_j \dot{\to} y_{\mathsf{S}(i)}$. If $\xi(y_{\mathsf{S}(i)})$ is the case, then

$$\mathsf{fromto}(\mathsf{simulate}(y), \mathsf{S}(\mathsf{position}(y,i)), \mathsf{position}(y,\mathsf{S}(i)))$$

is $\mathsf{verifier}(y_{\mathsf{S}(i)})$ with some possible added parts after the application of the second part of the procedure defining $\mathsf{TMsimulate}$ to the previous entries; as $\mathsf{verifier}(y_{\mathsf{S}(i)})$ is a proof and the potential added parts arising from the second part of the procedure do not change its proof nature (since they make use of an axiom and codify a sound reasoning, namely one use of *modus ponens*), we know that

$$\begin{aligned}&\mathsf{upto}(\mathsf{simulate}(y), \mathsf{position}(y,\mathsf{S}(i))) = \mathsf{upto}(\mathsf{simulate}(y), \mathsf{position}(y,i))\\ &* \mathsf{fromto}(\mathsf{simulate}(y), \mathsf{S}(\mathsf{position}(y,i)), \mathsf{position}(y,\mathsf{S}(i)))\end{aligned}$$

is also a proof. Now suppose that $j, m < \mathsf{S}(i)$ are such that $y_m = y_j \dot{\to} y_{\mathsf{S}(i)}$. If $m < i$, then, up to applications of the second part of the definition of $\mathsf{simulate}$ that do not interfere with the considered position,

$$\begin{aligned}&\mathsf{upto}(\mathsf{simulate}(y), \mathsf{position}(y,\mathsf{S}(i))) =\\ &\qquad \mathsf{upto}(\mathsf{simulate}(y), \mathsf{position}(y,i)) * \langle \ulcorner \mathsf{Pr}_\xi(\dot{y}_{\mathsf{S}(i)})\urcorner\rangle\end{aligned}$$

and, by construction of $\mathsf{simulate}$ and induction hypothesis[20], we have

$$\ulcorner \mathsf{Pr}_\xi(\dot{y_j})\urcorner, \ulcorner \mathsf{Pr}_\xi(\dot{y_j}) \to \mathsf{Pr}_\xi(\dot{y}_{\mathsf{S}(i)})\urcorner \in \mathsf{upto}(\mathsf{simulate}(y), \mathsf{position}(y,i)),$$

[19]In fact, here we need to use a provable form of strong induction for G_2, namely

$$\mathsf{G}_2 \vdash \varphi(0) \land (\forall i.(\forall j \le i.\varphi(j)) \to \varphi(\mathsf{S}(i))) \to \forall i.\varphi(i),$$

and $\varphi(i) := i < \mathsf{L}(y) \to \mathsf{Prf}_\xi((\mathsf{simulate}(y))_{\mathsf{position}(y,i)}, \mathsf{upto}(\mathsf{simulate}(y), \mathsf{position}(y,i))$. See [39, pp. 35, 63] for further details on strong induction.

[20]This is the step that requires strong induction for one to be sure that $\ulcorner \mathsf{Pr}_\xi(\dot{y_j})\urcorner$ was already present in the previous construction.

so $\mathtt{upto}(\mathtt{simulate}(y),\mathtt{position}(y,\mathsf{S}(i)))$ is a proof[21]. Now, suppose $m = i$. Then, we have that $(\mathtt{simulate}(y))_{\mathtt{position}(y,i)} = \ulcorner \mathtt{Pr}_\xi(\dot{y_m})\urcorner$ and

$$\mathtt{fromto}(\mathtt{simulate}(y),\mathsf{S}(\mathtt{position}(y,i)),\mathtt{position}(y,\mathsf{S}(i))) =$$
$$\langle \ulcorner \mathtt{Pr}_\xi(\dot{y_m}) \to (\mathtt{Pr}_\xi(\dot{y_j}) \to \mathtt{Pr}_\xi(\dot{y}_{\mathsf{S}(i)}))\urcorner, \ulcorner \mathtt{Pr}_\xi(\dot{y_j}) \to \mathtt{Pr}_\xi(\dot{y}_{\mathsf{S}(i)})\urcorner, \ulcorner \mathtt{Pr}_\xi(\dot{y}_{\mathsf{S}(i)})\urcorner\rangle$$

from the definition of simulate. As $\ulcorner \mathtt{Pr}_\xi(\dot{y_m}) \to (\mathtt{Pr}_\xi(\dot{y_j}) \to \mathtt{Pr}_\xi(\dot{y}_{\mathsf{S}(i)}))\urcorner$ is, by Assumption 1 on ξ defining G_2, an axiom[22] and, from the previously considered induction hypothesis,

$$\mathtt{Prf}_\xi((\mathtt{simulate}(y))_{\mathtt{position}(y,i)}, \mathtt{upto}(\mathtt{simulate}(y),\mathtt{position}(y,i)),$$

it follows, from what we observed, that

$$\mathtt{upto}(\mathtt{simulate}(y),\mathtt{position}(y,\mathsf{S}(i)))$$

is a proof.

In sum, we have established the desired result using induction; going outside G_2, we have confirmed that[23]

$$\mathsf{G}_2 \vdash \mathtt{Prf}_\xi(x,y) \to \mathtt{Prf}_\xi(\ulcorner \mathtt{Pr}_\xi(\dot{x})\urcorner, \mathtt{simulate}(y)),$$

and so $\mathsf{G}_2 \vdash \mathtt{Pr}_\xi(x) \to \mathtt{Pr}_\xi(\ulcorner \mathtt{Pr}_\xi(\dot{x})\urcorner)$. ⊣

We are now going to confirm that we can discharge any assumption made on Theorem 6.5.2 under the proviso that we only focus our attention to a finite set of axioms (here we are also including the logical axioms).

Definition 6.5.1. Given a finite set of axioms (here we are also considering this restriction to the logical axioms) $\mathcal{S} = \{\varphi_0, \dots, \varphi_n\} \subseteq \{\varphi | \mathsf{G}_2 \vdash \xi(\ulcorner\varphi\urcorner)\}$, we define $\mathtt{Prf}^{\mathcal{S}}(x,y) := \mathtt{Prf}_{\theta^{\mathcal{S}}}(x,y)$, where $\theta^{\mathcal{S}}(x) := \bigvee_{i=0}^{n} x = \ulcorner\varphi_i\urcorner$. Furthermore, we define $\mathtt{Pr}^{\mathcal{S}}(x) := \exists y.\mathtt{Prf}^{\mathcal{S}}(x,y)$.

Proposition 6.5.2. $\mathsf{G}_2 \vdash \forall x.\forall y.(\mathtt{Prf}^{\mathcal{S}}(x,y) \to \mathtt{Pr}_\xi(\ulcorner \mathtt{Pr}_\xi(\dot{x})\urcorner))$.

[21] Here we are using the internalized symbol $\in$ to express that a certain x is an element of the sequence y, i.e. $x \in y$; clearly, this is definable in G_2.

