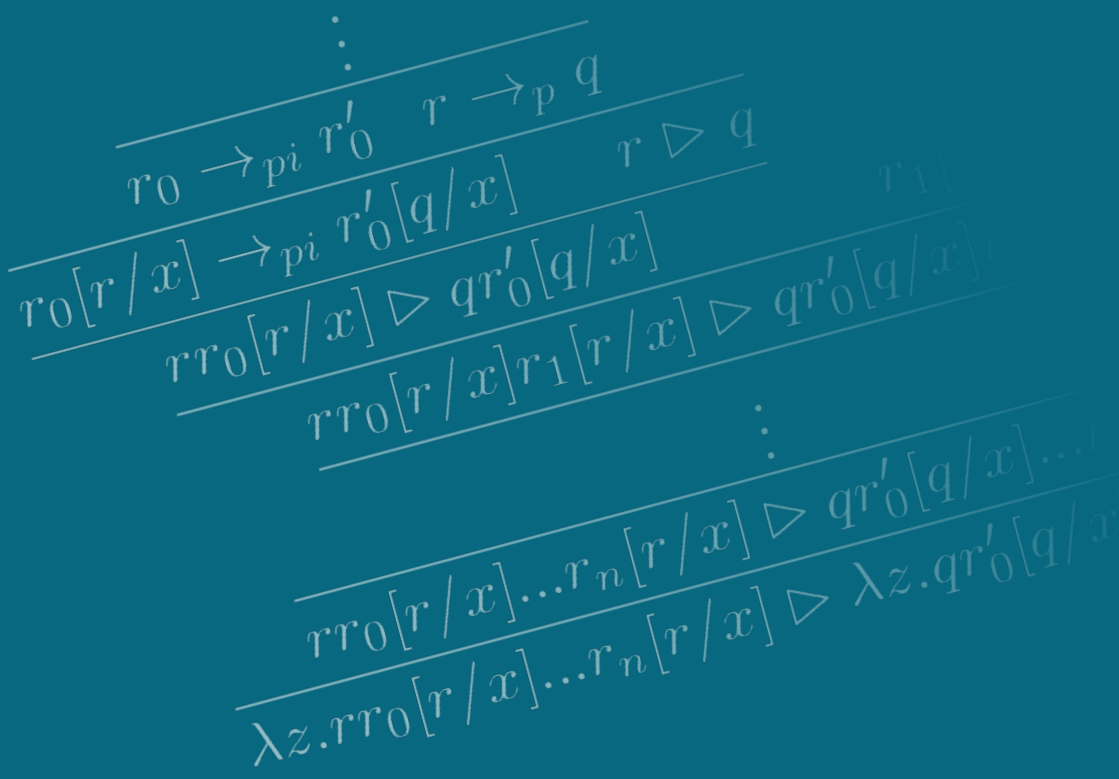
Duccio Pianigiani

Lectures in Proof Theory and Complexity



FIRENZE UNIVERSITY PRESS

USiena PRESS

Ricerca
Didattica
all'Università di Siena

UNIverSI

UNIverSI
Ricerca e Didattica all'Università di Siena
ISSN 3035-5915 (PRINT) | ISSN 3035-5931 (ONLINE)
– 4 –

UNIverSI. Ricerca e Didattica all'Università di Siena

Duccio Pianigiani

Lectures in Proof Theory and Complexity

FIRENZE UNIVERSITY PRESS | USIENA PRESS
2025

Lectures in Proof Theory and Complexity / Duccio Pianigiani. – Firenze : Firenze University Press; Siena : USiena Press, 2025.
(UNIverSI. Ricerca e Didattica all'Università di Siena ; 4)

https://books.fupress.com/isbn/9791221507782

ISSN 3035-5915 (print)
ISSN 3035-5931 (online)
ISBN 979-12-215-0778-2 (PDF)
ISBN 979-12-215-0779-9 (XML)
DOI 10.36253/979-12-215-0778-2

Graphic design: Alberto Pizarro Fernández, Lettera Meccanica SRLs
Front cover image: the derivation on the cover comes from the proof of representability of all partial computable functions in lambda calculus without types, and the construction is taken from page 71 of the volume.

Peer Review Policy
Peer-review is the cornerstone of the scientific evaluation of a book. All FUP - USiena PRESS's publications undergo a peer-review process by external experts under the responsibility of the Editorial Board and the Scientific Boards of each series (DOI 10.36253/fup_best_practice.3).

Referee List
In order to strengthen the network of researchers supporting FUP - USiena PRESS's evaluation process, and to recognise the valuable contribution of referees, a Referee List is published and constantly updated on FUP - USiena PRESS's website (DOI 10.36253/fup_referee_list).

Best Practice in Scholarly Publishing (DOI 10.36253/fup_best_practice)

The online digital edition is published in Open Access on www.fupress.com.

Published by Firenze University Press and USiena PRESS

Powered by Firenze University Press
Università degli Studi di Firenze
via Cittadella, 7, 50144 Firenze, Italy
www.fupress.com

This book is printed on acid-free paper
Printed in Italy

Table of contents

FUP Best Practice in Scholarly Publishing (DOI 10.36253/fup_best_practice)
Duccio Pianigiani, *Lectures in Proof Theory and Complexity*, © 2025 Author(s), CC BY-SA 4.0, published by Firenze University Press and USiena PRESS, ISBN 979-12-215-0778-2, DOI 10.36253/979-12-215-0778-2

Introduction

Duccio Pianigiani

> Proof theory begins where recursion theory ends.
>
> Georg Kreisel

These lecture notes constitute an excerpt of the contents of the course 'Formal Systems' taught by me for the master's degree course 'Applied Mathematics' and partly for the 'Language and Mind' master's degree course at the University of Siena, so it was created with an educational purpose, albeit for second-level courses. They are mainly oriented towards applications of *Proof Theory* (one of the macro-areas into which Mathematical Logic is divided) to *Computability Theory* and *Computational Complexity Theory*, albeit with inevitable entanglements with *Model Theory* and with *Category Theory*. Preliminary, I focus on some classical results concerning formal arithmetic, dating back to the classical era of the 1930s and 1940s, then in subsequent chapters inviting a comparison with more recent results. In this regard, I strongly emphasise the acceleration imparted also to logical study by the development of computer science: in addition to decidability in principle, indeed, computer science has forced attention to be paid to the classification of mathematical problems according to their degree of difficulty and the computational means that can be used in practice. I share the view that sees a continuity with Hilbertian foundational research, i.e. his finitism, or more broadly the various trends of the constructive mathematics (i.e. which requires a more explicit characterisation of proofs than classical mathematics) and these most recent developments which led to a narrower form of finitism, although only a few so-called 'ultraintuitionists' such as van Dantzig, Mannoury or Yessenin-Volpin had previously gone indeed so far as to consider exponentiation problematic before the contributions, between the 70s and the 80s, notably by Rohit Parikh and Edward Nelson. These led to the creation of the research area of *Bounded Arithmetic* and to the discovery of important connections to *Computational Complexity*. Central has become the (unsolved) question of whether $\mathsf{P} = \mathsf{NP}$, already formulated *in nuce* in the 1950s independently by Kurt Gödel and John Nash. Lastly, I outline some (in my opinion) relevant developments in the approach to classical problems such as incompleteness, that have provided a bridge to other areas of science beyond *Mathematical Logic*, as the emergence of the phenomenon of incompleteness in specifically *mathematical* (and not metamathematical) contexts, and in the field of information theory and algorithmic randomness. Far from thinking that this would be an exhaustive introduction, I also devote space to the deep connections between constructive (intuitionistic) logic, type theory (i.e. programming) and that great unifying theory of mathematics constituted by *Category Theory*. For reasons of space, I have not been able to go into all the topics in depth, but each chapter contains extensive bibliographical references and suggestions for independent study. The text could not be entirely self-contained and presupposes a certain familiarity with the first-order logic. For all preliminary notions I invite the reader to consult the very comprehensive, collective, free online handbook https://openlogicproject.org/

This publication was made possible thanks to the specific contribution of the University of Siena for its support of Open Access.

FUP Best Practice in Scholarly Publishing (DOI 10.36253/fup_best_practice)

Part I Formal Systems

1. From decidability to feasibility

1.1. Prerequisites: the language of first-order logic

We can introduce first-order languages through the concept of signature. Once the logical symbols $\neg = not$, $\wedge = and$, $\vee = or$, $\rightarrow = if...then$, $\exists = for\ some$, $\forall = for\ all$ have been established. The signature is essentially the list of non-logical symbols of the language; more formally, it is a quadruple $\langle \mathcal{F}, \mathcal{P}, \mathcal{C}, ar \rangle$, where $\mathcal{F}$ is the set of functional symbols of the language, $\mathcal{P}$ is the set of relational symbols, $\mathcal{C}$ is the set of constant symbols, and ar is a function that assigns each functional or predicative symbol its arity. For example, in the theory of ordered rings $\mathcal{F} = \{+, -, \cdot\}$, $\mathcal{P} = \{\leq\}$ and $\mathcal{C} = \{0, 1\}$; in Peano's first-order arithmetic theory, $\mathcal{F} = \{S, +, \cdot\}$, $\mathcal{R} = \{\leq\}$ and $\mathcal{C} = \{0\}$; in Zermelo-Fraenkel first order set theory, $\mathcal{R} = \{\in\}$ and no constant or function symbols occurs. With regard to a particular relation, namely identity (or equality) $=$, there are two options:

1. (More common) Include this relation among the logical symbols and always interpret $t = s$ as "t and s denote the same object", considering the identity axioms as logical rules (*theories with identity*).

2. Consider $=$ as a binary relation of the signature, interpretable as a set of pairs of elements of the domain, adding, however, axioms that specify that this relation must be an equivalence and a congruence (*theories without identity*).

Terms are builded from function symbols, constant symbols and variables. Thus, any constant c and variable x are terms, and if $t_0, ..., t_n$ are terms and f is a symbol of -arity n, then $f(t_0, ..., t_n)$ is a term. *Atomic formulas* are defined to be of the form $P(t_0, ..., t_n)$ where P is a k -ary relation symbol of the signature. In particular, for t, s terms, $t = s$ is an atomic formula. More complex formulas are inductively defined from atomic formulas and logical constants; hence any atomic formula is a formula, and if α and β are formulas, then so are $(\neg\alpha), (\alpha \wedge \beta), (\alpha \vee \beta), (\alpha \rightarrow \beta), \forall x\alpha$ and $\exists x\alpha$.

We therefore introduce Tarski's semantics for first-order logic in a somewhat pedantic manner, starting from the notion of *realisation*. A *realisation* is a pair $\mathcal{U} = \langle A, \rho_U \rangle$, where:

1. A is a non-empty domain of objects (the universe of discourse).

2. ρ_U is function that assigns a meaning to every *constant* of the language, whether individual, functional or predicative, while assigning to variables only a domain of variation, i.e. A.

Logics that admit empty domains and terms that do not denote any object are called 'free logics'. They also admit terms t that denote objects outside the (non-empty) domain for which, clearly, the principle $\forall x\phi(x) \rightarrow \phi(t)$ is not acceptable in the aforementioned domain. More precisely, a classical realisation satisfies these conditions:

1. For each individual constant c_i of the language,$\rho_U(c_i)$ is an object of A, which we will denote by c_i^U.
2. For each functional symbol f_i, $\rho_U(f_i)$ is *an operation* of the same -arity as f_i on A, which we also denote by f_i^U.

Duccio Pianigiani, University of Siena, Italy, duccio.pianigiani@unisi.it, 0000-0001-9441-7226

Referee List (DOI 10.36253/fup_referee_list)
FUP Best Practice in Scholarly Publishing (DOI 10.36253/fup_best_practice)

Book References DOI 10.36253/979-12-215-0778-2.references

3. For each predicative symbol P_i, $\rho_U(P_i)$ is a *relation* of the same -arity as P_i on A, which we denote by P_i^U. In particular, if $\rho_U(=)$ is precisely the set of pairs $< a, a >$ of elements of A, we say that it is a *normal* realisation.
4. For each variable x_i, $\rho_U(x_i) = A$.

Other presentations can be found in the scientific literature. For example, we can consider the structure (of the same signature as the language) obtained by explicitly writing, instead of ρ, the *interpretations* of the relational and functional symbols and constants according to ρ. If $\rho(R_i) = R_i^U$, $\rho(f_i) = f_i^U$, $\rho(c_i) = c_i^U$. We will denote the realisation as follows:

$$\mathcal{U} = \langle A, R_0^U, ..., R_n^U, f_0^U, ..., f_m^U, c_0^U, ..., c_k^U \rangle$$

Hence, many presentations start from the concept of an *algebraic-relational structure* of a certain signature and a language for this structure, i.e. a language of the same type of similarity. Notice that so far, having established a realisation, we are only able to interpret closed terms and statements in the strict sense. *Open* formulas are defined as those containing free variables. They are either *satisfied* or not, depending on how the free variables are interpreted in a given domain. Following the most orthodox approach, we give a simultaneous interpretation σ of all variables: by *evaluation in A* we mean a function $\sigma : Var \longrightarrow A$, i.e. an infinite sequence $\sigma = \langle \sigma(x_0), \sigma(x_1), \sigma(x_2), ... \rangle$ of objects of A. Since we do not know in advance which variables might appear in a formula, the simplest solution is in fact to require that an assignment assigns values to *all* variables. By *x_i-variant of σ* we mean a sequence:

$$\sigma(a/x_i) = \langle \sigma(x_0), \sigma(x_1), \sigma(x_2), ...\sigma(x_{i-1}), a, \sigma(x_{i+1})... \rangle$$

that is, a function that differs from σ only in the value it assigns to x_i. The evaluation of terms, induced by σ in a realisation $\mathcal{U}$ is the following:

1. $\rho_U^\sigma(c_i) = \rho_U(c_i) = c_i^U$, for every constant$c_i$.
2. $\rho_U^\sigma(x_i) = \sigma(x_i)$, for every variable$x_i$.
3. $\rho_U^\sigma(f(t_0, ..., t_n)) = \rho_U(f)(\rho_U^\sigma(t_0), ..., \rho_U^\sigma(t_n))$.

To formally define the notion of *truth*, we must first establish the meaning of the variables. At this point, however, we are able to evaluate the formulas; we arrive at the notion of truth, following Tarski, by passing through the relation of *satisfaction*. Some key notions of semantics will be the following:

1. *Satisfiability*: ϕ is satisfiable in $\mathcal{U}$ if there exists some assignment σ to the variables that satisfies ϕ in $\mathcal{U}$ (notation $\mathcal{U} \models_\sigma \phi$)
2. *Truth*: ϕ is true in $\mathcal{U}$ if and only if it is satisfied by every σ (notation $\mathcal{U} \models \phi$).
3. *Validity*: ϕ is universally valid if and only if it is true in every realisation (notation $\models \phi$).

The *statements*, which in this context are properly thoseformulas that are free of variables, will have the characteristic of being *satisfied by all assignments*, or *not satisfied by any assignment* of value to the variables, with respect toa certain *realisation*. A statement will be said to be *true in* ρ_U, or *false in* ρ_U.

We define the satisfaction relation σ satisfies ϕ in $\mathcal{U}$(notation $\rho_U^\sigma \models \phi$, or more commonly $\mathcal{U} \models_\sigma \phi$), inductively, through the following clauses:

1. $\mathcal{U} \models_\sigma P_i(t_0, ..., t_n)$ if and only if $\langle \rho_U^\sigma(t_0), ..., \rho_U^\sigma(t_n) \rangle \in \rho_U^\sigma(P_i)$
2. $\mathcal{U} \models_\sigma \neg\psi$ if and only if $\mathcal{U} \not\models_\sigma \psi$
3. $\mathcal{U} \models_\sigma \phi \wedge \psi$ if and only if $\mathcal{U} \models_\sigma \phi$ and $\mathcal{U} \models_\sigma \psi$ (similarly for $\vee$)
4. $\mathcal{U} \models_\sigma \phi \rightarrow \psi$ if and only if $\mathcal{U} \models_\sigma \neg\phi$ or $\mathcal{U} \models_\sigma \psi$
5. $\mathcal{U} \models_\sigma \exists x_j \phi$ if and only if $\mathcal{U} \models_{\sigma(a/x_j)} \phi$ for some $a \in \mathcal{U}$.

6. $\mathcal{U} \models_\sigma \forall x_j \phi$ if and only if $\mathcal{U} \models_{\sigma(a/x_j)} \phi$ for every $a \in \mathcal{U}$.