[22] This assumption is not strictly necessary to require, but it allows this proof to be much simpler. It can be substituted by the assumption that there is f such that $\mathsf{G}_2 \vdash \mathtt{Prf}_\xi(\ulcorner \mathtt{Pr}_\xi(x \dot{\to} y) \to \mathtt{Pr}_\xi(\dot{x} \dot{\to} \dot{y})\urcorner, \mathsf{f}(x,y))$. Let us briefly see why. Consider g such that $\mathsf{G}_2 \vdash \mathtt{Prf}_\xi(a \dot{\to} b, x) \wedge \mathtt{Prf}_\xi(b \dot{\to} c, y). \to \mathtt{Prf}_\xi(a \dot{\to} c, \mathsf{g}(x,y))$. We know that **C3** holds, so, by **C1** and **C4**, $\mathsf{G}_2 \vdash \forall x.\forall y.\exists z.\mathtt{Prf}_\xi(\ulcorner \mathtt{Pr}_\xi(\dot{x} \dot{\to} \dot{y}) \to (\mathtt{Pr}_\xi(\dot{x}) \to \mathtt{Pr}_\xi(\dot{y}))\urcorner, z)$. Using pairing and Theorem 6.3.2 to exploit the fact that $\mathtt{Prf}_\xi$ is $\Delta_0(\mathsf{G}_2)$, we can guarantee the existence of h such that $\mathsf{G}_2 \vdash \forall x.\forall y.\mathtt{Prf}_\xi(\ulcorner \mathtt{Pr}_\xi(\dot{x} \dot{\to} \dot{y}) \to (\mathtt{Pr}_\xi(\dot{x}) \to \mathtt{Pr}_\xi(\dot{y}))\urcorner, \mathsf{h}(x,y))$. Consequently, the mentioned axiom can be substituted by the use of $\mathsf{g}(\mathsf{f}(x,y),\mathsf{h}(x,y))$.

[23] We initially assumed $\mathtt{Prf}_\xi(x,y)$ and then used induction, this does not constitute a problem, since

$$\mathsf{G}_2 \vdash (\mathtt{Prf}_\xi(x,y) \to \forall i < \mathsf{L}(y).\Phi(i)) \leftrightarrow \forall i.(\mathtt{Prf}_\xi(x,y) \to (i < \mathsf{L}(y) \to \Phi(i))).$$

Proof. From Proposition 6.5.1, for $0 \leq i \leq n$,

$$\mathsf{G}_2 \vdash v = \ulcorner \varphi_i \urcorner \to \mathsf{Pr}_\xi(\ulcorner \dot{v} = \ulcorner \varphi_i \urcorner \urcorner).$$

As $\mathsf{G}_2 \vdash v = \ulcorner \varphi_i \urcorner \to \xi(v)$, then $\mathsf{G}_2 \vdash \mathsf{Pr}_\xi(\ulcorner \dot{v} = \ulcorner \varphi_i \urcorner \urcorner) \to \mathsf{Pr}_\xi(\ulcorner \xi(\dot{v}) \urcorner)$, and so $\mathsf{G}_2 \vdash v = \ulcorner \varphi_i \urcorner \to \mathsf{Pr}_\xi(\ulcorner \xi(\dot{v}) \urcorner)$. Having in mind that $\mathsf{G}_2 \vdash \xi \to \mathsf{Pr}_\xi$ and what we just concluded, $\mathsf{G}_2 \vdash \forall v.\theta^{\mathcal{S}}(v) \to \mathsf{Pr}_\xi(\ulcorner \mathsf{Pr}_\xi(\dot{v}) \urcorner)$. From Theorem 6.3.1, this entails the existence of a function-symbol $\mathtt{verifier}^{\mathcal{S}}$ such that

$$\mathsf{G}_2 \vdash \theta^{\mathcal{S}}(x) \to \mathsf{Prf}_\xi(\ulcorner \mathsf{Pr}_\xi(\dot{x}) \urcorner, \mathtt{verifier}^{\mathcal{S}}(x)).$$

The rest of the result follows from a construction similar to the one presented for the proof of Theorem 6.5.2; so we can create a function $\mathtt{simulate}^{\mathcal{S}}$ such that

$$\mathsf{G}_2 \vdash \mathsf{Prf}^{\mathcal{S}}(x, y) \to \mathsf{Prf}_\xi(\ulcorner \mathsf{Pr}_\xi(\dot{x}) \urcorner, \mathtt{simulate}^{\mathcal{S}}(y)).$$

⊣

Corollary 6.5.3. $\mathsf{G}_2 \vdash \mathsf{Pr}^{\mathcal{S}}(x) \to \mathsf{Pr}_\xi(\ulcorner \mathsf{Pr}_\xi(\dot{x}) \urcorner)$.

Proof. Immediate. ⊣

Definition 6.5.2. We say that ξ_0 is a *reasonable numeration* if there are formulas $\varphi_0, \dots, \varphi_n$ such that, provably in G_2,

$$\begin{aligned}
&\xi_0(v) \leftrightarrow \Big(\exists \mathtt{Formula}(x).\exists \mathtt{Formula}(y).v = x \dot{\to} (y \dot{\to} x)\Big) \vee \\
&\Big(\exists \mathtt{Formula}(x).\exists \mathtt{Formula}(y).\exists \mathtt{Formula}(z). \\
&v = (x \dot{\to} (y \dot{\to} z)) \dot{\to} ((x \dot{\to} y) \dot{\to} (x \dot{\to} z))\Big) \vee \\
&\Big(\exists \mathtt{Formula}(x).\exists \mathtt{Formula}(y).v = (\dot{\neg} y \dot{\to} \dot{\neg} x) \dot{\to} (x \dot{\to} y)\Big) \vee \\
&\Big(\exists \mathtt{Formula}(y).\exists \mathtt{Variable}(x).\exists \mathtt{Term}(t).v = \dot{\forall} x.y \dot{\to} \mathsf{sub}(y, t)\Big) \vee \\
&\Big(\exists \mathtt{Formula}(y).\exists \mathtt{Variable}(z).\exists \mathtt{Variable}(x).v = (\dot{\forall} x.(y \dot{\to} z)) \dot{\to} ((\dot{\forall} x.y) \dot{\to} (\dot{\forall} x.z))\Big) \vee \\
&\Big(\exists \mathtt{Formula}(y).\exists \mathtt{Variable}(x).\mathtt{notfree}(x, y) \wedge v = y \dot{\to} \dot{\forall} x.y\Big) \vee \\
&\Big(\exists \mathtt{Variable}(x).v = x \dot{=} x\Big) \vee \Big(\exists \mathtt{Variable}(x).\exists \mathtt{Variable}(y).\exists \mathtt{Formula}(z). \\
&v = (x \dot{=} y) \dot{\to} (\mathsf{sub}(z, x) \dot{\to} \mathsf{sub}(z, y))\Big) \vee \left(\bigvee_{i=0}^{n} v = \ulcorner \varphi_i \urcorner\right).
\end{aligned}$$

(We are considering $\dot{\forall}$ such that $\mathsf{G}_2 \vdash (\dot{\forall} \ulcorner x \urcorner . \ulcorner \varphi \urcorner) = \ulcorner \forall x.\varphi \urcorner$, not to be confused with the similar $\overset{\bullet}{\overbrace{\forall \vec{x}.}}$; in fact, $\mathsf{G}_2 \vdash (\dot{\forall} \ulcorner x \urcorner . \ulcorner \varphi \urcorner) = \overset{\bullet}{\overbrace{\forall x.}} \ulcorner \varphi \urcorner$.) We define the following sets:

$$\begin{aligned}
\mathcal{A} := \{&x \dot{\to} (y \dot{\to} x), (x \dot{\to} (y \dot{\to} z)) \dot{\to} ((x \dot{\to} y) \dot{\to} (x \dot{\to} z)), \\
&(\dot{\neg} y \dot{\to} \dot{\neg} x) \dot{\to} (x \dot{\to} y), y \dot{\to} \mathsf{sub}(y, t), (\dot{\forall} x.(y \dot{\to} z)) \dot{\to} ((\dot{\forall} x.y) \dot{\to} (\dot{\forall} x.z)), \\
&y \dot{\to} \dot{\forall} x.y, x \dot{=} x, (x \dot{=} y) \dot{\to} (\mathsf{sub}(z, x) \dot{\to} \mathsf{sub}(z, y))\},
\end{aligned}$$

and $\mathcal{B} := \{\mathtt{Formula}(x), \mathtt{Variable}(x), \mathtt{Term}(x), \mathtt{notfree}(y, x)\}$.

In the previous definition, ξ_0 enumerates a finitely axiomatizable theory of first-order arithmetic; [28, p. 112] confirms that the logical basis we presented is enough.