By the *Coincidence lemma*, if σ, τ are two assignments that coincide on the free variables of t. Then $\rho^\sigma_\mathcal{U}(t) = \rho^\tau_\mathcal{U}(t)$. Furthermore, if σ, τ are two assignments that coincide on the free variables of a formula ϕ, then $\mathcal{U} \models_\sigma \phi$ if and only if $\mathcal{U} \models_\tau \phi$. In particular, if ϕ is a statement, i.e. it has no free variables, then if $\mathcal{U} \models_\sigma \phi$ *for some* σ, then $\mathcal{U} \models_\sigma \phi$ *for every* σ. Note that two assignments σ, τ clearly coincide in assigning values to all variables in the set X, if X *è empty*. A statement is therefore *true in all evaluations* associated with a given realisation, or it is not true in *any* of them ("true in ρ_U", or "false in ρ_U"). A *formula* ϕ can be satisfied by *all* the evaluations associated with a given realisation (we will write $\mathcal{U} \models \phi$ and say that it is true in that realisation), by *none* of them, or only by *some* of them. A formula ϕ is universally (or logically) valid if it is true in all realisations. As regards deductive systems, in this volume we will introduce those most commonly used in *Proof Theory*, i.e. those *à la* Gentzen (many rules and few axioms). However, it is useful to recall a typical axiomatic presentation (in Frege-Hilbert style), consisting of many logical axioms and few deduction rules. In axiomatic derivations, a sequence of sentences counts as a correct derivation from a set of hypothesis, if every sentence in it either is an axiom, or is an element of the given set of hypothesis, or come from previous formulas of the sequence by a rule of the calculus. Let us give a typical example of axiomatisation:

1. $p_0 \to (p_1 \to p_0)$
2. $(p_0 \to (p_1 \to p_2)) \to ((p_0 \to p_1) \to (p_0 \to p_2))$
3. $(p_0 \wedge p_1) \to p_0$
4. $(p_0 \wedge p_1) \to p_1$
5. $p_0 \to (p_1 \to (p_0 \wedge p_1))$
6. $p_0 \to (p_0 \vee p_1)$
7. $p_1 \to (p_0 \vee p_1)$
8. $(p_0 \to p_2) \to ((p_1 \to p_2) \to ((p_0 \vee p_1) \to p_2)$
9. $(p_0 \to p_1) \to ((p_0 \to \neg p_1) \to \neg p_0)$
10. $\neg\neg p_0 \to p_0$

Propositional rules:

$$\frac{\phi \quad \phi \to \psi}{\psi} \qquad \frac{\phi}{\phi\sigma}$$

namely Modus Ponens and uniform substitution, where $\sigma$is a substitution: 1) $p\sigma = \sigma(p)$; 2) $\bot\sigma = \bot$; 3) $(\phi \wedge \psi)\sigma = \phi\sigma \wedge \psi\sigma$ (analogously for other connectives). Alternatively, due to the fact that every formula derivable from 1)-10) can be derived by applying the substitution rule only to the axioms, axioms are very often given as *axiom schemes*, i.e. replacing propositional variables with metavariables to represent infinitely many individual axioms, incorporating, so to speak, the substitution rule, which will therefore be omitted. By removing axiom 10, we obtain the *Intuitionistic Propositional Calculus* IPC.

First-order extension. In this case, we add some additional axioms and rules to the axiomatic basis of propositional calculus:

$$\forall x \phi(x) \to \phi(t) \qquad \phi(t) \to \exists x \phi(x)$$

and rules:

$$\frac{\phi \to \psi(z)}{\phi \to \forall x \psi}$$

universal generalization where where z does not appear among the free variables of ϕ.

$$\frac{\phi(z) \to \psi}{\exists x \phi \to \psi}$$

existential generalization where z does not appear among the free variables of ψ. In fact, there are many (obviously equivalent) ways to axiomatise classical first-order logic. Often, only axioms

for implication and negation are given, plus those for the universal quantifier, since connectives, like quantifiers, are interdefinable. This is not the case for intuitionistic logic, where the above-mentioned logical operators are all independent and that is why we have reported an axiomatisation for the full language. We use the symbology "$\Gamma \vdash \phi$" to indicate derivability from hypothesis in the set Γ (perhaps empty) in a given deductive system, of the formula ϕ. For systems that satisfy axioms 1. and 2., the deduction theorem holds:

$$\Gamma \cup \{\gamma\} \vdash \psi \Rightarrow \Gamma \vdash \gamma \to \psi$$

In this volume, we will also discuss classical modal logic systems, i.e., extensions of this logic with operators $\Box$ = *necessarily* and $\Diamond$ = *possibly*. The semantics for this logic are those of Kripke models. The prerequisites will be provided at the time.

1.2. The concept of formal system

The discovery of antinomies in *naïve* set theory led many matematicians in the early 20th century to abandon Cantor's idea of the actual infinite and set-theory. On the contrary, the eminent logician and mathematician David Hilbert, one of the most remarkable figures in mathematics of the entire century, believed that "Cantorian paradise" (i.e. set theory) had to be defended, and considered it his personal duty to defend mathematics against skeptics, clarifying and justifying the mathematician's use of the infinite. Hilbert's reaction to the antinomies concerning the existence of certain sets, was to claim that a set of mathematical objects (e.g. set of all real numbers) exists only if we can prove that the corresponding axiom system for them is free of contradictions, and a proof of this has to be done combinatorially or constructively. Infinitary mathematics must be based on safe *finitary* mathematics, but what is the finitary, in some sense (essentially kantian) intuitive mathematics? Skolem's *Primitive Recursive Arithmetic* PRA, for example, was a candidate to embody the finist methods. Actually there are many different versions of this theory. Among the most common is this one: the first-order language of PRA contains symbols for any primitive recursive function and the relative recursive defining axioms. In this extended language, the theory allows for induction over open formulas. We will see also that we may think of it as the theory $I\Sigma_1$ in the more restricted language of *Peano Arithmetic*. In Tait (1968), the author argues that finitist arithmetic coincides with PRA, even though this identification is largely conventional and disputable. However, after the discovery of Gödel's incompleteness results, wanting to keep the core of this programme, as we will see, it was necessary to expand Hilbert's finististic standpoint. Many issues we will deal with have arisen during the controversy around the *legitimacy of abstract objects* as the infinite sets, which led on the other hand to the clarification of the concept of *finite*, *effective*, *algorithmic* procedures. Hilbert never formulated his program exactly, however, in a nutshell, we can say that it consisted of the following steps:

1. Formalize all infinitistic mathematics in an higher-order formal system of mathematical logic.

2. Show the consistency of this system with finitist tools, which have purely combinatorial character. We will see that this implies the elimination, in principle, of the ideal objects (e.g. uncountable sets) from the proof of concrete, "real" statements (certain universal sentences $\forall x \psi(x)$, where $\psi(x)$ is a decidable predicate): the justification for the infinitary mathematics lies in its conservativity over real, finitisitically meaningful mathematics.

Gödel's second incompleteness theorem (1931) showed that this program was unattainable: no consistent, recursively axiomatized system that contains sufficient mathematics proves its own consistency. The first incompleteness theorem, on the other hand, showed that for a wide class of formal theories, there are true sentence that can be expressed in the language of these theories, but that are neither provable, nor refutable in them. Lastly, Hilbert and Ackermann in 1928 also formulated the *Entscheidungsproblem* (thet means "decision problem"), i.e. the problem of devising an algorithm for deciding the validity of statements of first-order logic, but in 1936, Alonzo Church and Alan Turing independently showed that a general solution to this problem is impossible. To show this, they gave a deep formal treatment of the intuitive notion of "effectively computable

method" which gave rise to the modern abstract models of computation, which we will account for in the book. This section is devoted to recalling some fundamental metatheoretical concepts related to formalised mathematical theories.

Formal mathematical theories. A formal mathematical (first order) theory is given by the following ingredients:

1. A first order language L.
2. A set of logical axioms (formulas of L).
3. A set of inference rules.
4. A set T of formulas containg the logical axioms and closed under the inference rules, i.e. if $\Gamma \subseteq \mathsf{T}$ and ϕ is a formula of the language derivable from Γ by means of the logical axioms and the inference rules (write $\Gamma \vdash \phi$), then $\phi \in \mathsf{T}$.
5. If a set Γ exists such that $\mathsf{T} = \{\phi | \Gamma \vdash \phi\}$, then Γ is a set of *non-logical axioms* for T and T is the set of *theorems*. As usual, we will write $\mathsf{T} \vdash \phi$ to mean that ϕ is derived from the axioms of T by means of logical axioms and inference rules of the theory.

In this book we will actually favour formal systems in the style of Gentzen, i.e. many rules and few axioms (see on p. 174). For an axiomatic theory T, let $Thm(\mathsf{T}) = \{\phi | \mathsf{T} \vdash \phi\}$ the set of theorems and $Ref(\mathsf{T}) = \{\phi | \mathsf{T} \vdash \neg\phi\}$ the set of refutable sentences. We say that the theory T is *consistent* iff $Thm(\mathsf{T}) \cap Ref(\mathsf{T}) = \emptyset$. We say that T is *syntactically complete* if for each sentence ϕ of the language of T, either $\phi \in Thm(\mathsf{T})$, or $\phi \in Ref(\mathsf{T})$. We say that T is *decidable*, if the set $Thm(\mathsf{T})$ is algorithmically decidable. Lastly, we say that S is an *extension* of T iff the language of T is a subset of the language of S and all theorems of T are also theorems of S.

Among the various meanings in logic of the word "completeness" it is worth remembering the following:

1. *Syntactic completeness.* An axiomatic theory is complete in this sense, if for all sentence of its language ϕ, either ϕ, or $\neg\phi$ is a theorem of it.
2. *Completeness with respect to the truth.* If for instance $\mathsf{T} = \mathsf{PA}$, then, if ϕ is true in the standard (intendend) model $\mathbb{N}$ of natural numbers, then it is a theorem of T.
3. *Semantic completeness.* A sentence ϕ is a theorem of a certain theory, if and only if it is true in *all* models $\mathfrak{M}$ of that theory. First order theories are complete in this third sense.

Gödel's first theorem states that for many theories, even weaker than PA, properties 1. and 2. do not apply.

Decidability. A theory is called *decidable* if the set of its theorems, i.e. the sentences derivable in it, is *decidable*, that is, there is a mechanical procedure which enables one to decide whether an arbitrary given sentence of the language of the theory is a theorem or not. A theory is *axiomatizable* if there is a decidable set of axioms for it. If a theory is axiomatizable and *syntactically complete*, then it is *decidable*, although the converse does not always hold: there are *incomplete* theories which are *decidable* (for instance, the theory ACF of algebraically closed fields is decidable, but not complete: it becomes complete when we fix the characteristic and we get an ACF_p). All theories which contain Robinson arithmetic Q (see below) or that interpret it are both incomplete and undecidable. Contrasting with the undecidability of Q we have Tarski's decidability result of the theory of ordered *Real Closed Fields*.

Axiomatizability. The theory Q is *finitely axiomatized*. However, the set of axioms of our formal systems will be often *infinite*, but we will be able to determine by an effective (algorithmic) procedure whether a certain formula is an axiom or not, namely, the theory will be *recursively axiomatizable*. Or, at least, we will be able to enumerate all and only the axioms: we say in this case that the theory has a *recursively enumerable set of axioms*. Craig's theorem (see p. 69) clarifies the connection among these alternatives.

Theorem 1. *If a theory T is recursively axiomatizable (often called simply "axiomatizable") and syntactically complete, then is decidable.*

Proof. Recall that if a theory is axiomatizable, then the set of its theorems is recursively (or "computably") enumerable (see e.g. Enderton (2001) 156-57). Suppose now T is complete and has

a computable set of axioms. If T is *inconsistent*, it is clearly computable (Algorithm: "just say yes"). If T is *consistent*, to decide whether or not a sentence ϕ is in T, simultaneously search for a derivation of ϕ from T and a derivation of $\neg\phi$. Since T is complete, you are bound to find one or the other; and since T is consistent, if you find a derivation of $\neg\phi$, there is no derivation of ϕ.
QED

The complete theory of a model.

As we have already mentioned in the prerequisites, a structure (or model) $\mathcal{U}$ for a first-order language consists of a non-empty domain equiped with distinguished relations, functions and designed individuals on it, that interpret respectively the relational, functional and constant symbols of the language. We denote $Th(\mathcal{U})$ the so-called *complete theory* of the model $\mathcal{U}$, namely the set of true sentences defined on it. Clearly, if L is the language of the model, then for any $\phi \in L$, either $\phi \in Th(\mathcal{U})$, or $\neg\phi \in Th(\mathcal{U})$. It is typical in this regard to contrast these two historical results:

1. First order theory of *Real Closed Ordered Fields* RCOF is recursively (or effectively) axiomatizable and complete in the first sense. For a celebrated theorem due to Tarski:
$$\mathsf{RCOF} = Th(\langle \mathbb{R}, +, -, \cdot, 0, 1, < \rangle)$$
2. The first Gödel theorem, states at the opposite that no consistent, effectively axiomatizable and sufficiently powerful theory of arithmetic is complete in the first two senses. Hence $Th(\langle \mathbb{N}, +, \cdot, S, 0 \rangle)$, although complete, *is not effectively axiomatizable.*

The First Incompleteness Theorem can be actually formulated as follows.

Theorem 2. (Gödel 1931) *There is no decidable set of axioms for the complete theory of the standard model.*

Proof. If a theory is *consistent* and *represents (or binumerates) all computable functions*[1], then is *undecidable*. If a theory is *axiomatizable* and *complete* then is *decidable*. Then if a theory is *consistent*, *axiomatizable* and *represents all recursive functions*, it *cannot be complete*. Now consider that the complete theory of the standard model is complete and represents all computable functions: hence cannot be axiomatizable.
QED

For "sufficiently powerful theories", we will mean extensions of the theory of formalized arithmetic theory Q (Robinson Arithmetic), a finitely axiomatized theory in the language $L = \{\overline{0}, S, +, \cdot\}$ whose axioms are[2]:

1. $Sx \neq \overline{0}$
2. $Sx = Sy \to x = y$
3. $x + \overline{0} = x$
4. $x + Sy = S(x + y)$
5. $x \cdot \overline{0} = \overline{0}$
6. $x \cdot Sy = x \cdot y + x$
7. $x \neq \overline{0} \to \exists y(Sy = x)$.

In we replace axiom 7. with the axiom scheme of induction:

$$\phi(\overline{0}) \wedge \forall x(\phi(x) \to \phi(Sx)) \to \forall x \phi(x)$$

we get the theory (not finitely axiomatizable) called "first-order Peano arithmetic", denoted by PA.

Strong fragments of Peano Arithmetic. Recall this classification of formulas (here of the language of PA), called *Arithmetical Hierarchy.* A formula is called a *bounded formula* if it contains only bounded quantifiers $\forall x \leq t$ or $\exists x \leq t$. The set of bounded formulas is denoted Δ_0. For $n \geq 0$ the classes Σ_n and Π_n of first-order formulas are inductively defined by:

[1] For instance, Odifreddi (1989-1999), 40-1 uses the terminology "represents, weakly represents" , while in Hájek and Pudlák (1993), 155 is used the terminology "binumeratess, numerates".

[2] Hoping not to cause confusion, and to avoid further burdening the notation, we will use in all the book the symbol "=" for equality, for conversion and to mean equal by definition. We hope the reader will grasp the different meanings intuitively.

1. $\Sigma_0 = \Pi_0 = \Delta_0$,
2. Σ_{n+1} is the set of formulas obtained by prepending an arbitrary block of existential quantifiers and bounded universal quantifiers to Π_n-formulas.
3. Π_{n+1} is the set of formulas obtained by prepending an arbitrary block of universal quantifiers and bounded existential quantifiers to Σ_n-formulas.

We want to point out that the correct notation would be for example Π_n^0 and Σ_n^0 where the superscript "0" means that we are talking about first-order formulas (while the superscript "1" would mean that we are talking about second-order formulas). If we do not write anything it will be implicit that we are talking about first-order formulas. Recall that in classical logic it is possible to rewrite formulas to bring all quantifiers to the front (*prenex normal form*), hence each formula is in one of the above form. A formula is Δ_n in the considered theory (e.g. PA) if it is provably equivalent in the theory to both a Σ_n formula and a Π_n formula. Note that if a formula is in Π_n, then is also in Σ_{n+1}. A corresponding hierarchy of subtheories of PA is defined by restricting the induction axiom to a fixed level of the arithmetic hierarchy. We denote $I\Sigma_n$ Peano Arithmetic with induction restricted to the class Σ_n.

The so called *strong fragments* of Peano Arithmetic, denoted by $I\Sigma_n$, or $I\Pi_n$, for $n > 1$, are obtained by restricting the induction schema to the classes Σ_n, resp. Π_n. These fragments are *finitely axiomatizable* (see Hájek and Pudlák (1993), 77-81) because there exists a *universal formula* for them (this is not the case for the full PA). A universal formula for Σ_n is a Σ_n-formula $\Psi(x, y)$ such that for each Σ_n- formula $\theta(x)$ there exists a number e such that:

$$I\Sigma_1 \vdash \theta(x) \leftrightarrow \Psi(x, \overline{e})$$

where $\overline{e} = \overbrace{SSS...S\overline{0}}^{e-times}$. Hence we can collapse infinitely many instances of induction in a *unique* axiom:

$$\forall y(\Psi(\overline{0}, y) \wedge \forall x(\Psi(x, y) \rightarrow \Psi(x + 1, y)) \rightarrow \forall x\Psi(x, y))$$

Why the theory Q? This theory was introduced by Raphael Robinson in 1950 and in fact, it is not adequate to the formalization of arithmetic, due to the absence of the induction principle. However we will see what is the tecnical reason of interest: it *binumerates* the recursive functions (see 1). We will see that Q, although very weak, still is *essentially undecidable* (all consistent extensions of it are also undecidable).

It should be remarked that we can go much weaker in complexity, e.g. by considering a theory named R, introduced in Tarski, Mostowski and Robinson (1953) which is strictly weaker than Q, is still essentially undecidable, binumerates the recursive functions, but is not finitely axiomatizable. In some way Q is a minimal finitely axiomatizable theory in which every recursive function is "binumerable" or ("representable"). The importance of this theory is also due to foundational reasons in the current debate on constructive mathematics, where some go beyond the traditional finitist or intuitionist positions. For them, it represents a sort of paradigm. In this respect, it is worth mentioning the influential Nelson (1986) where the *interpretability in* Q is synonymous with *safety*. The author's radical finitism poses a barrier in the *totality of exponentiation*, a fact he didn't believe in. His strictly finitist program implies the denial of the existence of certain large natural numbers as 2^{65536} and claims that we should try to develop mathematics under the assumption of the denial of the totality of exponentiation. We will see (Parikh's theorem 111) that the totality of exponentiation cannot be proved in very weak theories.

A theory is *essentially incomplete* if all its recursively axiomatizable extensions are incomplete. Gödel (Rosser) theorem in fact says that a certain weak base theory Q is essentially incomplete (i.e also incompletable) in the first two sense. Robinson proved that Q is also *essentially undecidable*, i.e. that any consistent theory that extends (or interprets) Q is undecidable. A proof of incompleteness that uses these properties of recursively enumerable and recursive sets rather than self-reference can be called *structural*.

No axiomatic subtheory of Q obtained by removing any one of its axioms remains essentially undecidable (see Tarski, Mostowski and Robinson (1953)). Let us remove for example the axiom

$S(x) = S(y) \to x = y$. Call Q^- the theory thus obtained and consider the structure:

$$\mathcal{M} = \langle \{0, 1\}, 0, S, +, \cdot \rangle$$

where 0 and $\cdot$ have their usual meaning, but $S(\overline{0}) = S(\overline{1}) = 1$; $0 + 0 = 0$ and

$$x + y = \begin{cases} 0 & if\ x = y = 0 \\ 1 & if\ x = 1\ or\ y = 1 \end{cases} \tag{1}$$

Note that all the other axioms are true in this structure. Let $Th(\mathcal{M})$ the set of sentences true in this structure, observing that it is an extension of Q^- and it is *decidable*: indeed it proves

$$\forall x \phi(x) \leftrightarrow \phi(\overline{0}) \wedge \phi(\overline{1}), \quad \exists x \phi(x) \leftrightarrow \phi(\overline{0}) \vee \phi(\overline{1})$$

The statement that first Incompleteness theorem by Gödel holds for the consistent axiomatizable theory T is actually equivalent to the each of the statements: "T is essentially incomplete" and "T is essentially undecidable". The equivalence depends in particular on the facts that every *incomplete decidable* theory has a consistent, *decidable complete* extension in the same language, and that it is well known that every consistent *axiomatizable* and *complete* theory is *decidable* (if the theory is not consistent, the algorithm tha decides it is the one that says always YES; otherwise, being the theory axiomatizable, we can effectively make a list of its correct derivations: if a sentence does not appear in the list, since the theory is complete, it will appear its negation) .

In more recent years (see, among others, Visser (2009)), attention has also been paid to the "other" Robinson arithmetic R, axiomatized by the following schemes (where the language is $0, S, +, \cdot, \leq$ and the numerals $\overline{n}$ are defined by means of the successor S):

1. $\overline{m} + \overline{n} = \overline{m + n}$
2. $\overline{m} \cdot \overline{n} = \overline{m \cdot n}$
3. $\overline{m} \neq \overline{n}$, if $m \neq n$.
4. $\forall x(x \leq \overline{n} \to x = \overline{0} \vee ... \vee x = \overline{n})$
5. $\forall x(x \leq \overline{n} \vee \overline{n} \leq x)$

If we remove the axiom 5. the remaining theory R_0 is no longer essentially undecidable. However, if $\leq$ is not assumend as primitive and $x \leq y$ is defined as $\exists z(z + x = y)$, then the theory R_0 is still essentially undecidable. If in axiom 4. we put $\leftrightarrow$ in place of the implication, the theory is essentially undecidable, but not minimal essentially undecidable (for a more detailed discussion, see Jones (1983)).

The theory R is *intepretable* (in the sense of the definition on p. 112) in Q, but not viceversa: if Q were interpretable in R, then it would be intepretable in a finite fragment of R, but finite fragments of R have finite models ("locally finitely satisfable") where, on the contrary, Q has only infinite models.

Theorem 3. (Ryll-Nardzewski 1952) *There is no finite set of axioms* S *of the language of* PA*, such that* $\mathsf{S} \vdash \phi$ *if and only if* $\mathsf{PA} \vdash \phi$.

Proof. A direct argument appeals to Gödel's second result: reasoning by contradiction if PA were finitely axiomatizable, and therefore had finite instances of theinduction axiom, then would be equivalent to some $I\Sigma_k^0$, i.e. PA, with the induction restricted to formulas of the form $\exists x_0 \forall x_1 \exists x_2 ... Q x_{k-1} \theta$ for some fixed k. However it is provable that $\mathsf{PA} \vdash Con(I\Sigma_k^0)$, for all k. Hence we would have $I\Sigma_k^0 \vdash Con(I\Sigma_k^0)$, in other words $\mathsf{PA} \vdash Con(\mathsf{PA})$, against Gödel's second theorem. QED

The first order logic has earned only with time the fame of "standard logic". From a famous theorem in Lindström (1969), we now know that countable compactness and downwards Löwenheim-Skolem characterize first-order logic, in the sense that it is the strongest logic to fulfils both, where:

1. *Compactness* (K. Gödel 1930, A. I. Malčev 1937)): "Γ has a model iff each of its finite subset $\Sigma \subseteq \Gamma$ has a model", and

2. *Löwenheim-Skolem theorems* (L. Löwenheim 1915, T. Skolem 1920):

   (a) (Löwenheim-Skolem downwards). Let Σ a set of sentencese of the first order language $\mathcal{L}$ and let its cardinality $|\mathcal{L}| = \kappa$; let $\kappa < \lambda$. If Σ has a model of cardinality λ, then Σ has a model of cardinality $\kappa \leq \alpha < \lambda$.

   (b) (Löwenheim-Skolem upwards). Let Σ a set of sentences of the first order language $\mathcal{L}$ and let its cardinality $|\mathcal{L}| = \kappa$. If Σ has a model of cardinality $\lambda \geq \kappa$ then Σ has also a model of cardinality μ, for all $\mu \geq \lambda$.

It is worth making a comparison with the *second order* Peano arithmetic PA_2 (in the language with the only functional symbol S, the successor, were $Sn = n + 1$, and a constant 0), a finitely axiomatizable theory with axioms:

1. $\forall x \neg (Sx = 0)$

2. $\forall x \forall y (Sx = Sy \rightarrow x = y)$

3. $\forall X ((X(0) \wedge \forall x (X(x) \rightarrow X(Sx)) \rightarrow \forall y X(y)))$

The privileged interpretation consists of the second-order structure:

$$\mathcal{N} = \langle \mathbb{N}, \mathcal{P}(\mathbb{N}), \overline{0}, \overline{S} \rangle$$

From a theorem due to Dedekind (1887), it follows that all *principal* models $\langle U, F, a, g \rangle$ of this theory (i.e. those models in which the second-order variables vary over *all* the power-set of the domain U, namely $F = \mathcal{P}(U)$) are isomorphic: that is to say, with respect to this semantics, PA_2 is *categorical*. It is well known that (because of compactness theorem and Löwenheim-Skolem theorems) categoricity *does not hold* if we formulate Peano arithmetic in the First Order Logic, and that there exist non standard models, besides the standard one. But the categoricity of $\mathsf{PA}_2$is strictly dependent from the choice of the above semantic of *principal* models. A more *general* interpretation of the second-order language is a structure:

$$\mathcal{U} = \langle \{\mathcal{F}_i | i \in \mathbb{N}\}, P_0, ..., P_{i_j}, \{a_i | i \in I\} \rangle$$

where, for all i, $\mathcal{F}_i$ is a family of $i - ary$ relations on $A^i = A \times ... \times A$, where are interpreted the relational variables. If for all i, $\mathcal{F}_i$ coincides with the whole-power set $\mathcal{P}(A^i)$, then we say that the interpretation is *principal*. For *general* interpretations, Leon Henkin (1950), showed the semantic completeness (i.e. the compactness) and Löwenheim-Skolem and thus the non-categoricity. So, what is second order logic with general semantic? The "nothing but" thesis says: Second-order logic with the general semantics is nothing but first-order logic (many-sorted) together with the *comprehension axioms*. Thus a sentence is valid in the general semantics iff it is logically implied (in first-order logic) by the set of comprehension axioms:

$$\exists P \forall u_0 ... \forall u_n (P(u_0, ..., u_n) \leftrightarrow \phi)$$

where ϕ is a second order formula not containing P. These tell us what sets exist. The second order language is therefore treated often as a two-sorted first order language and in this case a model consists of two sets, where the elements of the first domain are interpreted as numbers and the elements of the second domain are interpreted as sets.

1.3. Models of computation: Turing machines

The computational model proposed by Alan Turing has some highly intuitive features. Comparing the various approaches that emerged at the time, Gödel stated the following:

I was completely convinced only by Turing's paper (K. Gödel, 1968).

In Turing (1936) a computation model is introduced, obtained by analyzing the activity of an idealized human agent ("computator") that performs computations with pencil and paper writing symbols on a squared paper, running operations "so elementary, which is not easy to imagine that they can be further broken down". It is assumed (*finiteness conditions*) that the number of symbols, as well as the number of squares observed at any one moment, and the number of "mental states" of computator are *finite*. The operations which the idealized human runs are change the symbol in the observed square and move to a different set of squares; operations depend only on the internal state of computator and the symbol scanned. Turing suggested that any computable function, in an informal way, was "computorable" (*Turing Thesis*). In fact he proved with a rigorous argument (although not in a formal system) that every function "computorable" is computable by a Turing machine (*Turing theorem*). Important matematicians immediately accepted Turing's analysis. Within a very short time, the various models of computation proposed at the time (Turing machines, recursive functions, lambda calculus...) were shown to be equivalent (see Soare (2014) for an historical reconstruction), but in an interview with W. Aspray, in 1985, Kleene declared:

> Turing's definition of computability was intrinsically plausible, whereas with the other two [recursive functions and λ-definability], a person became convinced only after he investigated and found, much by surprise, how much could be done with the definition.

The Turing machine was shortly thereafter taken as a model for many things: in Gödel's opinion, Turing's definition fulfils all defining properties of a formal mathematical system (see Feferman (2006)). According to some exponents of the computationalist theory of mind, starting from McCulloch and Pitts (1943), the Turing machine might even provide a model for the mind. Other, starting from Lucas (1961), on the contrary, claim that Gödel's first theorem refutes the mechanistic thesis that the human mind is, or can be accurately modeled as a Turing machine. However, as regards this discussion, Gödel himself was cautious, and wrote that from his theorem actually simply followed this dichotomy:

> *Either* mathematics is incompletable ... that is to say, the human mind (even within the realm of pure mathematics) infinitely surpasses the powers of any finite machine, *or* else there exist absolutely unsolvable diophantine problems (Gödel (1951), 310).

The Turing model has been used also to study the computational complexity of problems and to define complexity classes. This, we will see, among other things, provides a formal definition to the intuitive notion of "feasible" computation as *polynomial time* computation. We consider first a very simple example of one-tape machines. The machine's tape is potentially *infinite*. This corresponds to an assumption that the *memory* of the machine is (potentially) *infinite*. The cells are all marked with a symbol. In each instant the machine is in a state q_i and after writing something, the head can slide to the right ($= R$) or left ($= L$).

Formally, a deterministic Turing machine, in the simplest version presented here, is a set of instructions based on this alphabet:

1. *Tape symbols* $\{\alpha_0, ..., \alpha_n, *\}$. The symbols which are different from $*$ ("the blank") are called *the alphabet of the machine* (or *input alphabet*).

2. *Internal states* $\{q_0, ..., q_n\}$. We distinguish an *initial state* q_0.

3. *Action symbols* L, R, standing for "left" and "right".

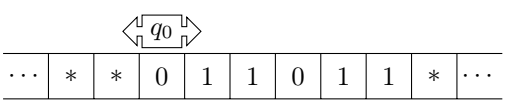


Figure 1. Tape of a Turing machine.

The tape is divided into *cells*, and each cell contains one symbol from the tape alphabet. There is a special blank symbol $*$. At any instant, all but finitely many cells hold $*$. The tape alphabet contains the blank symbol in addition to the alphabet of inputs, but may contain other symbols as well. However, the tape alphabet *cannot be the same as the input alphabet.* The tape alphabet always contains the blank symbol, but the input alphabet cannot contain this symbol. Indeed, if the blank symbol were part of the input alphabet, the Turing machine would never know when the input actually ends. At the beginning the head is positioned on the first symbol on the left different from $*$ and the machine is in the state q_0. Instructions are often given by quadruples. Quadruple Turing machines have three kinds of instructions, namely $q_i\alpha\beta q_j$ ("if You are in the state q_i reading α, print β in its place and go in state q_j") or $q_i\alpha A q_j$ ("if You are in state q_i reading α, go in direction A ($= L, R$) and change the state in q_j"), where α and β are tape symbols. Pure erasing is a special kind of printing $*$. A set of instructions $\Im$ is *consistent*, if for all quadruples we have that if $q_i\alpha A q_j \in \Im$ and $q_i\alpha B q_k \in \Im$, then $A = B$ and $q_j = q_k$. If at a given instant the machine is in a state q_j reading α, but there is no instruction on how to proceed, the computation ends. We can distinguish states with a special role called *final states*. Turing's original formalism of instructions actually used *quintuple*, whereas the quadruple approach is due to Post: the quintuple variation prints and moves at each step, while the quadruple variation does one or the other but not both. Post's version of Turing machine, when in a given state, either prints or moves and so its rules are more elementary. However a quadruple program can be converted into an equivalent quintuple program. There are just two kinds *quintuple* instructions and the sequence $q_i\alpha\beta A q_j$ must be read: "in state q_i read α, write in its place β, go in direction A and change the state in q_j", where A can be $R =$"go right", $L =$"go left"(sometimes also the istruction $I =$"do not move" is added, that does not add power). A more general definition of a single-tape Turing machine $\mathcal{M}$ can therefore be that it is a sequence $\mathcal{M} = \langle \Gamma, \Sigma, Q, *, F, q_o, \delta \rangle$ where Γ is the tape alphabet, $\Sigma \subset \Gamma$ is the *inputs alphabet* (a subset of tape symbols), $* \notin \Sigma$ is a special symbol ("blank") in Γ, Q is a finite set of states, $F \subseteq Q$ is a set of *final states* (e.g. $F = \{q_{f_0}, ..., q_{f_n}\}$), δ is the trasition function. and q_0 is the initial state. An *instantaneous configuration* of the machine is a string of the form $a_0a_1...a_iqa_{i+1}a_{i+2}...a_n$ where $a_0a_1...a_ia_{i+1}a_{i+2}...a_n$ is the portion of the tape between the leftmost and rightmost *nonblanks* (the remaing cells are *blank*), q is the state of the machine and a_{i+1} is the symbol scanned by the tape head. If C_j, C_i are configurations, then $C_j \longrightarrow C_i$ denotes the transition from the first to the second configuration (one step of the computation). Instead we denote with $C_j \longrightarrow^* C_i$ a finite sequence of configurations that begins with C_j and ends with C_i. A computation is just a finite sequence of such configurations. Turing machines can be used either to calculate functions (*transducers*), or to decide "YES" or "NO"problems. In the latter case, when at some point we reach a configuration C_n such that we can not go any further, we have these possible cases:

1. $C_n = ...\alpha q_f \beta...$, where q_f is a final state, then the computation *accepts*.

2. $C_n = ...\alpha q_j \beta...$, where q_j is *not* a final state, then the computation *rejects*.

Recall that, in general, the computation may also *not terminate* and go on forever; therefore it is necessary to distinguish between *recognition* and *acceptance*[3] Let Σ^* be the set of all finite strings (*words*) of an alphabet Σ. We call *language* (or *problem*) a subset of Σ^*:

1. A machine $\mathcal{M}$ *recognizes* (i.e. decides) a language $L \subseteq \Sigma^*$, if for all $\sigma \in \Sigma^*$ there is a *maximal computation* (i.e. such that we can no longer go on) $q_0\sigma \longrightarrow^*_{\mathcal{M}} wq_jz$ such that $q_j = q_f$ iff $\sigma \in L$.
2. A machine *accepts* a language $L \subseteq \Sigma^*$ if for every $\sigma \in L$ it is able to establish this membership; but if $\sigma \notin L$, although it do not says incorrectly that $\sigma \in L$, it can give a negative answer, or go on to calculate forever. In other words, this machine accepts L iff $L = \{\sigma \in \Sigma^* | q_0\sigma \longrightarrow^*_{\mathcal{M}} wq_fz\}$, where q_f is a final state.

[3] The the terminology of the literature is not uniform and the meaning of these terms often does not coincide with ours.

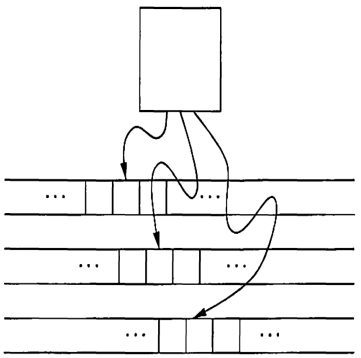


Figure 2. A multitape Turing machine

A language is called *decidable*, if there is a machine that recognizes it; is called *semidecidabile* (or computably enumerable) if there is a machine that accepts it. We say the *partial function* $f : \Sigma^* \longrightarrow \Sigma^*$ is *computable*, if there exists a machine $\mathcal{M}$ such that if $\sigma \in \Sigma^*$ and $f(\sigma) = y$, then the machine halts on σ and the final configuration has the form $\varepsilon q_f y$, where q_f is a final state and ε is the empty string and the machine does not halt on any string not in the domain of f.

Alternative definitions of Turing machines abound, all having the same power. For example, a *multitape* Turing machine is like an ordinary Turing machine, but with several tapes. Each tape has its own head for reading and writing. Every multitape Turing machine has an equivalent single-tape Turing machine (see Ausiello, Gambosi and d'Amore (2002) 181-92). A multitape machine (see fig.2) working in time $t(n)$ has an equivalent one-tape machine working in time $O(t(n)^2)$.