Theorem 6.5.3. (*Under Assumption 1*) *Suppose that ξ_0 is a reasonable numeration. Assume also that, for $f(\vec{x}) \in \mathscr{A}$ and $\Phi(\vec{x}) \in \mathscr{B}$:*

A: $\mathsf{G}_2 \vdash \Phi(\vec{x}) \to \mathtt{Pr}_\xi(\ulcorner \Phi(\dot{\vec{x}}) \urcorner)$;

B: $\mathsf{G}_2 \vdash \mathtt{Pr}_\xi(\ulcorner \dot{v} = \overset{\bullet}{\overbrace{f(\vec{x})}} \urcorner) \to \mathtt{Pr}_\xi(\ulcorner \dot{v} = f(\dot{\vec{x}}) \urcorner)$.

Then, $\mathsf{G}_2 \vdash \mathtt{Pr}_{\xi_0}(x) \to \mathtt{Pr}_\xi(\ulcorner \mathtt{Pr}_{\xi_0}(\dot{x}) \urcorner)$.

Proof. Let us focus on the first element of the disjunction characterizing ξ_0. From Proposition 6.5.1, we know that

$$\mathsf{G}_2 \vdash v = x \overset{\bullet}{\to} (y \overset{\bullet}{\to} x) \to \mathtt{Pr}_\xi(\ulcorner \dot{v} = \overset{\bullet}{\overbrace{x \overset{\bullet}{\to} (y \overset{\bullet}{\to} x)}} \urcorner).$$

So,

$$\mathsf{G}_2 \vdash (\exists \mathtt{Formula}(x).\exists \mathtt{Formula}(y).v = x \overset{\bullet}{\to} (y \overset{\bullet}{\to} x)) \to$$
$$(\exists \mathtt{Formula}(x).\exists \mathtt{Formula}(y).\mathtt{Pr}_\xi(\ulcorner \dot{v} = \overset{\bullet}{\overbrace{x \overset{\bullet}{\to} (y \overset{\bullet}{\to} x)}} \urcorner);$$

by **A**,

$$\mathsf{G}_2 \vdash (\exists \mathtt{Formula}(x).\exists \mathtt{Formula}(y).v = x \overset{\bullet}{\to} (y \overset{\bullet}{\to} x)) \to (\exists x.\exists y.$$
$$\mathtt{Pr}_\xi(\ulcorner \mathtt{Formula}(\dot{x}) \urcorner) \wedge \mathtt{Pr}_\xi(\ulcorner \mathtt{Formula}(\dot{y}) \urcorner) \wedge \mathtt{Pr}_\xi(\ulcorner \dot{v} = \overset{\bullet}{\overbrace{x \overset{\bullet}{\to} (y \overset{\bullet}{\to} x)}} \urcorner).$$

By **B**, it follows

$$\mathsf{G}_2 \vdash (\exists \mathtt{Formula}(x).\exists \mathtt{Formula}(y).v = x \overset{\bullet}{\to} (y \overset{\bullet}{\to} x)) \to (\exists x.\exists y.$$
$$\mathtt{Pr}_\xi(\ulcorner \mathtt{Formula}(\dot{x}) \urcorner) \wedge \mathtt{Pr}_\xi(\ulcorner \mathtt{Formula}(\dot{y}) \urcorner) \wedge \mathtt{Pr}_\xi(\ulcorner \dot{v} = \dot{x} \overset{\bullet}{\to} (\dot{y} \overset{\bullet}{\to} \dot{x}) \urcorner).$$

Using the derivability conditions, we can conclude[24]

$$\mathsf{G}_2 \vdash (\exists \mathtt{Formula}(x).\exists \mathtt{Formula}(y).v = x \overset{\bullet}{\to} (y \overset{\bullet}{\to} x)) \to$$
$$\exists x.\exists y.\mathtt{Pr}_\xi(\ulcorner \mathtt{Formula}(\dot{x}) \wedge \mathtt{Formula}(\dot{y}) \wedge \dot{v} = \dot{x} \overset{\bullet}{\to} (\dot{y} \overset{\bullet}{\to} \dot{x}) \urcorner);$$

again from the derivability conditions, we obtain[25]

$$\mathsf{G}_2 \vdash (\exists \mathtt{Formula}(x).\exists \mathtt{Formula}(y).v = x \overset{\bullet}{\to} (y \overset{\bullet}{\to} x)) \to$$
$$\mathtt{Pr}_\xi(\ulcorner \exists \mathtt{Formula}(x).\exists \mathtt{Formula}(y).\dot{v} = x \overset{\bullet}{\to} (y \overset{\bullet}{\to} x) \urcorner).$$

[24] Here we use the fact, from $\mathsf{G}_2 \vdash A(\vec{x}) \to (B(\vec{x}) \to A(\vec{x}) \wedge B(\vec{x}))$, we can prove $\mathsf{G}_2 \vdash \mathtt{Pr}_\xi(\ulcorner A(\dot{\vec{x}}) \urcorner) \wedge \mathtt{Pr}_\xi(\ulcorner B(\dot{\vec{x}}) \urcorner) \to \mathtt{Pr}_\xi(\ulcorner A(\dot{\vec{x}}) \wedge B(\dot{\vec{x}}) \urcorner)$.

[25] In this step, we are using $\mathsf{G}_2 \vdash A(\vec{x}) \to \exists \vec{x}.A(\vec{x})$ to conclude $\mathsf{G}_2 \vdash \mathtt{Pr}_\xi(\ulcorner A(\dot{\vec{x}}) \urcorner) \to \mathtt{Pr}_\xi(\ulcorner \exists \vec{x}.A(\vec{x}) \urcorner)$.

We can prove in a similar way for the other elements of the disjunction characterizing ξ_0 (in particular, the last element of the disjunction was, in a sense, analyzed in the proof of Proposition 6.5.2). All this yields, using the derivability conditions[26], $\mathsf{G}_2 \vdash \xi_0(v) \to \mathsf{Pr}_\xi(\ulcorner \xi_0(\dot{v})\urcorner)$, and consequently

$$\mathsf{G}_2 \vdash \xi_0(v) \to \mathsf{Pr}_\xi(\ulcorner \mathsf{Pr}_{\xi_0}(\dot{v})\urcorner).$$

It is important to observe that ξ_0 is a $\Delta_0(\mathsf{G}_2)$-formula, since each element of the disjunction can be written as a Δ_0-formula; for example, again focusing on the first element,

$$\begin{aligned}\mathsf{G}_2 \vdash &(\exists \mathtt{Formula}(x).\exists \mathtt{Formula}(y).v = x \mathbin{\dot{\to}} (y \mathbin{\dot{\to}} x)) \leftrightarrow \\ &(\exists x \leq v.\exists y \leq v.\mathtt{Formula}(x) \wedge \mathtt{Formula}(y) \wedge v = x \mathbin{\dot{\to}} (y \mathbin{\dot{\to}} x)).\end{aligned}$$

From Theorem 6.3.1, we can guarantee the existence of a function-symbol `verifier` such that $\mathsf{G}_2 \vdash \xi_0(x) \to \mathsf{Prf}_\xi(\ulcorner \mathsf{Pr}_{\xi_0}(\dot{x})\urcorner, \mathtt{verifier}(x))$. The rest of the proof follows from the construction used in the proof of Theorem 6.5.2. ⊣

Theorem 6.5.4. (*Under Assumption 1*) *Suppose that* ξ_0 *is a reasonable numeration. Assume also that, for* $f(\vec{x}) \in \mathscr{A}$ *and* $\Phi(\vec{x}) \in \mathscr{B}$*:*

C: $\mathsf{G}_2 \vdash \exists \Phi(\vec{x}).\mathsf{Pr}_\xi(\ulcorner \dot{v} = \dot{\overbrace{f(\vec{x})}} \urcorner) \to \mathsf{Pr}_\xi(\ulcorner \exists \Phi(\vec{x}).\dot{v} = f(\vec{x})\urcorner)$.