Definition 1. *With writing $f(n) \in O(g(n))$ we mean that there are n_0, c such that for all $n \geq n_0, f(n) \leq c \cdot g(n)$ (f is asymptotically bounded by g). If the other way round does not hold, we say that g grows faster than f. The set $O(g(n))$ is the order of growth of g.*

The Turing machine model perhaps may seem unrealistic. In 1971, for example, Hartmanis, Cook and Reckhow defined the model of random-access register machine (RAM model), for the purpose of introducing a theoretical model much closer in spirit to the design of Von Neumann architecture of modern computers. Actually all *reasonable* models define the same set of computable functions. We will show that also deterministic and *indeterministic* machines are equivalent *about what they can calculate*, but in this case *not for the amount of resources used.* However, if we consider *deterministic* machines, it is widely believed that all effective classical deterministic "sequential" (i.e. the instructions are executed in a sequence) models are polynomially- related with regard to the resources required to compute:

> *Invariance Thesis*. There exists a standard class of (deterministic) machine models, which includes among others all variants of Turing Machines and all variants of RAM machine models that simulate each other with Polynomially bounded overhead in time, and constant factor overhead in space (see Van Emde Boas (1990)).

More precisely, let us define the worst-case time complexity $t_M(n)$ of a machine (see 6 for the definition), which essentially means that for any w of length n the machine M take at most $t_M(n)$ steps to halt and return an output, and the worst space complexity $s_M(n)$, which means that for any w of length n the machine M needs to visit at most $s_M(n)$ cells to halt and return an output. In particular a machine $\mathcal{M}_1$ *run faster* than a machine $\mathcal{M}_2$, if $t_{\mathcal{M}_1}(n) \in O(t_{\mathcal{M}_2}(n))$ but not conversely.

A model of computation $\mathcal{C}$ simulates a model of computation $\mathcal{D}$ with *time overhead* $\alpha_t(x)$ and *space overhead* $\alpha_s(x)$, if for every machine $M_i \in \mathcal{D}$ there exists a machine $M_{i^*} \in \mathcal{C}$ such that

M_{i^*} simulates M_i and when a function $f(x)$ is computed by M_i in time $t(n)$ and space $s(n)$ then M_{i^*} requires $\alpha_t(t(n))$ time and $\alpha_s(s(n))$ space for computing it. It can be shown, for instance, that there exists a simulation of the *RAM* model by the Turing machine model with cubic time overhead and constant space overhead: $\alpha_t(t(n)) \in O(t(n)^3)$ and $\alpha_s(s(n)) \in O(s(n))$ (Slot and Emde Boas 1988). Hence these models can be simulated each other with a limited (polynomial) use of space and time. *Without taking into account the cost of mutual simulation*, many other models of computation are equivalent in power to Turing Machines. We deepen here the case of *non-deterministic* machines. Although simulable, the non-deterministic model is not simulable *efficiently*. Other models, as Randomized computation, Parallel computation, Analog computers, DNA computers, Quantum computers are excluded as well.

Deterministic Turing machines cannot include two or more quintuples (or quadruples) with the same first two elements: now let's drop this restriction. In his seminal 1936 paper, Turing also defined an extension of his "automatic machines" that he called "choice machines", which are now more commonly known as *nondeterministic Turing machines*. The execution of a nondeterministic Turing machine is not determined entirely by its input and its transition function; rather, at each step of its execution, the machine can (so to speak) *choice* the next move. Hence, a nondeterministic Turing machine allows for the possibility of *more than one next move from a given configuration*. In other words the *non-deterministic* machines are obtained by dropping the requirement of *coherence*. A non-deterministic algorithm can be interpreted as an algorithm that has the additional ability to *guess* a continuation in the scope of a finite set of possible continuations. This is the *nondeterministic* move. Hence every configuration may yield *more than one* next configuration. In contrast to the deterministic Turing machines, for which a computation is a sequence of configurations, a computation of a nondeterministic TM is a *tree of configurations* that can be reached from the initial configuration. The *root* of the tree is the initial configuration of the machine. The *children* of a node are the configurations that can follow in one move. The highest number of quintuples of the machine having the same first two elements is called the *degree of nondeterminism*. A node is a *leaf* if there is no next move. A *path* from the root to a leaf is an *accepting* computation if and only if the leaf is an *accepting configuration*. A *deterministic* computation corresponds to a particular sequence of choices. If *some* branch is an accepting computation (i.e. leads to a configuration of the form xqy with $q \in F$), then the machine accepts its input. It *rejects* its input, if *all branches* (i.e. for *every* choice of steps) are rejecting computations (i.e lead to a final configuration xqy with $q \notin F$).

A frequent and more intuitive presentation of the "physical model" of these machines, is indeed the one that adds the unit of reading and writing of the deterministic machine, a peripherial called *guessing-module* having it own write-only head. Suppose that the tape cells are numbered with integers $... -3, -2, -1, 0, 1, 2, 3...$ at the beginning the input is written starting from cell 1 to the right, the read-write head is scanning cell 1, the write-only head is scanning cell -1, and the finite state control is inactive. The *guessing module* writes, from right to left, a symbol at a time starting from the cell -1, then halts (*guessing stage*, actually the process may or may not terminate). At this point the control unit enters into play (*checking stage*) and the guessed string can be examined during the (purely deterministic) checking stage. The machine adopts the initial state q_0 and works as ordinary machine with the combination of the "guessed word" and the original input word, now as its input. During this second stage the guessing module and its guessing head are passive. An alternative way to consider a non-deterministic machines is therefore to say that a machine has the task of *guessing* the right solution to a problem.

The intuitive idea behind this approach to non-deterministic model is to provide a *solution-verifier* model, rather than a *problem-solver*: given a *guessed* solution to an instance of the problem, the algorithm can deterministically *verify* that it is a correct solution to the problem. Formally, a *verifier* for a language L is a *deterministic* Turing machine $\mathcal{V}$ such that $L = \{\sigma | \exists c(\mathcal{V} \text{ accepts } \langle \sigma, c \rangle)\}$. In other words, a verifier for L is a deterministic Turing machine that takes in an input x and an *evidence* c, and checks if c witnesses that x is in L. The additional information provided by c is also called *proof* or *certificate*. A nondeterministic computation can be therefore considered as divided into two phases: guess+verify. The *non-deterministic* moves guess a solution; then there is a *deterministic* subroutine which checks if there is a computation that accepts input. We call

verifier the above model,

Theorem 4. *The following are quivalent:*

1. *L is recursively enumerable (i.e. accepted by a deterministic Turing machine).*
2. *L is accepted by a non-deterministic Turing machine.*
3. *there exists a verifier for L.*

Proof. $1 \Rightarrow 2$ because a deterministic machine is a fortiori an indeterministic one. $2 \Rightarrow 3$: let us consider an accepting sequence of configurations of a nondeterministic machine; we encode this sequence and let c be a code of it. Let V be the verifier that uses c as a certificate. It imitates the nondeterministic machine by consulting c to know the next move to do. Clearly there exists c such that V accepts $\langle x, c\rangle$ if there exists a computation of the nondeterministic machine that accepts x. $3 \Rightarrow 1$: let V be a verifier for a language $L \subseteq \Sigma^*$ and let us consider the algorithm A that, on input $x \in \Sigma^*$, for each $k = 1, 2, 3...$ simulates V on the pair $\langle x, c\rangle$ for all $c \in \Sigma^*$ of length less or equal to k, for k steps and accepts if V accepts within this time. Now, if if the algorithm A on input x accepts, then this means that V itself accepts $\langle x, c\rangle$ and therefore $x \in L$. On the other hand, if $x \in L$, then V, being a verifier for L, accepts $\langle x, c\rangle$ for some c within t steps, for some t. Then the above program A accepts x at stage $k = max\{|c|, t\}$. Hence L is the language accepted by a deterministic algorithm A. QED

Actually we can simulate the nondeterministic model by the deterministic model, but *with an exponential cost.* If this loss of efficiency is inherent in some way, or due to the limits of our current knowledge, is a very important and unresolved issue. Essentially, it is the famous problem $\mathsf{P} = \mathsf{NP}$ we will discuss below.

Remark 1. *We remark that complexity classes that we will see, as* P *and* NP*, are defined with reference to the model Turing machines. However this does not imply a restriction, as the the most important complexity classes are invariant with respect to the computation model considered.*

We now show how to simulate a non deterministic computation with a deterministic machine. Let d be the degree of indeterminism of the machine. If for instance $d = 3$, then a computation is a subtree of the complete ternary tree as the following:

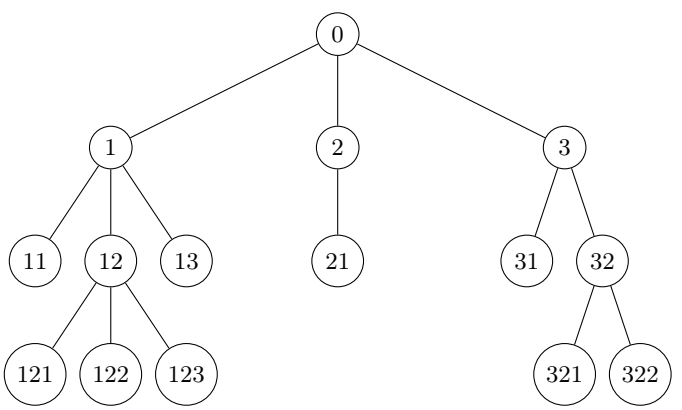


Figure 3. A nondeterministic computation

Each node is labelled with a string denoting a branch of nondeterministic computation. To prove the equivalence between nondeterministic and deterministic machine, we use a deterministic machine with three tapes. Suppose that in the nondeterministic machine each node in the computation tree has at most d children: in *tape 1* of the deterministic machine we copy the initial configuration of the nondeterministic machine, in *tape 2* we simulate the computation tree, typing in lexicographic order all finite strings of numbers from 1 to d that code nodes of the complete

d-ary tree, recalling that the computation tree is actually a *subtree* of this tree, and in *tape 3* we perform the computation. The algorithm works in phases, one phase for each string of numbers $j_1, ..., j_k$:

1. generates $j_1, ..., j_k$ on the second tape.
2. Copy the input from the first to the third tape,
3. Set $n = 1$ and let q the actual state of the nondeterministic machine and a the symbol scanned on tape 3:
   (a) if $\delta(q, a) = \emptyset$ or the number of elements of the set $\delta(q, a)$ is less than j_n, STOP (negative outcome).
   (b) Otherwise, take the j_n--th triple and apply the transition to the third tape.

Now, if there is a path of length k in the nondeterministic computation leading to a final configuration, that is to say a configuration $...xqy...$ where $q \in F$, then there is surely a phase of the simulation associated with a sequence of length k on the second tape that identifies this path and the simulation of this sequence will lead to a final state. If, on the other hand, such a path does not exist, then the associated deterministic machine can never reach a final state. But if the nondeterministic machine accepts a string in k steps then the deterministic machine requires $\leq k \cdot d^k$ steps. Actually, if there is a branch of the nondeterministic machine that accept in k steps, and d is the degree of indeterminacy, then the deterministic machine has to process at most

$$\sum_{j=0}^{k} d^j = 1 + d + d^2 + ... + d^k = \frac{d^{k+1} - 1}{d - 1} \leq k \cdot d^k$$

strings for simulating the computation.

Universality. A key property of Turing's model is universality: *there is a machine capable of simulating any other machine*. The possibility of a "universal" machine, was once considerd counterintuitive, because the parameters of the universal machine are fixed (alphabet size, number of states, and number of tapes), while the corresponding parameters for the machine being simulated could be much larger. The method by which the problem has been overcome is that of *coding*. We start here by considering a very simple deterministic one-tape kind of machines, with input alphabet $\{0, 1\}$ and tape alphabet $\{0, 1, *\}$. Since each machine $\mathcal{M}$ with a finite alphabet can be simulated by a machine $\mathcal{M}$ with tape alphabet $\{0, 1, *\}$, nothing is lost if therefore we consider, in our simple example, only machines whose input language is a binary language (see e.g. Du, Ko (2014) pp. 25-32 or Hopcroft, Motwani and Ullman (2000) pp. 377-81) and this argument may be extended to any fixed alphabet to get a binary encoding of a $k - tapes$ machine over this alphabet and also to nondeterministic machines. As we will see in the following chapters, the method of assigning a unique, unambiguously decodable numerical code to each symbol of a formal language, assigning natural numbers to symbols and then combining them into a single number to encode a string of symbols, was devised by Kurt Gödel in view of his incompleteness theorems, and has had wide applications. In our case, we can assign a numerical code to each state, tape symbol and direction and simply write each quintuple or quadruple as a sequence of five or four codes separated by some symbol between one code and another and between one string representing an instruction and another. Even better, identifying binary strings with natural numbers, such encoding can actually be done by means of binary strings[4]. Hence *Turing machines can ultimately be coded as binary strings*. Let therefore $\mathcal{M} =< \Pi, \Sigma, Q, q_0, \delta, \{q_f\} >$, where $\Pi = \{0, 1, *\}$, $Q = \{q_0, ..., q_{k+1}\}$, $q_f = q_{k+1}$. Let us consider this encoding:

$$\ulcorner 0 \urcorner = 0, \ulcorner 1 \urcorner = 00, \ulcorner * \urcorner = 000, \ulcorner q_i \urcorner = 0^{i+1}, \ulcorner L \urcorner = 0, \ulcorner R \urcorner = 00$$

A quintuple $q_i 01 R q_k$ will be coded by the binary string $0^{i+1}1010010010^{k+1}$. If the instructions of $\mathcal{M}$ are $\alpha_0, ..., \alpha_n$, then, if $\ulcorner \alpha_j \urcorner$ is the code of α_j, we set

$$\ulcorner \mathcal{M} \urcorner = 111 \ulcorner \alpha_0 \urcorner 11 \ulcorner \alpha_1 \urcorner 11 ... 11 \ulcorner \alpha_n \urcorner 111$$

[4] We identify a binary string σ with the natural number n s.t. the base-two representation of $n + 1$ is 1σ.

Each string encodes at most one machine. However, note that there can be many encodings of this kind for the same Turing machine: actually, since different orderings of the istructions give different codes, there are $m!$ equivalent codes for a machine of m istructions. We have coded all Turing machines in binary, but not all finite binary strings have the shape of a code. However, we can associate a Turing machine to *every* binary string, simply by associating the *trivial machine that rejects all inputs*, to the binary strings not having the shape of a code. In this way we can enumerate all Turing machines according to the lexicographic order of their index $\mathcal{M}_\varepsilon, \mathcal{M}_0, \mathcal{M}_1, \mathcal{M}_{00}, \mathcal{M}_{01}, \mathcal{M}_{000}$.... We have an *enumeration* of the one tapes machine, i.e. that every binary string codes a Turing machine and each machine is coded at least by one string.

We describe now a three-tapes machine $\mathcal{U}$ that simulates $\mathcal{M}$, where the first tape[5] contains the input code $\ulcorner\mathcal{M}\urcorner^\frown\sigma$, the second tape is used to simulate the tape of $\mathcal{M}$ and is initialized to contain the binary string σ and the third tape contains the current state code, say 0^i of $\mathcal{M}$. The machine $\mathcal{U}$ works as follows:

1. preliminary check the first tape content, to see e.g. if it has a prefix of the form of a legal code of a deterministic machine. If no, STOP, reject. Otherwise:
2. *Initialize the first, the second and the third tapes.* Copy σ in the second tape $\ulcorner q_0\urcorner$ in the third tape. The first tape head is positioned on the first symbol of $\ulcorner\mathcal{M}\urcorner$, that of the second tape on the first symbol of σ and the third on the $\ulcorner q_0\urcorner$'s leftmost symbol as well.
3. If the third tape contains $\ulcorner q_f\urcorner$, STOP and accept. Otherwise:
   (a) if the symbol currently scanned in second tape is for example 1 and the content of the third tape is $\ulcorner q_i\urcorner$, then the first tape head scans $\ulcorner\mathcal{M}\urcorner^\frown\sigma$ starting from the left up to the second block 111, searching for a substring starting with $110^{i+1}1001$; if it is not found STOP, reject. Otherwise:
   (b) let e.g. that substring be $110^{i+1}1001010010^{k+1}$: then write $\ulcorner q_k\urcorner$ in the third tape, replace the content of the second tape with $\ulcorner 0\urcorner$, move this head to the right and go back to step 3.

It is immediate to verify that $\mathcal{M}$ accepts σ, then $\mathcal{U}$ accepts $\ulcorner\mathcal{M}\urcorner^\frown\sigma$. If $\mathcal{M}$ does not converge on σ (i.e. goes on forever), then $\mathcal{U}$ does not converge on $\ulcorner\mathcal{M}\urcorner\sigma$. Lastly, if $\mathcal{M}$ converges on σ but does not accept, the same does $\mathcal{U}$ on $\ulcorner\mathcal{M}\urcorner^\frown\sigma$.

Halting problem. It would be useful to have an algorithm able to decide, given a program and given a particular input, if the program terminates on that input. We have seen that an algorithm $\mathcal{U}$ exists that halts (accepting or rejecting) on $\ulcorner\mathcal{M}\urcorner^\frown\sigma$, when $\mathcal{M}$ halts (accepting or rejecting) on σ. But $\mathcal{M}$ could also loop: is there an algorithm U that also *halts* and *reject* $\ulcorner\mathcal{M}\urcorner^\frown\sigma$, when $\mathcal{M}$ *loops* on σ? Alan Turing proved in 1936 that a general algorithm to solve the halting problem for all possible program-input pairs cannot exist (it also follows the unsolvability of *Enscheidungsproblem.* Many problems indeed are proved to be unsolvable *by reducing* to them the *Halting Problem*). Indeed, suppose that there is a machine V that decides the set:

$$HALT = \{\ulcorner\mathcal{M}\urcorner^\frown\sigma | \mathcal{M} \textit{ halts on } \sigma\}$$

i.e. that on input $\ulcorner\mathcal{M}\urcorner^\frown\sigma$, halts and *accepts*, if $\mathcal{M}$ halts on σ, or halts and *rejects*, if $\mathcal{M}$ does not halt (loops) on σ.

Build another machine D such that on input σ write $\sigma^\frown\sigma$ on its tape and runs V on $\ulcorner\mathcal{M}_\sigma\urcorner^\frown\sigma$, and accepts, if V rejects, or loops, if V accepts. Hence:

$$D \text{ halts on } \sigma \text{ iff } V \text{ rejects } \sigma^\frown\sigma \text{ iff } \mathcal{M}_\sigma \text{ loops on } \sigma$$

Hence D behaves differently from every machine on at least one input. However, being in the above enumeration, D is in turn some $\mathcal{M}_{\sigma^*}$, so that $D(\sigma^*) = \mathcal{M}_{\sigma^*}(\sigma^*)$ halts iff $\mathcal{M}_{\sigma^*}(\sigma^*)$ does not halt (contradiction). Hence, the machine V cannot exist.

1.4. Refinements of the Church-Turing thesis

In 1934, Church stated an early version of the "Church thesis":

[5] We use here the symbol $^\frown$ for concatenation of strings.

> A function is effectively calculable if and only if it is λ-definable.

Influenced by Gödel's lectures on recursive functions, Church later reformulated his thesis, replacing the $\lambda-$definable functions with general recursive functions. Actually Church proposes his thesis as a "correct *definition* of effective computability". On this line were also Turing and Gödel, that is to provide a *correct definition* of effective computability. In this regard Post says that Church initially proposed the following:

> A function of positive integers is effectively computable only if recursive.

as a "working hypothesis", but criticize Church for having later transformed it in a *definition*, because "to mask this identification under a definition ... blinds us to the need of its continual verification" (Post (1936), p. 105). Only Kleene in 1943 started to use the expression "Church thesis", that is, as a hypothesis that needs further evidence to ascertain the truth:

> Thesis I. Every computable function (effectively decidable predicate) is general recursive.

In Kleene (1967), p. 232 he summarized:

> ...this claim we call Turing's thesis. It was shortly shown by Turing 1937 that his computable functions are the same as the λ-definable functions, and hence the same as the general recursive functions. So Turing's and Church's theses are equivalent. We shall usually refer to them both as Church's thesis, or in connection with that one of its three versions which deals with "Turing machines" as the Church-Turing thesis.

As for the use of the thesis of Church-Turing, a distinction can be made (see Odifreddi (1989-1999) vol. I, p. 103) between:

1. *essential use in metamathematics*: for example, if I show that a function is not recursive, then the Church-Turing thesis says it is not computable by any means. To demonstrate the unsolvability of a problem, I can therefore translate into a function, demonstrating that this is not recursive (from the recursive unsolvability is deduced the unsolvability in an absolute sense).

2. *Non essential use*. For example when, as a shortcut, we give an algorithm and then we resort to the thesis of Church-Turing to say that it corresponds to a recursive function ("proof by Church thesis"). Instead of producing a rigorous recursive definition, we prefer an informal argument, when the possibility of its formalization is evident.

Although this thesis is highly corroborated today, having withstood numerous attacks, it is worth seeing the development of its interpretation. Today, have been proposed some variants of Church's thesis. Kreisel (1971) drew a distinction between the Church's thesis and a stronger thesis:

> He called Church's *Superthesis* a stronger version of Church's Thesis, in which one not only claims that certain mathematical tasks are *equivalent* to recursive ones, but rather that each such *task* is *equal* to some program for an idealized computer: to each mechanical rule or algorithm is assigned a more or less specific programme, modulo trivial conversions, which can be seen to define the same computation process as the rule (Odifreddi (1996), p. 403).

In Kreisel's opinion, Turing established a version of the superthesis for the notion of mechanically computable function. Circumscribing the analysis to deterministic discrete mechanical devices, Gandy (1980) recommended to distinguish the thesis of Church and Turing from what he called 'Thesis M':

> *Thesis* M: Whatever can be calculated by a machine is Turing-machine-computable.

The Church-Turing thesis does not imply the thesis M. Turing in fact started from the analysis of a computation performed by a human being, but a thesis concerning procedures executable by a human has no implication concerning the extent of the procedures that machines are capable of carrying out "since, for example, there might be, among a machine's repertoire of atomic operations, operations that no human being who is working effectively is able to perform" (Copeland (2000)). To justify the thesis M, in order to specify what is meant by 'mechanically computable', Gandy considered only discrete machines with restrictions specified by a set of axioms (the most important are *local causation* and *unique assembly*) motivated on the basis of considerations drawn from special relativity and quantum mechanics.

> *Gandy's thesis.* Each computable function from a machine that satisfies the axioms of Gandy is Turing-computable.

A stronger version of the Church-Turing thesis states that any algorithmic process is simulated *efficiently* by a Turing machine. In the 1970s Robert Solovay and Volker Strassen worked out, however, a probabilistic primality test using an efficient algorithm, which was the first of its kind. They used *randomness* as an essential part of the algorithm. No efficient deterministic test for primality was known in that time, thus:

> it seemed as though computers with access to a random number generator would be able to efficiently perform computational tasks with no efficient solution on a conventional deterministic Turing machine (Nielsen, Chuang (2010) p. 6).

The probabilistic Turing machine began to look to many as the reference model. It is in particular provable that if it is possible to calculate a function with k elementary operations in a model of computation, then it is possible to calculate the function with a probabilistic machine by $p(k)$ elementary operations, for some polynomial $p(x)$. From which this modification of the strong version of the Church Turing thesis (see Nielsen, Chuang (2010) p. 140):

> *Strong Church–Turing thesis*: Any model of computation can be simulated on a probabilistic Turing machine with at most a polynomial increase in the number of elementary operations required.

In the early 80s of the 20th century Feynman highlights some essential difficulties in simulating quantum systems on classical computers and proposes as a program to develop computers based on the principles of quantum physics. In the 80's David Deutsch poses the question of whether a quantum computer can efficiently solve computational problems that have no efficient solution even with probabilistic Turing machines. In 1994 Peter Shor gave a significant response to this problem, creating a quantum algorithm that can factor a compound integer N in polynomial time in the number of digits $O(\log N)$ of N. Whereas the computations are also physical processes implemented by a physical system Deutsch poses the question of whether the laws of physics can be used to prove a version of the thesis of Church-Turing and proposed this variant (although very different from the original thesis, which speaks of effective methods, rather than finitely realisable physical systems):

> *CDT Thesis (Church, Turing, Deutsch).* Every finitely realisable physical system can be simulated by a universal model computing machine operating by finite means.

where Deutsch suggests that the reference computation model is, however, the Quantum Computer, but according to Nielsen, Chuang (2010) it is not clear whether Deutsch's notion of a *Universal Quantum Computer* is sufficient to efficiently simulate an arbitrary physical system. In the following chapters, we will be mainly interested in this thesis, known as *Edmonds-Cobham-Cook-Karp thesis*:

> A function is feasibly computable if and only if it is computed by some deterministic machine in polynomial time. A problem is feasibly decidable in case it is decidable by a deterministic Turing machine using a polynomial amount of computation time.

The seminal work Hartmanis and Stearns (1965) provided for the first time a precise definition of the time and space complexity of an algorithm, defining the concept of a complexity class. This statement is similar in form to Church's Thesis[6], however, it must be emphasised that effective computability it is not practical or feasible computability.

1.5. Turing on incomputability and the undecidability

In these two sections, we wish to establish a comparison between the proof of the *undecidability* theorem given in Turing (1936) on the one hand, and Cook (1971) and Levin (1973) independently proved *unfeasibility* theorem, on the other hand, as a paradigmatic example of the transition between two phases in the development of logic and at the same time emphasise a *trait-d'union* in the idea of logical formalisation of the computation of a Turing machine, taking a cue from a famous letter from Gödel to Von Neumann, in which research into feasibility is presented as a natural continuation of research into decidability, after the negative results of Church and Turing. In order to do this, without loss of generality, in both cases we adopt the simplest deterministic machine model, i.e. the one-tape, in which instructions are given *à la Post* by quadruples following Davis, Sigal and Weyuker (1994) pp. 451-57. For other proposals see e.g. Sipser (2006) pp. 276-82 or, for a more detailed version of it Ausiello, Gambosi and d'Amore (2002) pp. 329-35 and with regard to the Turing theorem, the Open Logic Project (2015) pp. 501-09. To show that the decision problem (*Entscheidungsproblem*) is unsolvable, Turing applied its famous result on the halting problem in this form:

> *Unsolvability of the halting problem.* There is no algorithm that, given a description of an arbitrary program Φ that computes a partial numerical function and a number n, decides whether Φ with input n halts in finitely many steps.

This result has a wide range of applications and we will show that many problems are algorithmically unsolvable by "coding" the halting problem into these problems. A problem A is reduced to a problem B, when a solution to B can be used to solve A; we prove that the solution to the *Entscheidungsproblem* could be used to find a solution for the halting problem (that is unsolvable!).

Strategy: given a Turing machine and a natural number n we want to build a set of first-order formulas Γ and a first-order formula ψ such that $\Gamma \vdash \psi$ if and only if the machine halts on input n. If it were possible to decide whether $\Gamma \vdash \psi$, we could also decide the halting problem. Since the latter is unsolvable, then there are Γ and ψ for which we are not able to decide whether ψ follows from Γ.

Preliminary will be an approproate description of the machine:

1. consider the boxes of the tape as numbered with integers:

$$... - 2, -1, 0, 1, 2, ...$$

2. instants of discrete time will be denoted by natural numbers: at 0 instant, the machine scans the square 0.

3. the alphabet of the machine will have symbols $S_0, S_1, ..., S_r$, where S_0 is the blank.

4. The predicative language chosen to describe the machine will contain predicative symbols $Q_i(t, x) =$" at time t, in the state q_i, it is scanned the square x"; $S_j(t, y); =$" at the instant t, the contents of the square y is the symbol S_j". In particular there will be the successor function symbol $S(x) = x + 1$, the constant $\overline{0}$ and the relational symbols $=$ and $<$ and relative axioms. In particular axioms for $<$ say that this relation is transitive, antireflexive and that S is total, injective and that $x < S(x)$.

[6] Beware that in some texts the Edmonds-Cobham thesis means another statement, namely: *the time complexities in any two reasonable and general models are polynomially related*, and it is also known as *extended Church-Turing thesis*.

We denote from $\overline{n}$ the numeral $\overbrace{S(S(S(...(S(0))...)))}^{n-times}$. Since we dont have subraction, the negative integers may be formalized as follows:

$$Q_i(t,-p) \leftrightarrow \exists z(Q_i(t,z) \wedge \overbrace{S(S(S...S(z)...)))}^{p-times} = 0)$$

Representation of the initial conditions. in the initial state q_0 the machine observes a sequence of n symbols S_1 (so we represent the input n); the rest of the tape is white:

$$Q_0(\overline{0},\overline{0}) \wedge \bigwedge_{0 \leq i < n} S_1(\overline{0},\overline{i}) \wedge \forall y(\overline{n} < y \rightarrow S_0(\overline{0},y))$$

Representation of the final conditions.

I want a formula ψ such that ψ is true in the moment in which the machine halts. Note that the machine halts if there is a time t in which the machine, lying in state q_i, reads S_j, but no instruction begins with q_iS_j. Since both the number of states and symbols is finite, we can form the disjunction:

$$\bigvee_{i,j} \exists t \exists x(Q_i(t,x) \wedge S_j(t,x))$$

of the q_i, S_j such that no instruction starts with q_iS_j. This will be our ψ.

Representation of instructions.

The instruction $q_iS_jS_kq_m$ is formalized as follows:

$$\forall t \forall x(Q_i(t,x) \wedge S_j(t,x) \rightarrow (Q_m(t+1,x) \wedge S_k(t+1,x) \wedge$$

$$\wedge \forall y(y \neq x \rightarrow \bigwedge_{i \leq r}(S_i(t,y) \rightarrow S_i(t+1,y))$$

The instruction $q_iS_jRq_m$ is formalized as follows:

$$\forall t \forall x(Q_i(t,x) \wedge S_j(t,x) \rightarrow (Q_m(t+1,x+1) \wedge \forall y \bigwedge_{i \leq r}(S_i(t,y) \rightarrow S_i(t+1,y))$$

We leave it as an exercise to the reader to formalise the instruction $q_iS_jLq_m$. The set Γ will contain all these axioms relative to the instructions and to the initial conditions, moreover the axioms for successor, and preorder and identity. Note that according to the intended interpretation $\Gamma \vdash \psi$ implies that the machine halts on input n. The reverse will be demonstrated by induction. It will follow the equivalence between (a) $\Gamma \vdash \psi$ and (b) " The machine halts on input n ", but being (b) undecidable, it will also undecidable (a). First of all we represent the state of the machine at instant s as the conjunction of these three formulas:

1. $Q_i(\overline{s},\overline{p})$
2. $S_{j_0}(\overline{s},\overline{p_0}) \wedge ... \wedge S_j(\overline{s},\overline{p}) \wedge ... \wedge S_{j_v}(\overline{s},\overline{p_v})$
3. $\forall y((y \neq \overline{p_0} \wedge ... \wedge y \neq \overline{p} \wedge ...y \neq \overline{p_v}) \rightarrow S_0(\overline{s},y))$

Suppose the machine halts at instant s reading S_j in square p and lying in state q_i; the description of this situation is $Q_i(\overline{s},\overline{p}) \wedge S_j(\overline{s},\overline{p})$, that implies ψ. In essence, to demonstrate that from Γ follows ψ is enough therefore to show that for every $s \geq 0$, if the machine does not halts before s, then from Γ follows a description of the machine at stage s.

Induction If $s = 0$, then it is obvious, since Γ contains the initial state of the machine. For the inductive step, suppose that at s the machine has not stopped; for (IH) we are able to derive from Γ a description of the machine at instant s. Suppose, then, that at s, the machine scans the square p containing the symbol S_j, and is in the state q_i. Because the machine does not stop, the possible continuations will be as follows:

1. if the instruction is $q_iS_jS_kq_m$, then Γ will contain the formula:

$$\forall t \forall x(Q_i(t,x) \wedge S_j(t,x) \rightarrow (Q_m(t+1,x) \wedge S_k(t+1,x) \wedge$$

$$\wedge \forall y(y \neq x \rightarrow \bigwedge_{i \leq r}(S_i(t,y) \rightarrow S_i(t+1,y))$$

But this, together with the inductive hypothesis, that is, that at s the machine is described by 1., 2., 3., and the axioms for $<, =$ and successor, implies:

(a) $Q_m(\overline{s+1}, \overline{p}) \wedge S_{j_0}(\overline{s+1}, \overline{p_0}) \wedge ... \wedge S_k(\overline{s+1}, \overline{p}) \wedge ... \wedge S_{j_v}(\overline{s+1}, \overline{p_v})$

(b) $\forall y(y \neq \overline{p_0} \wedge ...y \neq \overline{p} \wedge ...y \neq \overline{p_v}) \rightarrow S_0(\overline{s+1}, y)$

That is, a description of the machine at stage $s+1$.

2. If the instruction is $q_iS_jRq_m$, then Γ will contain the axiom:

$$\forall t \forall x(Q_i(t,x) \wedge S_j(t,x) \rightarrow (Q_m(t+1,x+1) \wedge \forall y \bigwedge_{i \leq r}(S_i(t,y) \rightarrow S_i(t+1,y))$$

similarly to the previous case, we obtain:

(a) $Q_m(\overline{s+1}, \overline{p+1}) \wedge S_{j_0}(\overline{s+1}, \overline{p_0}) \wedge ... \wedge S_k(\overline{s+1}, \overline{p+1}) \wedge ... \wedge S_{j_v}(\overline{s+1}, \overline{p_v})$

(b) $\forall y(y \neq \overline{p_0} \wedge ...y \neq \overline{p+1} \wedge ...y \neq \overline{p_v}) \rightarrow S_0(\overline{s+1}, y)$

That is, a description of the machine at $s+1$.

3. If the instruction is $q_iS_jLq_m$, we leave this case as an exercise for the reader.

In any case, Γ implies a description of the machine at $s+1$. Thereby ends the inductive step. We conclude by induction that for all non-negative number s, if the machine does not stop before s, then Γ implies a description of it at instant s. Therefore, if the machine halts at s, then its description will contain the expression $Q_i(s,p) \wedge S_j(s,p)$, that implies ψ. That is, if the machine halts on input n, then from Γ follows ψ.

1.6. Cook and Levin's theorem and the unfeasibility

Computability in polynomial time has become synonymous with feasibly decidability. Church's Thesis concerns the concept of effective computability. Analogously, the thesis put forward by Cobham, Edmonds and Cook concerns the feasible computations, by identifying efficient computablility with *computable in polynomial time*. Note that, like the Church–Turing thesis, this is not a theorem. For the continuation of the discussion on this topic, it is necessary to specify the notion of computational efficiency and the complexity measures we will adopt to quantify it. We consider one-tape machines and we define:

1. the *running time* of a deterministic machine $\mathcal{M}$ on an input x is the number of steps $time_M(x)$ made by the machine until it halts (note that it can also be infinite[7]).
2. The *working space* of a deterministic machine $\mathcal{M}$ on an input x is the number of cells $space_M(x)$ visited at least once by the head during the computation (note that it can also be infinite and that can be finite although the machine runs forever).
3. the *time complexity* of the machine is:

$$t_M(n) = max\{time_M(x) || x| = n\}$$

i.e. the maximum number of steps that M makes for inputs of lengths n, before halting.