Then, $\mathsf{G}_2 \vdash \mathsf{Pr}_{\xi_0}(x) \to \mathsf{Pr}_\xi(\ulcorner \mathsf{Pr}_{\xi_0}(\dot{x})\urcorner)$.

Proof. Similar to the previous proof. ⊣

6.6 A form of internal completeness in G_2

We start this section with a useful proposition that allows us to center the discussion of internal Σ_1-completeness in G_2 in the study of that form of completeness for function-symbols of G_2.

Proposition 6.6.1. *The two following statements are equivalent:*

A: *For all* $\Sigma_1(\mathsf{G}_2)$*-formulas* $\varphi(\vec{x})$, $\mathsf{G}_2 \vdash \varphi(\vec{x}) \to \mathsf{Pr}_\xi(\ulcorner \varphi(\dot{\vec{x}})\urcorner)$*;*

B: *For all function-symbols* f, $\mathsf{G}_2 \vdash f(\vec{x}) = y \to \mathsf{Pr}_\xi(\ulcorner f(\dot{\vec{x}}) = \dot{y}\urcorner)$.

Proof. Clearly, **A** $\implies$ **B**. Let us prove the conserve implication. Consider $\varphi(\vec{x})$ a $\Sigma_1(\mathsf{G}_2)$-formula. There is a Σ_1-formula $\varphi_0(\vec{x}) := \exists \vec{y}.\psi(\vec{x},\vec{y})$, with ψ a bounded formula, such that $G \vdash \varphi \leftrightarrow \varphi_0$. From Proposition 6.3.1, we know that $G \vdash \psi(\vec{x},\vec{y}) \leftrightarrow \chi_\psi(\vec{x},\vec{y}) = \overline{1}$. So,

[26]Namely, $\mathsf{G}_2 \vdash \mathsf{Pr}_\xi(\ulcorner A(\dot{\vec{x}})\urcorner) \vee \mathsf{Pr}_\xi(\ulcorner B(\dot{\vec{x}})\urcorner) \to \mathsf{Pr}_\xi(\ulcorner A(\dot{\vec{x}}) \vee B(\dot{\vec{x}})\urcorner)$.

by assumption, $\mathsf{G}_2 \vdash \chi_\psi(\vec{x},\vec{y}) = \overline{1} \to \mathsf{Pr}_\xi(\ulcorner \chi_\psi(\dot{\vec{x}},\dot{\vec{y}}) = \overline{1}\urcorner)$; using the derivability conditions, we can thus conclude $\mathsf{G}_2 \vdash \psi(\vec{x},\vec{y}) \to \mathsf{Pr}_\xi(\ulcorner \psi(\dot{\vec{x}},\dot{\vec{y}})\urcorner)$. Hence[27], $\mathsf{G}_2 \vdash \exists \vec{y}.\psi(\vec{x},\vec{y}) \to \mathsf{Pr}_\xi(\ulcorner \exists\vec{y}.\psi(\dot{\vec{x}},\vec{y})\urcorner)$. ⊣

Having the previous proposition in mind, we might just study condition **B**; in fact, we are going to study a slight variation of it, namely

$$\mathsf{G}_2 \vdash f(\vec{x}) = \overline{n} \to \mathsf{Pr}_\xi(\ulcorner f(\dot{\vec{x}}) = \overline{n}\urcorner),$$

with f a function-symbol in a suitable class and $n \in \mathbb{N}$. We proceed our aim with a simple proposition.

Proposition 6.6.2. *Suppose that ξ is such that $\mathsf{G}_2 \vdash \forall \mathsf{Term}(x).\forall\mathsf{Term}(y).x \le y \to \xi(x \mathbin{\dot{\le}} y)$. Then, $\mathsf{G}_2 \vdash \forall x.\forall y.x \le y \to \mathsf{Pr}_\xi(\ulcorner \dot{x} \le \dot{y}\urcorner)$.*

Proof. Similar to the proof of Proposition 6.5.1. ⊣

Throughout the rest of this section, we are going to assume that ξ is a numeration of a theory T satisfying the conditions of the previous result.

Definition 6.6.1. We define the class $\mathsf{FuncFin}$ of function-symbols of G_2 by recursion:

Basis Case: $\mathsf{S}, +, \mathsf{min}, \mathsf{max}, \pi^k_j \in \mathsf{FuncFin}$.

Finite Composition: If $f, g_0, \ldots, g_k \in \mathsf{FuncFin}$ and[28], for each $n \in \mathbb{N}$, there are $\vec{m_{i,j}} \in \mathbb{N}$, with $1 \le i \le \ell$, such that[29]

$$\mathsf{G}_2 \vdash \mathsf{COMP}[f; g_0, \ldots, g_k](\vec{x}) = \overline{n} \leftrightarrow \bigvee_{i=1}^{\ell}\left(\bigwedge_{j=0}^{k} g_j(\vec{x}) = \overline{\vec{m_{i,j}}}\right),$$

then $\mathsf{COMP}[f; g_0, \ldots, g_k] \in \mathsf{FuncFin}$.

Finite Bounded Recursion: If f is defined by bounded recursion from t_0, t_1, $t_2 \in \mathsf{FuncFin}$ and, for each $n \in \mathbb{N}$, there are $g_0, \ldots, g_k \in \mathsf{FuncFin}$ already defined and $\vec{m_i} \in \mathbb{N}$, with $1 \le i \le \ell$, such that

$$\mathsf{BR}[t_0, t_1; t_2](x) = \overline{n} \leftrightarrow \bigvee_{i=0}^{k} g_i(x) = \overline{\vec{m_i}};$$

then $\mathsf{BR}[t_0, t_1; t_2] \in \mathsf{FuncFin}$.

Theorem 6.6.1 (Form of Internal Completeness for $\mathsf{FuncFin}$). *For each $n \in \mathbb{N}$, if $f \in \mathsf{FuncFin}$, then $\mathsf{G}_2 \vdash f(\vec{x}) = \overline{n} \to \mathsf{Pr}_\xi(\ulcorner f(\dot{\vec{x}}) = \overline{n}\urcorner)$.*

---

[27]This follows from the following fact: from $\mathsf{G}_2 \vdash \Phi(\vec{x}) \to \exists\vec{x}.\Phi(\vec{x})$, we can conclude $\mathsf{G}_2 \vdash \mathsf{Pr}_\xi(\ulcorner\Phi(\dot{\vec{x}}) \to \exists\vec{x}.\Phi(\vec{x})\urcorner)$, and so $\mathsf{G}_2 \vdash \mathsf{Pr}_\xi(\ulcorner\Phi(\dot{\vec{x}})\urcorner) \to \mathsf{Pr}_\xi(\ulcorner\exists\vec{x}.\Phi(\vec{x})\urcorner)$, hence $\mathsf{G}_2 \vdash \exists\vec{x}.\mathsf{Pr}_\xi(\ulcorner\Phi(\dot{\vec{x}})\urcorner) \to \mathsf{Pr}_\xi(\ulcorner\exists\vec{x}.\Phi(\vec{x})\urcorner)$.

[28]We call this other condition *finiteness*.

[29]If $\ell = 0$ we assume $\bigvee_{i=1}^{\ell}\Phi_i := \bot$.

Proof. Let us prove by induction on Definition 6.6.1.

Basis case: **S:** If $n = 0$, then $\mathsf{G}_2 \vdash \mathsf{S}(x) = 0 \to \perp$, and so $\mathsf{G}_2 \vdash \mathsf{S}(x) = 0 \to \mathsf{Pr}_\xi(\ulcorner \mathsf{S}(\dot{x}) = \overline{n} \urcorner)$. Assume $n > 0$. Then, either $\mathsf{G}_2 \vdash \overline{n} = \overline{2k+1}$ or $\mathsf{G}_2 \vdash \overline{n} = \overline{2(k+1)}$. Firstly, let us assume $\mathsf{G}_2 \vdash \overline{n} = \overline{2k+1}$. Then, $\mathsf{G}_2 \vdash \mathsf{S}(x) = \overline{n} \to \mathsf{S}(x) = \overline{2k+1}$, and so $\mathsf{G}_2 \vdash \mathsf{S}(x) = \overline{n} \to \mathsf{S}(x) = \overline{2k} + \overline{1}$, thus $\mathsf{G}_2 \vdash \mathsf{S}(x) = \overline{n} \to \mathsf{S}(x) = \mathsf{S}(\overline{2k})$. Therefore, $\mathsf{G}_2 \vdash \mathsf{S}(x) = \overline{n} \to x = \overline{2k}$, and using Proposition 6.5.1, $\mathsf{G}_2 \vdash \mathsf{S}(x) = \overline{n} \to \mathsf{Pr}_\xi(\ulcorner \dot{x} = \overline{2k} \urcorner)$. This implies that $\mathsf{G}_2 \vdash \mathsf{S}(x) = \overline{n} \to \mathsf{Pr}_\xi(\ulcorner \mathsf{S}(\dot{x}) = \overline{2k} + \overline{1} \urcorner)$, and so $\mathsf{G}_2 \vdash \mathsf{S}(x) = \overline{n} \to \mathsf{Pr}_\xi(\ulcorner \mathsf{S}(\dot{x}) = \overline{2k+1} \urcorner)$, consequently $\mathsf{G}_2 \vdash \mathsf{S}(x) = \overline{n} \to \mathsf{Pr}_\xi(\ulcorner \mathsf{S}(\dot{x}) = \overline{n} \urcorner)$.

Now let us assume $\mathsf{G}_2 \vdash \overline{n} = \overline{2(k+1)}$. Then, $\mathsf{G}_2 \vdash \mathsf{S}(x) = \overline{n} \to \mathsf{S}(x) = \overline{2(k+1)}$, and so $\mathsf{G}_2 \vdash \mathsf{S}(x) = \overline{n} \to x = \overline{2(k+1)} \dot{-} \overline{1}$. Hence, $\mathsf{G}_2 \vdash \mathsf{S}(x) = \overline{n} \to x = \overline{2k+1}$; again using Proposition 6.5.1, $\mathsf{G}_2 \vdash \mathsf{S}(x) = \overline{n} \to \mathsf{Pr}_\xi(\ulcorner \dot{x} = \overline{2k+1} \urcorner)$. Consequently, $\mathsf{G}_2 \vdash \mathsf{S}(x) = \overline{n} \to \mathsf{Pr}_\xi(\ulcorner \mathsf{S}(\dot{x}) = \mathsf{S}(\overline{2k+1}) \urcorner)$, thus $\mathsf{G}_2 \vdash \mathsf{S}(x) = \overline{n} \to \mathsf{Pr}_\xi(\ulcorner \mathsf{S}(\dot{x}) = \overline{n} \urcorner)$.

+: If $n = 0$, then $\mathsf{G}_2 \vdash x + y = 0 \to x = 0 \wedge y = 0$; using Proposition 6.5.1 this yields $\mathsf{G}_2 \vdash x + y = 0 \to \mathsf{Pr}_\xi(\ulcorner \dot{x} = 0 \wedge \dot{y} = 0 \urcorner)$, so $\mathsf{G}_2 \vdash x + y = 0 \to \mathsf{Pr}_\xi(\ulcorner \dot{x} + \dot{y} = 0 \urcorner)$. Let us now assume $n > 0$. It is not hard to prove that

$$\mathsf{G}_2 \vdash x + y = \overline{n} \leftrightarrow \bigvee_{i=0}^{n} (x = \overline{i} \wedge y = \overline{n-i}). \tag{I}$$

So, from Proposition 6.5.1, $\mathsf{G}_2 \vdash x + y = \overline{n} \to \bigvee_{i=0}^{n} \mathsf{Pr}_\xi(\ulcorner \dot{x} = \overline{i} \wedge \dot{y} = \overline{n-i} \urcorner)$, and so[30] $\mathsf{G}_2 \vdash x + y = \overline{n} \to \mathsf{Pr}_\xi(\ulcorner \bigvee_{i=0}^{n} (\dot{x} = \overline{i} \wedge \dot{y} = \overline{n-i}) \urcorner)$, consequently, again from (I), $\mathsf{G}_2 \vdash x + y = \overline{n} \to \mathsf{Pr}_\xi(\ulcorner \dot{x} + \dot{y} = \overline{n} \urcorner)$.

min: It is not hard to see that $\mathsf{G}_2 \vdash \mathtt{min}(x, y) = \overline{n} \leftrightarrow (x = \overline{n} \wedge x \leq y) \vee (y = \overline{n} \wedge y \leq x)$. The result follows from Propositions 6.5.1 and 6.6.2.

max: Similar to the previous case.

Finite Composition: Assume, by induction hypothesis, that the result holds for function-symbols $f, g_0, \ldots, g_k \in \mathsf{FuncFin}$ and that these function-symbols satisfy the condition of finiteness of Definition 6.6.1. Then, there are $\vec{m_{i,j}} \in \mathbb{N}$, with $1 \leq i \leq \ell$, such that

$$\mathsf{G}_2 \vdash \mathsf{COMP}[f; g_0, \ldots, g_k](\vec{x}) = \overline{n} \leftrightarrow \bigvee_{i=1}^{\ell} \left(\bigwedge_{j=0}^{k} g_j(\vec{x}) = \overline{\vec{m_{i,j}}} \right). \tag{II}$$

So, we get by induction hypothesis that $\mathsf{G}_2 \vdash \mathsf{COMP}[f; g_0, \ldots, g_k](\vec{x}) = \overline{n} \to \bigvee_{i=1}^{\ell} \mathsf{Pr}_\xi(\ulcorner \bigwedge_{j=0}^{k} g_j(\dot{\vec{x}}) = \overline{\vec{m_{i,j}}} \urcorner)$. From a property previously used, we conclude

$$\mathsf{G}_2 \vdash \mathsf{COMP}[f; g_0, \ldots, g_k](\vec{x}) = \overline{n} \to \mathsf{Pr}_\xi(\ulcorner \bigvee_{i=1}^{\ell} (\bigwedge_{j=0}^{k} g_j(\dot{\vec{x}}) = \overline{\vec{m_{i,j}}}) \urcorner).$$

From (II) we obtain the desired property.

[30] This follows from the fact that $\mathsf{G}_2 \vdash \mathsf{Pr}_\xi(\ulcorner \varphi(\dot{\vec{x}}) \urcorner) \vee \mathsf{Pr}_\xi(\ulcorner \psi(\dot{\vec{x}}) \urcorner) \to \mathsf{Pr}_\xi(\ulcorner \varphi(\dot{\vec{x}}) \vee \psi(\dot{\vec{x}}) \urcorner)$.

Finite Bounded Recursion: Assume, by induction hypothesis, that the result holds for $t_0, t_1, t_2 \in \texttt{FuncFin}$ and that these function-symbols satisfy the conditions of finiteness of Definition 6.6.1. Then, for $n \in \mathbb{N}$, there are $g_0, \ldots, g_k \in \texttt{FuncFin}$ already defined and $\vec{m_i} \in \mathbb{N}$, with $1 \leq i \leq \ell$, such that

$$\mathsf{BR}[t_0, t_1; t_2](x) = \overline{n} \leftrightarrow \bigvee_{i=0}^{k} g_i(x) = \overline{\vec{m_i}}\,. \tag{III}$$

Using a previously made reasoning and (III), we obtain the desired result.