[7] A single step of computation is composed by the following actions: read the input symbol from the active cell, look up the transition rule associated with the current state and input symbol, overwrite the input symbol with the new symbol and change the current state according to the transition rule.

4. the *space complexity* of the machine is:

$$s_M(n) = max\{space_M(x)||x| = n\}$$

As for nondeterministic Turing machines, $time_M(x)$ and $space_M(x)$ are defined as the minima among the accepting paths. Moreover:

(a) $t_M(n) = max\{time_M(x)||x| = n \text{ and } M \text{ accepts } x\}$ and

(b) $s_M(n) = max\{space_M(x)||x| = n \text{ and } M \text{ accepts } x\}$

In case no x of length n is accepted by M, some convention is added to these measures. In case of time complexity, add $n + 1$, the time needed just for reading the input. In case of space, if no x of length n is accepted by M, put $s_M(n) = 1$. If no accepting path exist, these functions are undefined (we follow Du, Ko (2014) pp. 19-21).

1. $\mathsf{DTIME}(f(n))$=set of languages decided by a deterministic Turing machine $\mathcal{M}$ in time $t_M(n) \leq f(n)$,
2. $\mathsf{NTIME}(f(n))$= set of languages accepted by a nondeterministic Turing machine in time $t_M(n) \leq f(n)$,
3. $\mathsf{DSPACE}(f(n))$=set of languages decided by a deterministic Turing machine $\mathcal{M}$ in space $s_M(n) \leq f(n)$,
4. $\mathsf{NSPACE}(f(n))$= set of languages accepted by a nondeterministic Turing machine in space $s_M(n) \leq f(n)$,

Theorem 5. *The following relations hold:*

$$\mathsf{DTIME}(f(n)) \subseteq \mathsf{NTIME}(f(n)) \subseteq \mathsf{DSPACE}(f(n)) \subseteq \mathsf{DTIME}(2^{c \cdot f(n)})$$

Proof. See e.g. Ausiello, Gambosi and d'Amore (2002), 300-03. QED

Now we define the complexity classes:

1. $\mathsf{P} = \cup_{k=0}^{\infty}\mathsf{DTIME}(n^k)$ (or $\mathsf{DTIME}(Poly)$) is the set of languages that can be decided by a deterministic Turing machine in polynomial time in the dimension of the input,
2. $\mathsf{NP} = \cup_{k=0}^{\infty}\mathsf{NTIME}(n^k)$ (or $\mathsf{NTIME}(Poly)$) is set of languages that can be accepted by a nondeterministic machine in polynomial time in t he dimension of the input.
3. (PSPACE and $\mathsf{NPSPACE}$ are defined analogously).

By definition, any language in P or NP is decidable. However some problems are *hard to solve*, but *easy to check*. We consider P the set of decision problems solvable quickly and NP the set of problems *verifiable* quickly: however, even today, *it is not known whether* $\mathsf{P} \neq \mathsf{NP}$.

Definition 2. *A total function f is said to be polynomial-time computable if there is a Turing machine $\mathcal{M}$ that computes f, and a polynomial $p(n)$, such that the number of steps in the computation by the machine with input x is bounded by $p(|x|)$.*

We say that Q is polynomial-time reducible to L (in symbols $Q \leq_p L$) if and only if there exists a function f computable in polynomial time such that for each x, $x \in Q$ iff $f(x) \in L$. Furthermore L is said $\mathsf{NP} - hard$ if for all $Q \in \mathsf{NP}$, $Q \leq_p L$. If also $L \in \mathsf{NP}$, then we say that L is NP-complete.

A popular NP-complete problem is the so-called *traveling salesman problem*. Remember that a Hamiltonian cycle is a cycle through a graph that visits all the nodes of the graph once and only once. Let us suppose we have n cities and their distances: we are asked to find the shortest tour that touches all cities exactly once (with a final return to the starting point). In another version, given a number B, the question is whether there is a tour that touches all cities at most once, shorter than or equal to B. In a more mathematical terms, given a complete, undirected and weighted graph (*complete* means that each pair of graph vertices is connected by an edge and *weighted* means that a weight is associated with each edge), it is asked to find a *Hamiltonian circuit* with the least weight (or, in the second version, less weight or equal to B). The problem is

solvable by brute force by considering $\sim n!$ possible permutations, but are not known algorithms that work in polynomial time. Little hope to find: the traveling salesman problem is NP -complete. On the other hand, however, it is obvious that the problem can be verified in polynomial time: if you give me a tour, I reckon its cost by adding up the cost of all their edges; then I check if it is less than or equal to B.

Intuitively, the NP-complete problems are the *most difficult* in the class NP. Clearly $\mathsf{P} \subseteq \mathsf{NP}$, but the problem $\mathsf{P} = \mathsf{NP}$ is still unsolved. However if there is a language L that is NP-complete and belong to P, then $\mathsf{P} = \mathsf{NP}$. Schematically, we prove this second statement by showing that, under the hypothesis that there exist a language L that is NP-complete and also belong to P, it follows that for any Q, if $Q \in \mathsf{NP}$, then $Q \in \mathsf{P}$.

It actually suffices that the number of steps be bounded by $c \cdot |x|^r$ for some r (see Davis, Sigal and Weyuker (1994), 441-43). Now, being L an NP-complete set, then if $Q \in \mathsf{NP}$, also we have $Q \leq_p L$, namely, $Q = \{x | f(x) \in L\}$ for some polynomial-time computable function $f(x)$. Still by hypothesis $L \in \mathsf{P}$. Then we show that $Q = \{x \in \Sigma^* | f(x) \in L\}$ is also in P. Let $\mathcal{M}$ be a machine that decides L in time (say) bounded by $c \cdot |x|^r$ (where $|x|$ denotes the length of x) and let $\mathcal{V}$ a machine that computes $f(x)$ in time bounded by $a \cdot |x|^s$. Let us consider the composition $\mathcal{W}$ of these machines: first run $\mathcal{V}$ on x to compute $f(x)$, then run $\mathcal{M}$ to decide whether $f(x) \in L$.

Note that the machine $\mathcal{V}$ cannot print more symbols that steps in the computation; hence $|f(x)| \leq |x| + a \cdot |x|^s$ and then W needs less than $b \cdot |x|^{rs}$ steps (for some b) to decide Q. It follows that $Q \in \mathsf{P}$.

Let us consider now the set SAT of *satisfable* propositional formulas (in conjunctive normal form). Although to analize the full truth-table has an exponential cost (in the number of propositional variables), it is not hard, thinking to the check+guess method, to realise that if somebody gives me a line of the truth table, I can rapidly check whether or not this line satisfies a formula. Actually $\mathsf{SAT} \in \mathsf{NP}$, because a nondeterministic algorithm that accept a formula ϕ can be defined, that works in time $a \cdot |\phi|^2$ *Cook's and Levin's theorem* (1971) says that actually SAT is NP-complete. From the previous result it follows therefore that

$$\mathsf{P} = \mathsf{NP}\ iff\ \mathsf{SAT} \in \mathsf{P}$$

Therefore, if someone found an efficient method to establish whether a formula of propositional calculus is satisfable, she would have solved this famous problem.

The first "formulation" of the problem if $\mathsf{P} = \mathsf{NP}$, a key issue in computational complexity, albeit implicitly, dates back to Kurt Gödel, in a letter to Von Neumann in 1956, although the importance of the problem raised was clear only with the work of Stephen Cook in the 1970s. In short, this is Gödel's argument: the unsolvability of the *Entscheidungsproblem* implies that the problem whether a formula ϕ of first order language has a proof is undecidable: but what, if we consider the problem whether ϕ has a proof of at most n symbols? This in principle is decidable, but at what cost? Actually Gödel considered that a proof of his conjecture would constitute a refinement of Hilbert's *Entscheidungsproblem*:

> Despite the unsolvability of the *Entscheidungsproblem*, the mental effort of the mathematician in the case of yes/no questions could be completely replaced by machines. One would indeed here to simply select an n so large that, if the machine yields no result, there would then also be no reason to think further about the problem. (Gödel's Collected Works (1986-2005) vol. V, 135).

Let us write $\vdash^n \phi$ for "the formula ϕ is provable in the predicate calculus with at most n *symbols*". Gödel asked what is the complexity of the set:

$$\mathsf{T} = \{\langle \phi, n \rangle | \vdash^n \phi\}$$

In particular, whether it is decidable by a Turing machine in time n^2. According to a widespread opinion, the polynomial-time computation is a good formalization of practical, realistic, feasible (by humans or computers) computability, a narrower requirement than the decidability in principle.

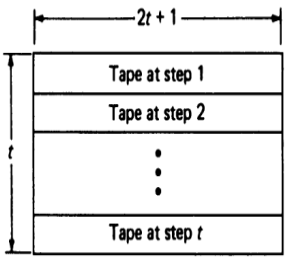


Figure 4. Let $t = p(|u|)$.

However Gödel was thinking of *linear* and *quadratic* time as corresponding to *feasibility*. Note that if our language contains k-symbols the problem is solvable brutally making a list of all proofs with at most n symbols, then see which is right for us; but this procedure has obviously a complexity of order k^n. T is in NP since a nondeterministic algorithm for verifying it consists of guessing a proof of $\leq n$ symbols and then verifying in polynomial time if the guessed proof is valid (see Buss (1995)). Today we know that the problem is NP-complete: hence if the hypothesis of Gödel was right, therefore, it would also be in P and as a result we would have $\mathsf{P} = \mathsf{NP}$. Indeed, Cook showed that the satisfability problem SAT can be reduced to it: let un consider a boolean formula $\Phi(p_0, .., p_n)$ and replace each propositional variable p_i with $x_i = y$. Call $\Phi^*(x_0, ..., x_n, y)$ this new formula. It is provable that the boolean formula is satisfable if and only if the formula:

$$\exists y(\exists z(z \neq y) \rightarrow \exists x_0, ..., \exists x_n \Phi^*(x_0, ..., x_n, y))$$

is provable with a proof of length $|\Phi|^c$, for some constant c.

But what is the length of proofs? There are infinitely many proofs of fixed *number of lines*, but finitely many proofs with *fixed number of occurrences of symbols*; in the above argument we considered the total amount of occurrences of symbols in a proof (its *size*). For instance, a variable x_n can be considered as a symbol x followed by $n-$marks and therefore of size $n + 1$. If we adopt the *number of formulas* as a measure of the length of proofs, it is undecidable for sequent calculus (see Buss (1991)). In general, many results in the study of complexity of proofs are established only for the case of symbol-length and it is not known whether those results can be transferred to the case of step-length.

We now come to the proof of the result discovered independently by Cook and Levin. The time required for a precise proof that the problem SAT is in NP is $O(|x|^2)$. This is shown for instance in Davis, Sigal and Weyuker (1994), 448, which we also follow in the precise exposition of hardness.

Theorem 6. (Cook and Levin 1971) SAT *is* NP *hard.*

Proof. Let $L \in \mathsf{NP}$ and M be a nondeterministic Turing machine accepting L in time $p(|x|)$ on input $x \in L$ (for some polynomial $p(n)$). We show that there is a *polynomial time computable function* that translates each input u into a propositional formula in conjunctive normal form δ_u such that:

$$M \text{ accepts } u \text{ iff } \delta_u \in \mathsf{SAT}$$

This δ_u will be a simulation of the computation of M on input u. To simplify matter, let us assume that an accepting computation halts *exactly* after $P(|u|)$ steps. To do this, we allow repetitions of the same configuration $C \Rightarrow C$. After $p(|u|)$ steps the scanned squares are at most $p(|u|)$ (going left or right). In the following we consider therefore a matrix (or *array*) of $2 \cdot p(|u|) + 1 \times p(|u|)$ cells, sufficient to exibit all information needed (see fig.4).

The first row of this array represents the initial configuration of the machine, in state q_1, scanning the symbol blank immediately to the left of the first symbol of u:

$$\overbrace{*...q_1*}^{t+1-times} \; s_{u_1}....s_{u_z} \; \overbrace{*...*}^{t-|u|-times}$$

where $z = |u|$ and $u = s_1....s_z$ and $t = p(|u|)$. Let the set of states of L be $Q = \{q_1, ..., q_m\}$ where q_1 is the initial state and q_m an accepting state and let the set of tape symbols be $S = \{s_0, s_1, ..., s_r\}$, where $* = s_0$. We use a set of propositional atoms $p_{h,j,k}$ and $w_{i,j,k}$, where $1 \leq h \leq m$, $0 \leq i \leq r$, $1 \leq j \leq 2 \cdot p(|u|)$ and $1 \leq k \leq p(|u|)$ and we consider an assignement of values such that:

1. $p_{h,j,k}$ is true iff M is in state q_h scanning the jth cell at step k.
2. $w_{i,j,k}$ is true iff the symbol s_i is in the jth cell at step k.

Let us say that the length of each of these atoms is of the order of complexity[8] $O(t)$. Lastly, it will be useful this abbreviation:

$$\Theta\{x_i | 1 \leq i \leq e\} = \bigwedge_{1 \leq l < f \leq e} (\neg x_l \vee \neg x_f) \wedge \bigvee_{1 \leq l \leq e} x_l$$

which is true iff exctly one of $x_1, ..., x_e$ is true. Since there are $e(e-1)/2$ pairs (l, f) with $1 \leq l < f \leq e$ and the second conjunct contains e literals, the whole formula has $O(e^2 t)$ symbols.

We formalize now the computation (where for "length"we mean the number of symbols):

1. The machine in state q_1 scans s_0 immediately to the left of u (the first row of the array):

$$\bigwedge_j w_{0,j,1} \wedge \bigwedge_j w_{u_j, p(|u|)+j+1,1} \wedge \bigwedge w_{0,p(|u|)+|u|+j+1,1} \wedge p_{1,p(|u|)+1,1}$$

   The length of this formula is $O(t^2)$.

2. At each step k there is a unique state and a unique scanned symbol:

$$\Theta\{p_{h,j,k} | 1 \leq h \leq m, 1 \leq j \leq 2 \cdot p(|u|) + 1\}$$

   The length is $O(t^4)$.

3. Each entry on the array contains exactly one symbol:

$$\bigwedge_k \bigwedge_j \Theta\{w_{i,j,k} | 0 \leq i \leq r\}$$

   This formula is of length $O(t^3)$.

4. Each configuration after the first is obtained by applying the istructions. Let us assume that the program is made of the following istructions:
   (a) $q_{i_a} s_{j_a} s_{k_a} q_{e_a} \quad (a = 1, 2, 3...r_a)$
   (b) $q_{i_b} s_{j_b} R q_{e_b} \quad (b = 1, 2, 3...r_b)$
   (c) $q_{i_c} s_{j_c} L q_{e_c} \quad (c = 1, 2, 3...r_c)$

[8] This bound depends on how we encode the index sequences and therefore the variables. In later editions the authors modified it, but with an efficient (binary) encoding, the length of these variables can be kept of order $O(t)$. However, here this constitutes a minor point.

Let moreover:

$$NOTHEAD(j,k) = \bigvee_i (w_{i,j,k} \wedge w_{i,j,k+1}) \wedge \bigwedge_h \neg p_{h,j,k}$$

(M is not scanning the jth cell at the k step) and let:

$$IDENT(j,k) = \bigvee_h \bigvee_i (p_{h,j,k} \wedge p_{h,j,k+1} \wedge w_{i,j,k} \wedge w_{i,j,k+1})$$

(M scans the jth cell at kth and $k+1th$ steps; the state and the symbol scanned is the same in both these successive configurations). Moreover let:

$$\alpha(j,k) = \bigvee_a (p_{i_a,j,k} \wedge w_{j_a,j,k} \wedge w_{k_a,j,k+1} \wedge p_{e_a,j,k+1})$$

(The $k+1th$ step comes from the k step applying one istruction from the group (a)). Analogously define $\beta(j,k)$ and $\gamma(j,k)$ relative to the groups (b) and (c)). Thus (4.) is defined as:

$$\bigwedge_k \bigwedge_j (NOTHEAD(j,k) \vee IDENT(j,k) \vee \alpha(j,k) \vee \beta(j,k) \vee \gamma(j,k))$$

This formula has length $O(t^3)$.

5. The $p(|u|)$-th configuration is terminal (recall that q_m is an accepting state):

$$\bigvee_j p_{m,j,p(|u|)}$$

This formula has length $O(t^2)$.

Now take δ_u as the conjunction of (1.)-(5.) and note that if M accepts u, then $\delta_u \in \mathsf{SAT}$, because the above assignement of truth values to the propositional atoms makes δ_u true. Conversely, if there is an interpretation that makes δ_u true, then we can reconstruct our array: by (3.) for each j,k there is a unique i such that $w_{h,j,k}$ is true: from this we can *uniquely* reconstruct the array in figure. By (2.) there are unique q_h and j such that $p_{h,j,k}$ is true. Hence each row can be made into a configuration of the machine. By (1.), the configuration corresponding to the first row of the array is an initial configuration. By (4.), for each row of the array after the first, the corresponding configuration is identical to it or results from it using one of the quadruples of the program. Finally, by (5.) the entire sequence of configurations constitutes an accepting computation. Thus, u is accepted by M.