⊣

Chapter 7

Conclusions and Future Work

The main topic of this thesis was the study of *notions of provability*, *videlicet* formulas $B(x,y)$ that satisfy $T \vdash \varphi \iff \exists n \in \mathbb{N}. T \vdash B(\ulcorner\varphi\urcorner, \overline{n})$. The following are examples of studied illustrious notions of provability: the usual notion of provability (Pr_T), k-provability ($T \vdash_{k \text{ steps}} \cdot$), and s-symbols provability ($T \vdash_{s \text{ symbols}} \cdot$). We presented general results concerning notions of provability—see, for example, Theorems 4.3.1 and 5.7.1—and we studied particular notions of provability.

In Chapter 2, we presented new relations between the First Incompleteness Theorem (G1), Undefinability of Truth, and recursion. Our approach is general enough for one to get the "big picture", but specific enough for one not to get lost with too much generality—in that chatper, we are always in the realm of (first-order) theories of arithmetic. We related G1 with Rice's Theorem (a major theorem in recursion) by developing a version of Kleene's Normal Form with formulas and provability; we studied the interplay between G1 and the non-recursiveness of truth via recursion; and we presented a general arithmetical form of the Diagonalization Lemma that is responsible for several major results, *scilicet* G1 the Undefinability of Truth, and Hilbert-Bernays Paradox.

We studied the decidability of k-provability in PA in Chapter 3—the relation 'being provable in PA with at most k steps'—and the decidability of the proof-skeleton problem—the problem of deciding if a given formula has a proof that has a given skeleton (the list of axioms and rules that were used). The decidability of k-provability for the usual Hilbert-style formalization of PA is still an open problem, but it is known that the proof-skeleton problem is undecidable for that theory. Using new methods, we presented a characterization of some numbers k for which k-provability is decidable, and we presented a characterization of some proof-skeletons for which one can decide whether a formula has a proof whose skeleton is the considered one. These characterizations are natural and parameterized by unification algorithms.

In Chapter 4, we studied Kreisel's Conjecture: if, for all $n \in \mathbb{N}$, $\mathsf{PA} \vdash_{k \text{ steps}} \varphi(\overline{n})$, then $\mathsf{PA} \vdash \forall x.\varphi(x)$. For a theory of arithmetic T, given a recursive function h, $T \vdash_{\leq h} \varphi$ holds if there is a proof of φ in T whose code is at most $h(\#\varphi)$ (this notion depends on the underlying coding). We created $P^h_T(x)$ a predicate for $\vdash_{\leq h}$ in T. We showed that there is a sentence φ and a total recursive function h such that $T \vdash_{\leq h} \mathsf{Pr}_T(\ulcorner \mathsf{Pr}_T(\ulcorner \varphi \urcorner) \rightarrow \varphi \urcorner)$, but $T \not\vdash_{\leq h} \varphi$, where Pr_T stands for the standard provability predicate in T. This statement is related to a conjecture by Montagna. We also studied variants and weakenings of Kreisel's Conjecture. By the use of reflection principles, we obtained a theory T^h_Γ that extends T such that a version of Kreisel's conjecture holds: given a recursive function h and $\varphi(x)$ a Γ-formula (where Γ is an arbitrarily fixed class of formulas) such that, for all $n \in \mathbb{N}$, $T \vdash_{\leq h} \varphi(\overline{n})$, then $T^h_\Gamma \vdash \forall x.\varphi(x)$. Several derivability conditions were studied for a theory to satisfy the following implication: if $T \vdash \forall x.P^h_T(\ulcorner \varphi(\dot{x}) \urcorner)$, then $T \vdash \forall x.\varphi(x)$. This corresponds to an arithmetization of Kreisel's conjecture. It was shown that, for certain theories, there is a function h such that $\vdash_{k \text{ steps}} \subseteq \vdash_{\leq h}$.

We studied Numeral forms of completeness and consistency for S^1_2 and other weak theories, like EA, in Chapter 5. This gave rise to: an exploration of the derivability conditions needed to establish the mentioned results; a presentation of a weak form of Gödel's Second Incompleteness Theorem without using 'provability implies provable provability'; a provability predicate that satisfies the mentioned derivability condition for weak theories; and a completeness result via consistency statements. Moreover, the Chapter 5 includes characterizations of the provability predicates for which the numeral results hold, having EA as the surrounding theory, and results on functions that compute finitist consistency statements. We ended Chapter 5 by drawing some philosophical implications of our results.

Chapter 6 was devoted to the study of the derivability condition 'provability implies provable provability', namely $\mathsf{Pr}_\xi(x) \rightarrow \mathsf{Pr}_\xi(\ulcorner \mathsf{Pr}_\xi(\dot{x}) \urcorner)$. As we mentioned, this condition is very sensitive to the underlying theory, for example it is an open problem if it holds for $\mathsf{I}\Delta_0$. We created a weak theory G_2 to study this condition; this is a theory for the class FLINSPACE. We also relate properties of G_2 to equality between computational classes.

There are several future lines of investigation having our thesis as a starting point:

- Express other results of recursion theory in the framework of Chapter 2;
- Extend the methods developed in Chapter 3 to a more general framework, by allowing a different type of algorithms;
- Explore the modal logic of P^h_T from Chapter 4;
- Find further consequences of Numeral Completeness from Chapter 5;
- Use the methods of Chapter 6 to further study the derivability condition 'provability implies provable provability'.

Bibliography

[1] M. Baaz and P. Pudlák. "Kreisel's Conjecture for L∃1". In: *Arithmetic, Proof Theory and Computational Complexity*. Ed. by P. Clote and J. Krajićek. Oxford University, 1993, pp. 29–59.

[2] L. Beklemishev. "A proof-theoretic analysis of collection". In: *Archive for Mathematical Logic* 37.5 (1998), pp. 275–296.

[3] L. Beklemishev. "Induction rules, reflection principles, and provably recursive functions". In: *Annals of Pure and Applied Logic* 85.3 (1997), pp. 193–242.

[4] J. L. Bell. "Incompleteness in a general setting". In: *Bulletin of Symbolic Logic* 13.1 (2007), pp. 21–30.

[5] A. Berarducci and B. Intrigila. "Combinatorial principles in elementary number theory". In: *Annals of Pure and Applied Logic* 55.1 (1991), pp. 35–50.

[6] R. Blanck and A. Enayat. "Marginalia on a theorem of Woodin". In: *Journal of Symbolic Logic* 82.1 (2017), pp. 359–374.

[7] G. S. Boolos, J. P. Burgess, and R. C. Jeffrey. *Computability and Logic*. Cambridge University Press, 2007. isbn: 9781107049277.

[8] V. Brattka, G. Gherardi, and A. Pauly. "Weihrauch Complexity in Computable Analysis". In: *Handbook of Computability and Complexity in Analysis*. Ed. by V. Brattka and P. Hertling. Springer International Publishing, 2021, pp. 367–417. isbn: 978-3-030-59234-9.

[9] P. Bruscoli and A. Guglielmi. "On the proof complexity of deep inference". In: *ACM Transactions on Computational Logic (TOCL)* 10.2 (2009), pp. 1–34.

[10] S. Buss. *Bounded arithmetic*. Vol. 86. Bibliopolis Napoli, 1986.

[11] S. Buss. *Handbook of Proof Theory*. Elsevier Science, 1998. isbn: 9780080533186.

[12] S. Buss. "On Gödel's theorems on lengths of proofs I: Number of lines and speedup for arithmetics". In: *Journal of Symbolic Logic* 59.3 (1994), pp. 737–756.

[13] S. Buss. "The undecidability of k-provability". In: *Annals of Pure and Applied Logic* 53.1 (1991), pp. 75–102.

[14] S. R. Buss. "On Herbrand's theorem". In: *International Workshop on Logic and Computational Complexity*. Springer. 1994, pp. 195–209.

[15] S. Cavagnetto. "The lengths of proofs: Kreisel's conjecture and Gödel's speed-up theorem". In: *Journal of Mathematical Sciences* 158.5 (2009), pp. 689–707.

[16] J. Cereceda. *An algorithm for the Faulhaber polynomials*. 2021. arXiv: `2103.08553 [math.NT]`.