We now shown that there is a polynomial-time computable function that maps each string u onto the corresponding CNF formula δ_u. Observe that the CNF formulas of (2.)-(5.) do not depend on u: a Turing machine can be constructed to write these on a tape in a number of steps proportional to the length of the expression, which is $O(t^4)$, and hence polynomial in $|u|$. As for (1.) some of its atoms do not depend directly on u; producing this part simply involves writing $O(t^2)$ symbols. The remaining atoms correspond in a one-one manner to the symbols making up u and can be produced in a number of steps proportional to $|u|$. This completes the proof of the Cook-Levin theorem as presented in Davis, Sigal and Weyuker (1994). QED

The Cook-Levin theorem allows one to prove that many other important problems are NP-complete, by showing that SAT can be efficiently reduced to them. A proof that an NP problem is NP-complete is a proof that the problem is far from feasible, unless every NP problem is in P, which is not known, but which is unlikely and which, if proved, would constitute a shocking novelty for computer science.

2. Abstract views of incompleteness

2.1. Models of computation: recursive functions

Computability and recursion are often identified. However in principle they are different concepts (see Soare (1996)): computability is a more general term, whereas the primary mathematical meaning of *recursion* has always been *definition by induction* (i.e., by recursion), namely defining a function f at an argument x using its own previously defined values. In the seminal paper Gödel (1931) the great Czech logician used the notion of a *primitive recursive function* (called *rekursive Funktion*) because these functions were easily representable in formal system PA for arithmetic, and were sufficient to enable him to code all the syntactic objects so that he could obtain self-reference and thereby incompleteness. The main property of this class of functions is the primitive recursion scheme, which yields an *inductive* definition of $f(n+1)$ using the preceding value $f(n)$ and previously defined functions g and h. Gödel realised, however, that the primitive recursive functions did not include all effectively computable functions, and in 1934 he proposed a wider class of functions based on an earlier suggestion of Herbrand. Gödel called these the *general recursive functions*.

Actually in the initial study of computability from 1931 to 1937 researchers considered only *total* computable functions. It was Kleene in 1938 who first proposed considering *partial* computable functions. Hence Kleene extended the class of primitive recursive functions adding a scheme of *Unbounded Search*, to introudce $\mu-recursive\ partial$ functions, i.e. the computation model whose fundamental concepts we will illustrate in this chapter (see again Soare (1996)).

Definition 3. *A function f from natural numbers to natural numbers is said partial if $Dom(f) \subseteq \mathbb{N}$; if in particular $Dom(f) = \mathbb{N}$, then the function is total.*

Definition 4. *A set $X \subseteq \mathbb{N}$ is called recursive iff its characteristic function is recursive:*

$$\chi_X(x) = \begin{cases} 1 & \text{if } x \in X \\ 0 & \text{if } x \notin X \end{cases}$$

is total recursive (that is computable). Moreover, X is called recursively enumerable (or computably enumerable) if either $X = \emptyset$, or coincides with the set of values of a total recursive function.

Examples of recursive sets are $\mathbb{N}$, $2\mathbb{N}$, finite sets, cofinite sets; moreover. if X, Y are recursive, then also $\overline{X}$, $X \cap Y$, $X \cup Y$, $X \setminus Y$ are recursive. Note the following difference:

1. The set of numbers n for which there exists a sequence of *exactly* n occurrences of the number "7" in the expansion of π is recursively enumerable: we generate the expansion π; every time we see a string of a certain number of digits "7", we count it and add the number to the list.
2. The set of numbers n for which there exists a sequence of *at least* n occurrences of the number "7" in the expansion of π is recursive: if n is in this set, then we know that at least n occurrences of "7" in π appears and therefore every $k < n$ is in turn in this set. It follows that such a set, either is all $\mathbb{N}$, or an initial segment of it. In both cases it is recursive.

Duccio Pianigiani, University of Siena, Italy, duccio.pianigiani@unisi.it, 0000-0001-9441-7226

Referee List (DOI 10.36253/fup_referee_list)
FUP Best Practice in Scholarly Publishing (DOI 10.36253/fup_best_practice)

Book References DOI 10.36253/979-12-215-0778-2.references

The class of primitive recursive functions has several fathers and mothers, from Dedekind to Skolem to Gödel. However it was Péter (1932) who defined primitive recursive the class of functions introduced by Gödel and studied this class of functions in depth.

Definition 5. *The class of primitive recursive functions is generated by the following axioms:*

1. *(Zero)* $Z(x) = 0$
2. *(Successor)* $S(x) = x + 1$
3. *(Projections)* $P_i^{n+1}(x_0, ..., x_n) = x_i$
4. *(Composition)* $f(x_0, ..., x_m) = g(h_0(x_0, ..., x_m), ..., h_n(x_0, ..., x_m))$, *where* $g, h_0, ..., h_n$ *are primitive recursive.*
5. *(Primitive recursion)*

   (a) $f(x_0, ..., x_m, 0) = g(x_0, ..., x_m)$

   (b) $f(x_0, ..., x_m, n + 1) = h(x_0, ..., x_m, n, f(x_0, ..., x_m, n))$

   where g, h *are primitive recursive.*

Many of the most familiar functions are primitive recursive. For example the famous Fibonacci function:

1. $f(0) = 0$
2. $f(1) = 1$
3. $f(n + 2) = f(n + 1) + f(n)$

Although at first glance it does not seem (to step 3. we have used two previous values of f, and not one), it is primitive recursive: the recursion scheme "on the course of values" used there, does not get out of primitive recursive functions.

Often are thus considered most complicated scheme of recursion, that however does not get out of primitive recursive functions. Strange as it may seem, the following scheme of double recursion is primitive recursive:

1. $\phi(0, n) = g(n)$
2. $\phi(m + 1, 0) = h(m)$
3. $\phi(m + 1, n + 1) = \psi(m, n, \phi(m, \gamma(m, n)), \phi(m + 1, n))$

where $g(x)$, $h(x)$, $\psi(x, y, u, v)$ and $\gamma(u, v)$ are given primitive recursive functions. However, note that in the right side of the definition, at point 3 in the first occurrence of $\phi(x, y)$ has been substituted in place of y a function $\gamma(u, v)$ given at the beginning. Compare this case with the next case (Ackermann-Péter function). The scheme ("nested recursion") applied in the definition of this function is not reducible to the primitive recursion (note that in the third clause of the definition, the A occurs twice, in two nested occurrences):

1. $A(m, 0) = m + 1$
2. $A(0, n + 1) = A(1, n)$
3. $A(m + 1, n + 1) = A(A(m, n + 1), n)$

For instance, $A(0, 1) = 1$, $A(3, 3) = 61$, $A(4, 4) = 2^{2^{2^{65536}}}$ ... The function $A(x, y)$ majorizes each primitive recursive function, i.e., for each f primitive recursive function, there exists a number n such that: $f(x_0, ..., x_k) < A(max\{x_0, ..., x_k\}, n)$. It follows that the Ackermann function is not primitive recursive: in fact, if it were, it would be primitive recursive also $f(n) = A(n, n) + 1$; therefore, there would exist a k such that $f(m) < A(m, k)$, hence $A(k, k) + 1 = f(k) < A(k, k)$, a contradiction. After the discovery of this function, a problem arose for Hilbert's school: is Ackermann's function finitist?

> We may however admit certain extensions of the scheme of recursion as well as the induction schema, without taking away what is characteristic of the method of recursive number theory (Hilbert and Bernays (1934) vol. I, p. 325).

> The nested recursion...appear to me to be finite in the same sense as primitive recursion, i.e. of one regards them as a statement of a computation procedure where one can recognize that the function defined by the respective process satisfies the recursion equations (Bernays (1970))

Subsequently, important logicians as Simpson and Tait have proposed to identify the finitary mathematics with Skolem's *Primitive Recursive Arithmetic* PRA. Kreisel proposed instead the identification of finitist functions with those provably total in PA. The detailed historical analysis conducted by Zach (1998) showed how Hilbert and Bernays regarded finitist arithmetic as partially but not necessarily completely formalised by primitive recursive arithmetic.

Here some example of primitive recursive functions.

1. *Sum.*
   (a) $x + 0 = x$
   (b) $x + (S(y)) = S(x + y)$
2. *Predecessor.*
   (a) $P(0) = 0$
   (b) $P(y + 1) = y$
3. *Truncated difference.*
   (a) $x \dot{-} 0 = x$
   (b) $x \dot{-} (S(y)) = P(x \dot{-} y)$
4. *Multiplication.*
   (a) $x \cdot 0 = 0$
   (b) $x \cdot (S(y)) = (x \cdot y) + x$
5. *Exponentiation.*
   (a) $x^0 = 1$
   (b) $x^{y+1} = x^y \cdot x$
6. *Absolute value.* $|x - y| = (x \dot{-} y) + (y \dot{-} x)$
7. *Factorial.*
   (a) $0! = 1$
   (b) $(y + 1)! = y! \cdot (y + 1)$
8. *Minimum* $min\{x, y\} = x \dot{-} (x \dot{-} y)$ and *Maximum* $max\{x, y\} = x + (y \dot{-} x)$.
9. *Sign functions*:
   (a) $sg(0) = 0$
   (b) $sg(x + 1) = 1$
   $\overline{sg}(x) = 1 - sg(x)$.

1. *Remainder* $Rem(x, y)$ "the remainder of the divisione of y by x":
   (a) $Rem(x, 0) = 0$
   (b) $Rem(x, y + 1) = S(Rem(x, y) \cdot sg(|y - S(Rem(x, y))|)$
2. *Quotient* $qt(x, y)$ ="quotient of the divisione of y by x":
   (a) $qt(x, 0) = 0$
   (b) $qt(x, y + 1) = qt(x, y) + \overline{sg}(|x - S(Rem(x, y))|)$
3. *Limited sum*:
   (a) $\sum_{y \leq 0} f(x, y) = f(x, 0)$
   (b) $\sum_{y \leq z+1} f(x, y) = \sum_{y \leq z} f(x, y) + f(x, z + 1)$
4. *Limited product* : $\Pi_{y \leq x} f(x, y)$ (analogous).

A relation is primitive recursive, if its characteristic function is. Let $\chi(\phi) = 1$ iff ϕ is true. Some examples of such relations are the following:

1. *Characteristic function of equality*: $\chi_{=}(x, y) = \overline{sg}(|x - y|)$
2. *Characteristic function of the relation* $<$: $\chi_{<}(x, y) = sg(y \dot{-} x)$
3. The relations obtained from the primitive recursive relations by means of connectives and *bounded quantifiers* in turn primitives recursive:
   (a) *Propositional connectives*: $\chi(\neg\phi) = 1 - \chi(\phi)$, $\chi(\phi \wedge \psi) = \chi(\phi) \cdot \chi(\psi)$.
   (b) *Bounded quantifiers*: $\chi(\forall y \leq x \theta(x, y)) = \Pi_{y \leq x} \chi(\theta)(x, y)$
   (c) *Exercise.* Define the characteristic function of $\vee$ and of the bounded existential quantifier.

4. *Divisibility* ("x divides y"): $x|y \leftrightarrow \exists z \leq y(x \cdot z = y)$
5. *Prime numbers* ("x is a prime number"):

$$x \geq 2 \wedge \forall y \leq x(y|x \rightarrow y = 1 \vee y = x)$$

6. *Operator of bounded minimization* ("the minumum y less or equal to x such that..."):

$$\mu y \leq z.R(x,y) = \begin{cases} min\ y.R(x,y) & \text{if } \exists y \leq z R(x,y) \\ 0 & \text{otherwise} \end{cases}$$

where $R(x,y)$ is primitive recursive.
Observe that: $\mu y \leq z.R(x,y) = \sum_{y \leq z}(y \cdot g(x,y))$, where:

$$g(x,y) = \begin{cases} 1 & \text{if } R(x,y) \wedge \forall z < y \neg R(x,z) \\ 0 & \text{otherwise} \end{cases}$$

7. *Definition by cases.* Let g_0, g_1, h be primitive recursive. Then also the following is primitive recursive:

$$f(x) = \begin{cases} g_0(x) & \text{if } h(x) = 0 \\ g_1(x) & \text{otherwise} \end{cases}$$

Just take $f(x) = g_0(x) \cdot \overline{sg}(h(x)) + g_1(x) \cdot sg(h(x))$.
8. *The sequence of prime numbers*, where $p(x) = p_x =$ the $x - th$ prime in increasing order:
   (a) $p_0 = 2$
   (b) $p_{x+1} = \mu y \leq p_x! + 1$("$y\ prime \wedge y > p_x$")

The bound to the minimization operator is essential, and is obtained from the proof of Euclid's theorem on the infinity of the prime numbers.

To encode finite sequences of numbers *by single numbers*, we will apply this well-known result:

> (*Fundamental theorem of arithmetic* or unique prime-factorization theorem). Every integer greater than 1 either is prime itself or is the product of prime numbers, and that this product is unique, up to the order of the factors.

There are actually many methods to encode the finite sequences in primitive recursive way; a possible coding is as follows (see Odifreddi (1989-1999), Vol. I, pp. 88-90) and exploits the notions and results just introduced:

1. *Sequence.* $x = \langle x_1, ..., x_n \rangle = p_0^n \cdot p_1^{x_1} \cdot ... \cdot p_n^{x_n}$. Note that the first exponent n, gives the length of the sequence.
2. *Projection.* $(x)_i = exp(x, p_i) = \mu v \leq x(p_i^v | x \wedge \neg(p_i^{v+1}|x))$
3. *Length.* $lh(x) = (x)_0$
4. $Seq(x)$="is a sequence". $Seq(x) \leftrightarrow \forall i \leq x(i > 0 \wedge (x)_i \neq 0 \rightarrow i \leq lh(x))$
5. *Concatenation.* If $Seq(x)$ and $Seq(y)$, then: $x*y = p_0^{lh(x)+lh(y)} \cdot \Pi_{i<lh(x)} p_{i+1}^{(x)_{i+1}} \cdot \Pi_{i<lh(y)} p_{lh(x)+i+1}^{(y)_{i+1}}$. Otherwise $x * y = 0$.

In a similar way we can encode the notion of initial segment x of a sequence y (*Exercise*). Notice that $Seq(x)$ says that the code of a finite sequence has the form:

$$x = p_0^n \cdot p_1^{x_1} \cdot ... \cdot p_n^{x_n} \cdot p_{n+1}^0 \cdot p_{n+2}^0 \cdot p_{n+3}^0 \cdot ...$$

where $n = lh(x)$ says that the sequence ends at x_n and all non-zero exponents (i.e. the elements of the sequence) occur among $x_1, ..., x_n$.

We now have the tools to prove (Péter) the closure of the set of primitive recursive functions under *recursion on the course of the values.*

Theorem 7. *The following schema of recursion on the course of the values is primitive recursive:*

1. $F(0,y) = g(y)$
2. $F(x+1,y) = h(\overrightarrow{F}(x,y),x,y)$

Proof. Let us show that the history of F, namely $\overrightarrow{F}(x,y) = \langle F(0,y),...,F(x,y)\rangle$, is primitive recursive, for g,h primitive recursive. It follows that also the recursion on the course of values is primitive recursive: In fact just define $\overrightarrow{F}$ as follows:

1. $\overrightarrow{F}(0,y) = \langle g(y)\rangle$
2. $\overrightarrow{F}(x+1,y) = \overrightarrow{F}(x,y) * < h(\overrightarrow{F}(x,y),x,y) >$

Therefore $F(x,y) = (\overrightarrow{F}(x,y))_{x+1}$.

QED

Recall that the primitive recursive functions are total. Not only, but we can say more about their totality.

Definition 6. *Let $f(x_0,...,x_k)$ be a total function and T an extension of Robinson arithmetic Q. We say that $f(x_0,...,x_k)$ is provably total in T, if there is a formula $\psi(x_0,...,x_k,y)$ such that:*

1. *$\psi(\overline{n_0},...,\overline{n_k},\overline{f(n_0,...,n_k)}))$ is true, for all numbers $n_0,...,n_k$ (where $\overline{n}$ is the numeral $\overbrace{SSS...S}^{n-times}\overline{0}$ denoting n).*
2. *$\mathsf{T} \vdash \forall x_0...\forall x_k \forall y \forall z(\psi(x_0,...,x_k,y) \wedge \psi(x_0,...,x_k,z) \rightarrow y = z)$.*
3. *$\mathsf{T} \vdash \forall x_0...\forall x_k \exists y \psi(x_0,...,x_k,y)$*

Theorem 8. (Parsons 1970) *Let $I\Sigma_1^0$ be the subtheory of Peano Arithmetic, obtained by restricting the induction schema to the Σ_1^0 formulas. Hence $f(x_0,...,x_k)$ is primitive recursive if and only iff it is provably total in $I\Sigma_1^0$.*

Proof. In Parsons (1970). See also Buss (1998), where this theorem is proved using the "witnessing" method, which we will illustrate in ch.7.4. QED

We have seen that not all computable total functions are primitive recursive. However, Ackermann's function, for instance, can be obtained by further expanding the class of recursive functions with the *regular minimization scheme*:

$$f(x) = \mu y.(g(x,y) = 0)$$

as long as g is total computable and $\forall x \exists y(g(x,y) = 0)$. Note that without this clause we go out from total functions. We calculate $g(x,0), g(x,1), g(x,2)...$ If a value y exists for which $g(x,y) = 0$, the computation sooner or later ends; but if we do not put this clause, the computation might never end. This is another way of incorporating induction, the minimum principle being equivalent to complete induction. Gödel and Kleene actually showed that this class, the total recursive functions, can be characterized also without the primitive recursion.