[17] P. Clote. "Computation Models and Function Algebras". In: *Handbook of computability theory* 140 (1999), pp. 589–681.

[18] P. Clote and J. Krajíček. "Open Problems". In: *Arithmetic, Proof Theory and Computational Complexity*. Ed. by P. Clote and J. Krajíček. Oxford University, 1993, pp. 1–19.

[19] P. Clote and E. Kranakis. "Machine Models and Function Algebras". In: *Boolean Functions and Computation Models*. Springer, 2002, pp. 413–495.

[20] P. Clote and G. Takeuti. "Exponential time and bounded arithmetic". In: *Structure in complexity theory*. Springer. 1986, pp. 125–143.

[21] S. Cook and R. Reckhow. "On the lengths of proofs in the propositional calculus (preliminary version)". In: *Proceedings of the sixth annual ACM symposium on Theory of computing*. 1974, pp. 135–148.

[22] S. Cook and R. Reckhow. "The relative efficiency of propositional proof systems". In: *Journal of Symbolic Logic* 44.1 (1979), pp. 36–50.

[23] A. Cordón-Franco, A. Fernández-Margarit, and F. Lara-Martín. "Provably total primitive recursive functions: theories with induction". In: *International Workshop on Computer Science Logic*. Springer. 2004, pp. 355–369.

[24] C. Cornaros. "On Grzegorczyk induction". In: *Annals of Pure and Applied Logic* 74.1 (1995), pp. 1–21.

[25] A. Das and I. Oitavem. "A recursion-theoretic characterisation of the positive polynomial-time functions". In: *27th EACSL Annual Conference on Computer Science Logic (CSL 2018)*. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik. 2018.

[26] C. Dimitracopoulos. "Subsystems of Peano Arithmetic and Classical Results of Number Theory". In: *Filomat* (1995), pp. 473–481.

[27] A. Ehrenfeucht and S. Feferman. "Representability of recursively enumerable sets in formal theories". In: *Archiv für mathematische Logik und Grundlagenforschung* 5.1-2 (1960), pp. 37–41.

[28] H. B. Enderton. *A mathematical introduction to logic*. Elsevier, 2001.

[29] W. Ewald. *From Kant to Hilbert Volume 2: A Source Book in the Foundations of Mathematics*. Oxford University Press, 1996.

[30] W. Ewald and W. Sieg (eds.) *David Hilbert's Lectures on the Foundations of Arithmetic and Logic 1917–1933*. Springer, 2013.

[31] W. M. Farmer. “A unification-theoretic method for investigating the k-provability problem”. In: *Annals of Pure and Applied Logic* 51.3 (1991), pp. 173–214.

[32] W. M. Farmer. “The Kreisel length-of-proof problem”. In: *Annals of Mathematics and Artificial Intelligence* 6.1-3 (1992), pp. 27–55.

[33] S. Feferman. “Arithmetization of metamathematics in a general setting”. In: *Fundamenta Mathematicae* 49.1 (1960), pp. 35–92.

[34] F. Ferreira. “Polynomial Time Computable Arithmetic”. In: *Contemporary Mathematics, W. Sieg (ed.)* Vol. 106. AMS, 1990, pp. 137–156. ISBN: 9780821851104.

[35] H. Friedman. “One hundred and two problems in mathematical logic”. In: *Journal of Symbolic Logic* 40.2 (1975), pp. 113–129.

[36] H. Gaifman. “Naming and Diagonalization, from Cantor to Gödel to Kleene”. In: *Logic Journal of the IGPL* 14.5 (2006-10), pp. 709–728.

[37] K. Gödel. “Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme I”. In: *Monatshefte für mathematik und physik* 38.1 (1931), pp. 173–198.

[38] A. Grzegorczyk. *Some classes of recursive functions*. Warszawa: Instytut Matematyczny Polskiej Akademi Nauk, 1953.

[39] P. Hájek and P. Pudlák. *Metamathematics of First-Order Arithmetic*. Perspectives in Logic. Cambridge University Press, 1998. ISBN: 9781316739457.

[40] M. Hallett and U. Majer (eds.) *David Hilbert's Lectures on the Foundations of Geometry 1891–1902*. Springer, 2004. ISBN: 9783540643739.

[41] J. van Heijenoort (ed.) *From Frege to Gödel: A Source Book in Mathematical Logic, 1879-1931*. Source books in the history of the sciences. Harvard University Press, 1967. ISBN: 9780674324497.

[42] P. Henk and A. Visser. “Interpretability suprema in Peano Arithmetic”. In: *Archive for Mathematical Logic* 56.5 (2017), pp. 555–584.

[43] J. Herbrand. “Sur la non-contradiction de l'arithmétique”. In: *Journal für die reine und angewandte Mathematik* 166 (1932), 1–8, translated in [41].

[44] G. Herman. “A new hierarchy of elementary functions”. In: *Proceedings of the American Mathematical Society* 20.2 (1969), pp. 557–562.

[45] D. Hilbert. “Beweis des tertium non datur, Nachrichten von der Gesellschaft der Wissenschaften zu Göttingen”. In: *Mathematisch-physikalische Klasse* (1931), pp. 120–125.

[46] D. Hilbert. “Die Grundlegung der elementaren Zahlenlehre”. In: *Mathematische Annalen* 104.1 (1931), pp. 485–494.

[47] D. Hilbert. “Die logischen Grundlagen der Mathematik”. In: *Mathematische Annalen* 88 (1923), pp. 151–165.

[48] D. Hilbert. “Prinzipien der Mathematik. Lecture notes by Paul Bernays. Winter-Semester 1917–18”. In: Bibliothek, Mathematisches Institut, Universität Göttingen. (1918) (the typescript is found in [30], pp. 31–270).

[49] W. Hodges. *A shorter model theory*. Cambridge university press, 1997.

[50] S. Homer and A. Selman. *Computability and Complexity Theory*. Graduate Texts in Computer Science. Springer, 2001. ISBN: 9780387950556.

[51] R. G. Jeroslow. “Redundancies in the Hilbert-Bernays Derivability Conditions for Gödel’s Second Incompleteness Theorem”. In: *Journal Symbolic Logic* 38.3 (1973-12), pp. 359–367.

[52] J. Joosten and A. Visser. “The interpretability logic of all reasonable arithmetical theories”. In: *Erkenntnis* 53.1 (2000), pp. 3–26.

[53] R. Kahle and I. Oitavem. “Lorenzen between Gentzen and Schütte”. In: *Paul Lorenzen Mathematician and Logician*. Ed. by G. Heinzmann and G. Wolters. Springer, 2021, pp. 61–73.

[54] M. Kikuchi. “Kolmogorov complexity and the second incompleteness theorem”. In: *Archive for Mathematical Logic* 36.6 (1997), pp. 437–443.

[55] M. Kikuchi and T. Kurahashi. “Generalizations of Gödel’s Incompleteness Theorems for Σ_n-definable Theories of Arithmetic”. In: *Review of Symbolic Logic* 10.4 (2017), pp. 603–616.

[56] M. Kikuchi and T. Kurahashi. “Generalizations of Gödel’s Incompleteness Theorems for Σ_n-definable Theories of Arithmetic”. In: *Review of Symbolic Logic* 10.4 (2017), pp. 603–616. DOI: 10.1017/S1755020317000235.

[57] J. Krajíček. “On the number of steps in proofs”. In: *Annals of Pure and Applied Logic* 41.2 (1989), pp. 153–178.

[58] J. Krajíček and P. Pavel. “The number of proof lines and the size of proofs in first order logic”. In: *Archive for Mathematical Logic* 27.1 (1988), pp. 69–84.

[59] J. Krajíček and P. Pudlák. “Propositional proof systems, the consistency of first order theories and the complexity of computations”. In: *Journal of Symbolic Logic* 54.3 (1989), pp. 1063–1079.

[60] G. Kreisel. “Hilbert’s Programme”. In: *Dialectica* 12.3,–4 (1958), pp. 346–372.

[61] T. Kurahashi. “A note on derivability conditions”. In: *Journal of Symbolic Logic* 85.3 (2020), pp. 1224–1253.

[62] T. Kurahashi. “Arithmetical Soundness and Completeness for Σ_2 Numerations”. In: *Studia Logica* 106.6 (2018-02), pp. 1181–1196. DOI: 10.1007/s11225-017-9782-4. URL: https://doi.org/10.1007%5C%2Fs11225-017-9782-4.

[63] T. Kurahashi. “Provability logics relative to a fixed extension of Peano Arithmetic”. In: *Journal of Symbolic Logic* 83.3 (2018), pp. 1229–1246.

[64] T. Kurahashi. *Rosser provability and the second incompleteness theorem*. arXiv: 1902.06863. 2019. arXiv: 1902.06863 [math.LO].

[65] P. Lindström. *Aspects of Incompleteness*. Lecture Notes in Logic. Cambridge University Press, 2017. ISBN: 9781107167926.

[66] J. M. Lourenço. *The NOVAthesis LaTeX Template User's Manual*. NOVA University Lisbon. 2021. URL: https://github.com/joaomlourenco/novathesis/raw/master/template.pdf.

[67] P. Lutz and J. Walsh. "Incompleteness and jump hierarchies". In: *Proceedings of the American Mathematical Society* 148.11 (2020), pp. 4997–5006.

[68] T. Miyatake. "On the length of proofs in a formal systems". In: *Tsukuba J. Math.* 4.1 (1980-06), pp. 115–125.

[69] R. Murawski. *Recursive Functions and Metamathematics: Problems of Completeness and Decidability, Gödel's Theorems*. Synthese Library. Springer, 1999. ISBN: 9780792359043.

[70] K. G. Niebergall. "Intensionality in philosophy and metamathematics". In: *Intensionality*. Ed. by R. Kahle. Lecture Notes in Logic. Cambridge University Press, 2005, pp. 123–159.

[71] P. Odifreddi. *Classical Recursion Theory: The Theory of Functions and Sets of Natural Numbers*. Elsevier Science, 1992. ISBN: 9780080886596.

[72] V. P. Orevkov. *Complexity of Proofs and Their Transformations in Axiomatic Theories*. Translations of Mathematical Monographs, Vol. 128. American Mathematical Society, 1993. ISBN: 0821845764.

[73] R. J. Parikh. "Some results on the length of proofs". In: *Transactions of the American Mathematical Society* 177 (1973), pp. 29–36.

[74] H. Pavel. "Theories very close to *PA* where Kreisel's Conjecture is false". In: *Journal of Symbolic Logic* 72.1 (2007), pp. 123–137.

[75] G. Priest. "On a paradox of Hilbert and Bernays". In: *Journal of philosophical logic* 26.1 (1997), pp. 45–56.

[76] P. Pudlák. "On the length of proofs of finitistic consistency statements in first order theories". In: *Studies in Logic and the Foundations of Mathematics*. Vol. 120. Elsevier, 1986, pp. 165–196.

[77] P. Pudlák. "On the lengths of proofs of consistency". In: *Collegium Logicum*. Springer, 1996, pp. 65–86.

[78] W. Rautenberg. *A Concise Introduction to Mathematical Logic*. Universitext (Berlin. Print). Springer, 2006. ISBN: 9780387302942.

[79] Rineke. "Efficient Metamathematics". PhD thesis. Universiteit van Amsterdam, 1993.

[80] R. Ritchie and P. Young. "Strong representability of partial functions in arithmetic theories". In: *Information Sciences* 1.2 (1969), pp. 189–204.

[81] H. Rose. *Subrecursion: Functions and Hierarchies*. Oxford logic guides. Clarendon Press, Oxford, 1984. ISBN: 9780198531890.

[82] S. Salehi and P. Seraji. "Gödel-Rosser's Incompleteness Theorem, generalized and optimized for definable theories". In: *Journal of Logic and Computation* 27.5 (2016-07), pp. 1391–1397. DOI: 10.1093/logcom/exw025.

[83] P. G. Santos. *Diagonalization in Formal Mathematics*. Springer Fachmedien Wiesbaden, 2020.

[84] P. G. Santos. *Diagonalization in Formal Mathematics*. BestMasters. Springer Fachmedien Wiesbaden, 2020. ISBN: 9783658291112.

[85] P. G. Santos and R. Kahle. "k-Provability in PA". In: *Logica Universalis* 15.4 (2021), pp. 477–516.

[86] P. G. Santos and R. Kahle. "Variants of Kreisel's conjecture on a new notion of provability". In: *Bulletin of Symbolic Logic* 27.4 (2021), pp. 337–350.

[87] M. Schirn and K. G. Niebergall. "Finitism = PRA? On a thesis of W. W. Tait". In: *Reports on Mathematical Logic* (2005-01).

[88] W. Sieg. "Herbrand analyses". In: *Archive for Mathematical Logic* 30 (1991), pp. 409–441.

[89] W. Sieg. *Hilbert's Programs and Beyond*. Oxford University Press, 2013.

[90] P. Smith. *An Introduction to Gödel's Theorems*. 2nd. Cambridge Introductions to Philosophy. Cambridge University Press, 2013.

[91] C. Smorynski. "The Incompleteness Theorems". In: *Handbook of Mathematical Logic*. Eighth. North-Holland, 1993, pp. 821–865. ISBN: 072042285X.

[92] C. Smoryński. *Self-reference and Modal Logic*. Springer-Verlag, 2012.

[93] R. Smullyan. *Diagonalization and Self-reference*. Logic Guides Series. Clarendon Press, 1994. ISBN: 9780198534501.

[94] R. I. Soare. *Recursively Enumerable Sets and Degrees: A Study of Computable Functions and Computably Generated Sets*. Perspectives in mathematical logic. U.S. Government Printing Office, 1987. ISBN: 9780387152998.

[95] G. Sommaruga-Rosolemos. "Fixed point constructions in Various Theories of Mathematical Logic". In: (1991).

[96] T. Strahm. "Theories with self-application and computational complexity". In: *Information and Computation* 185.2 (2003), pp. 263–297.

[97] R. Verbrugge and A. Visser. "A small reflection principle for bounded arithmetic". In: *Journal of Symbolic Logic* 59.3 (1994), pp. 785–812.

[98] A. Visser. "Another look at the second incompleteness theorem". In: *Review of Symbolic Logic* 13.2 (2020), pp. 269–295.

[99] A. Visser. "Friedman-reflexivity: interpreters as consistoids". In: *arXiv:2111.14413* (2021).

[100] A. Visser. "From Tarski to Gödel how to derive the second incompleteness theorem from the undefinability of truth without self-reference". In: *Journal of Logic and Computation* 29.5 (2019-03), pp. 595–604.

[101] A. Visser. "Peano Corto and Peano Basso: A study of local induction in the context of weak theories". In: *Mathematical Logic Quarterly* 60.1-2 (2014), pp. 92–117.

[102] A. Visser. "The absorption law or: how to Kreisel a Hilbert-Bernays-Löb". In: *Archive for Mathematical Logic* 60 (2021-12), pp. 441–468.

[103] A. Woods. "Some problems in logic and number theory and their connections". In: *Ph. D. Thesis*, *Univ. of Manchester* (1981).

[104] N. S. Yanofsky. "A Universal Approach to Self-Referential Paradoxes, Incompleteness and Fixed Points". In: *Bulletin of Symbolic Logic* 9.3 (2003), pp. 362–386.

[105] R. Zach. "Hilbert's Program". In: *The Stanford Encyclopedia of Philosophy*. Ed. by E. N. Zalta. Metaphysics Research Lab, Stanford University, 2019.

2023ON NOTIONS OF PROVABILITY PGS