Theorem 9. *The class of total recursive functions is the smallest class containing sum, product, projections, the characteristic function of equality and is closed under composition and under μ-recursion.*

We have so exhausted the notion of computability? Not really. We aim at a formalization of computable functions such that:

1. we can give an effective list $f_0, f_1, f_2, ...$ of their programs, containing all and only the programs of the computable functions

2. Such a list must be *uniformly effective*, in the sense that there is an algorithm Φ such that for all e, n:

$$\Phi(e, n) = f_e(n)$$

Now we will see that this cannot be done, if we limit ourselves to *total* functions, because of the phenomenon of *diagonalization*: for total recursive functions there is no *universal function* Φ as above. This method is applicable to any case where the sets of instructions can be effectively listed. Suppose that the above claims are satisfied by the total computable functions and take $h(x) = \Phi(x, x) + 1$. Also h is total, hence corresponds to some f_e in the list. But then we have a contradiction, since $f_e(e) = h(e) = \Phi(e, e) + 1 = f_e(e) + 1$. Hence h cannot be a member of the above list, which cannot be therefore exaustive. We must therefore abandon one of the above conditions. The problem does not arise if, while maintaining conditions 1. and 2., we admit partial functions. Let us admit therefore partial functions, i.e. that are not defined on some numbers. With the notation $\phi(x) \downarrow$ we mean that the function is defined on x, while the notation $\phi(x) \uparrow$ mean that the function is not defined at x. With writing $\phi(x) \simeq \psi(x)$ we understand now that either $\phi(x) \uparrow$ and $\psi(x) \uparrow$, or $\phi(x) \downarrow$ and $\psi(x) \downarrow$ and $\phi(x) = \psi(x)$. The minimization scheme is now as follows. If $\phi(x, y)$ is partial recursive:

$$\psi(x) \simeq \mu y(\phi(x, y) \simeq 0 \wedge \forall z \leq y \phi(x, z) \downarrow)$$

It is undefined if there is no such y. Adding this scheme to the primitive recursive functions (where equality has to be intended between partial functions) we have the model of Kleene's $\mu-$recursive functions[1].

The above definition can also be reformulated as follows: the partial recursive functions can be obtained by adding this minimization scheme, rather than what we added before:

$$\phi(x_0, ..., x_n) \simeq \mu y.R(x_0, ..., x_n, y)$$

where $R(x_0, ..., x_n, y)$ is a recursive relation and $\phi(x_0, ..., x_n) \uparrow$ if does not exist y satisfying $R(x_0, ..., x_n, y)$. The partial recursive functions are not closed with respect to this scheme, if $R(x_0, ..., x_n, y)$ is only recursively enumerable. Actually, if $R(x, y)$ is recursive, we can define

$$\phi(x_0, ..., x_n) \simeq \mu y.((\overline{sg}\chi_R(x_0, ..., x_n, y)) \simeq 0)$$

Theorem 10. (Kleene 1936) *Let $n > 0$; there are primitive recursive relations $T^n(x, x_0, ..., x_{n-1}, z)$ and $U(z)$ such that, for each recursive n-ary partial function ψ, there exists $e \in \mathbb{N}$ such that:*

$$\psi(x_0, ..., x_{n-1}) \simeq U(\mu z.T^n(e, x_0, ..., x_{n-1}, z))$$

$$\psi(x_0, ..., x_{n-1}) \downarrow \text{ iff } \exists z(T^n(e, x_0, ..., x_{n-1}, z))$$

The relation $T^n(e, x_0, ..., x_{n-1}, z)$ must be read as: "z encodes a computation of the function whose code is e on input $x_0, ..., x_{n-1}$".

Proof. The proof is very laborious and we refer to Odifreddi (1989-1999) pp. 90-6 for a detailed account. The basic idea is as follows: first of all, we associate numbers to the initial functions and to all scheme generating the recursive functions, encoding them *à la* Gödel. Then we define a computation tree, whose nodes are labelled by triples $\langle e, \langle x_0, ..., x_n \rangle, b \rangle$ where e is the code of a function f, the sequence $\langle x_0, ..., x_n \rangle$ represents its argument and b is the value of $f(x_0, ..., x_n)$. Internal nodes tell us inductively through which scheme, applied to the function that labels the children of a node, we arrived at the function that labels that node. Leaves correspond to initial functions and the root is labelled by the triple corresponding to the function whose computation

[1] In Odifreddi (1989-1999) p. 128 is made clear that the set of partial recursive functions *is not closed* under the simplified scheme $\phi(x) \simeq \mu y.(\psi(x, y) \simeq 0)$.

tree we are defining. The whole tree will be encoded like this: suppose that the node k has $n+1$ children, which are the roots of subtrees $T_0, ..., T_n$; then the tree generated by k has code $z = \langle k, \ulcorner T_0 \urcorner, ..., \ulcorner T_n \urcorner \rangle$. Finally, we define $T^n(e, x_0, ..., x_{n-1}, z)$ iff e encodes $\psi(x_0, ..., x_{n-1})$ and z encodes the computation tree whose root is labelled by $k = \langle \ulcorner \psi \urcorner, \langle x_0, ..., x_{n-1} \rangle, b \rangle$. With the use of the projections also define $U(z) = b$. Indeed, note that $(z)_1 = k$; but $k = \langle \ulcorner \psi \urcorner, x_0, ..., x_{n-1}, u \rangle$, hence $((z)_1)_1 = \ulcorner \psi \urcorner$, $((z)_1)_2 = \langle x_0, ..., x_{n-1} \rangle$ and so on. To translate this representation into a primitive recursive predicate that says that a given number encodes a computation tree, we in fact only need primitive recursive predicates and functions, mainly codes of sequences and projections. QED

We have in fact obtained the *enumeration theorem*: for each n, there exists a *universal function* concerning the class of partial recursive functions in n variables:

$$K^n(e, x_0, ..., x_{n-1}) \simeq U(\mu z T^n(e, x_0, ..., x_{n-1}, z))$$

Moreover, if we define $\phi_e^n(x_0, ..., x_{n-1}) \simeq K^n(e, x_0, ..., x_{n-1})$, then for all e each ϕ_e^n is partial recursive and each $n-ary$ partial recursive function correspond to some ϕ_e^n in the list. But we can also omit the reference to n-arity and see from the above that there is an *universal function* $K(u, x)$ that generates all partial functions, that is, such that, for each arity and for each partial recursive function $\phi(x_0, ..., x_{n-1})$, exists a code e for which we have:

$$\phi(x_0, ..., x_{n-1}) \simeq K(e, \langle x_0, ..., x_{n-1} \rangle)$$

Observes that $T^n(e, x_0, ..., x_{n-1}, y)$ implies $((y)_1)_1 = e$ and $((y)_1)_2 = \langle x_0, ..., x_{n-1} \rangle$. Then we place:

$$K(e, x) \simeq U(\mu y.T(y) \wedge ((y)_1)_1 = e \wedge ((y)_1)_2 = x)$$

where $T(y)$ is the primitive recursive predicate formalizing " y is a computation tree ", according to the previous construction. It has thus:

$$K(e, \langle x_0, ..., x_{n-1} \rangle) \simeq U(\mu y.T(y) \wedge ((y)_1)_1 = e \wedge ((y)_1)_2 = \langle x_0, ..., x_{n-1} \rangle)$$
$$\simeq \phi^n(x_0, ..., x_{n-1})$$

We write $\phi_e^n(x_0, ..., x_{n-1})$ for $K(e, \langle x_0, ..., x_{n-1} \rangle)$.

A universal Turing machine is essentially a machine that computes a similar function. It is also concluded from the above that there are exactly $\aleph_0$ *partial* recursive functions and exactly $\aleph_0$ *total* recursive functions: by the Church's thesis constant functions $c_n(x) \simeq n$ are total recursive and therefore there are at least $\aleph_0$ *total* recursive functions. Moreover, they cannot be more than $\aleph_0$, by the normal form theorem. It is also known that, being the set of functions (computable or not) from $\mathbb{N}$ to $\mathbb{N}$ of cardinality $2^{\aleph_0}$. This means that the computable functions are a small minority. We must make a few remarks about the *indices*. First of all each recursive partial function has infinite indices; for instance:

$$\psi_e(x) \simeq \psi_e(x) + Z(P_0^m(x))$$

And yet the indices of the two functions are different. The index for $\psi_e(x) + Z(P_0^m(x))$ is given by the sequence code $\langle 3, e, \langle 3, \langle 2, m, 1 \rangle, \langle 0 \rangle \rangle, e^+ \rangle$, where e^+ is the index of $+$.

Theorem 11. ("Padding lemma") *There is a total function $f(e, m, w)$ such that $\phi_e(x) \simeq \phi_{f(e,m,n)}(x)$, where $f(e, m, 0) < f(e, m, 1) < f(e, m, 2) < ...$ are all indexes of the same function.*

Proof. If $z = \langle 3, \langle 2, m, 1 \rangle, \langle 0 \rangle \rangle = \ulcorner Z(P_0^m(x)) \urcorner$,

let $g(e, m) = \langle 3, e, z, e^+ \rangle = \ulcorner \psi_e(x) + Z(P_0^m(x)) \urcorner$; observe that $\phi_e(x) \simeq \phi_{g(e,m)}(x)$. Let therefore $f(e, m, 0) = e$ and $f(e, m, n+1) = g(f(e, m, n), m)$.

QED

Theorem 12. (Theorem of parameters, or *s-m-n*-theorem) *Given $m, n \in \mathbb{N}$ there exists a function $s(e, x_0, ..., x_{n-1})$ such that:*

$$\phi_e(x_0, ..., x_{n-1}, w_0, ..., w_{m-1}) \simeq \phi_{s(e,x_0,...,x_{n-1})}(w_0, ..., w_{m-1})$$

Proof. We want constant functions $\phi_{f(0)}(w_0, ..., w_{m-1}) = 0$, $\phi_{f(1)}(w_0, ..., w_{m-1}) = 1$ etc. Thus, let us define:

1. $f(0) = \langle 3, \langle 2, m, 1\rangle, \langle 0\rangle\rangle$
2. $f(n+1) = \langle 3, f(n), \langle 1\rangle\rangle$

Then observe that:

1. $\phi_{f(0)}(w_0, ..., w_{m-1}) = Z(P_1^m(w_0, ..., w_{m-1}))$
2. $\phi_{f(n)}(w_0, ..., w_{m-1}) = \overbrace{SSS...S}^{n-times}(Z(P_1^m(w_0, ..., w_{m-1})) = n$

Hence, abbreviating $w = w_0, ..., w_{m-1}$ and $x = x_0, ..., x_{n-1}$:

$$\phi_e(x, w) \simeq \phi_e(\phi_{f(x_0)}(w), ..., \phi_{f(x_{n-1})}(w), P_1^m(w), ..., P_m^m(w))$$

Let therefore:

$$s(e, x_0, ..., x_{n-1}) = \langle 3, f(x_0), ..., f(x_{n-1}), \langle 2, m, 0\rangle, ..., \langle 2, m, m-1\rangle, e\rangle$$

And therefore we have $\phi_e(x, w) \simeq \phi_{s(e,x)}(w)$ QED

Many results that have been proved so far (enumeration theorem, *s-m-n* theorem, padding lemma) mention indices in their statements or in their proofs: do these results depend on the particular way in which indices were defined? We call acceptable a system that resemble the standard one, i.e. if it is possible to go effectively from the one to the other. More precisely, if ψ_e is the standard one, then a system of indices ζ_e is acceptable if there are computable functions g, f such that $\psi_e = \zeta_{f(e)}$ and $\psi_{g(e)} = \zeta_e$. In this respect we have this result:

Theorem 13. *An enumeration of the partial recursive functions $\{\phi_e\}_{e\in\mathbb{N}}$ is an accettable indexing if and only if:*

1. *for every partial recursive function ψ exists an index e such that $\psi \simeq \phi_e$.*
2. *There is a universal function.*
3. *The parameter theorem is fulfilled..*

It follows that every acceptable system of indices satisfies the *Recursion Theorem.*

Theorem 14. (Recursion Theorem) *Let f be total recursive; then there exists $n \in \mathbb{N}$ such that $\phi_n(x) \simeq \phi_{f(n)}(x)$, namely n and $f(n)$ compute the same function: we say that n is a fixed point of f.*

Proof. Let $u \in \mathbb{N}$ and let $\phi(x)$ be defined by the following instructions: apply the set of instructions P_u coded by u, to the input u; if the computation terminates and gives output w, then take the set P_w of the instructions encoded by w and apply it to x. If P_w applied to x halts and returns output z, put $\phi(x) \simeq z$. More formally, whereas the instructions to ϕ depends on u, let g the function that, given u, returns the code for these instructions:

$$\phi_{g(u)}(x) \simeq \begin{cases} \phi_{\phi_u(u)}(x) & \text{if } \phi_u(u)\downarrow \\ \uparrow & \text{otherwise} \end{cases}$$

Let $\phi_e(x) = f(g(x))$ and $n = g(e)$; then $\phi_n \simeq \phi_{g(e)} \simeq \phi_{\phi_e(e)} \simeq \phi_{f(g(e))} \simeq \phi_{f(n)}$ QED

Example. There is an index e such that for all n, $\phi_e(n) \simeq e$, that is, this function prints its name. Let $f(x, y) = x$; by the $s-m-n$ theorem there exists a function g such that $f(x, y) \simeq \phi_{g(x)}(y) = x$. For the theorem of fixed point exists n such that $\phi_{g(n)}(x) \simeq \phi_n(x)$. Thus:

$$n = f(n, y) \simeq \phi_{g(n)}(y) \simeq \phi_n(y)$$

Example. We can use the theorem of fixed point and the $s - m - n$ theorem to prove the Ackermann function. Let $K(e, x)$ a universal function and define by cases:

1. $g(n, 0, y) \simeq y + 1$
2. $g(n, x + 1, 0) \simeq K(n, \langle x, 1\rangle)$
3. $g(n, x + 1, y + 1) \simeq K(n, \langle x, K(n, \langle x + 1, y\rangle)\rangle)$

For the fixed point and for s-m-n, we find an n such that:

$$g(n, x, y) \simeq \phi_{h(n)}(x, y) \simeq \phi_n(x, y)$$

Thus:

1. $\phi_n(0, y) \simeq \phi_{h(n)}(0, y) \simeq g(n, 0, y) = y + 1$
2. $\phi_n(x + 1, 0) \simeq \phi_{h(n)}(x + 1, 0) \simeq g(n, x + 1, 0) \simeq K(n, \langle x, 1\rangle) \simeq \phi_n(x, 1)$
3. $\phi_n(x + 1, y + 1) \simeq \phi_{h(n)}(x + 1, y + 1) \simeq g(n, x + 1, y + 1) \simeq$
   $\simeq K(n, \langle x, K(n, \langle x + 1, y\rangle)\rangle) \simeq \phi_n(x, \phi_n(x + 1, y))$

As to the first example, functions that prints their own code as in that case are called *quines*. A *quine* is an easy example self-replicating computer program which takes no input and produces a copy of its own source code as its only output. The name "quine" was coined by Douglas Hofstadter, in his popular science book Hofstadter (1979), in honor of the philosopher Willard Van Orman Quine (1908-2000). There is a famous anecdote that when Descartes presented Christina of Sweden with the hypothesis that animals constitute a form of mechanical automaton, she pointed to a clock and exclaimed: "Let's see if it produces a child". Organisms, unlike machines, are self-organising systems that self- reproduce. The view of animals as machines (la *bête machine*), in other words, ran into serious difficulties in the face of the objection that living organisms, in general, unlike machines (as has long been assumed), have the capacity to reproduce themselves. Speaking more abstractly, self-replication is any behaviour of a dynamic system that leads to the construction of a copy of itself. Since the time of Descartes and La Mettrie, many things (starting with the concept of "machine" itself) have changed, but the opposition between vitalistic, organicistic and mechanistic conceptions of life is still strong. Certain mechanistic tenets conceive of living organisms as very complex machines programmed by genetic software, just as, conversely, attempts are made to describe living organisms in terms of self-replicating automatons. Anti-mechanistic conceptions instead emphasise the irreducible diversity between organisms and machines. The fixed point theorem and the diagonalization technique shed light on this problem (see also Rogers (1987) pp. 188-90 and Odifreddi (1989-1999) pp. 170-74 for a more in-depth discussion) and thanks to them it can be demonstrated that there must be a machine that, regardless of the input, constructs its own replica. The problem of self-replication of machines began to become a concrete and "engineering" problem with Von Neumann, who formalized the idea of cellular automata in order to create a theoretical model for a self-reproducing machine. He drew a general outline of his self-replicating automaton that anticipated some concepts of current cell biology, such as those of translation and transcription (the two fundamental stages of the protein synthesis process). In fact, the insights behind Von Neumann's original model anticipated some of Watson and Crick's (1953) discoveries concerning the functioning of DNA. At the Hixon Symposium in Pasadena, California on September 1948 he compared the functions of genes to self-reproducing automata; its universal constructor is a self-replicating machine in a cellular automata (CA) environment. Cellular automata are mathematical models he used to simulate complex systems representing real-world phenomena studied in physics and biology. In this research he made use of some concepts introduced by Turing, for example the Universal constructor, that is, able to build any machine starting from a its description, and in particular to self-reproduce, a concept inspired by the Universal Turing Machine (see Von Neumann (1951)).

2.2. Recursively enumerable sets

Informally, a set is computably enumerable (o recursively enumerable), if there is a computable procedure to list its elements.

Definition 7. *A set $X \subseteq \mathbb{N}$ is recursively enumerable (or computably enumerable) iff either $X = \emptyset$, or $X = Range(f) = \{f(0), f(1), f(2), ...\}$, for some f total recursive.*

Note that if X is recursive, then it is also recursively enumerable; in case of $X = \emptyset$, it is obvious; in case of X *finite*, e.g. $X = \{k_0, ..., k_n\}$, take f total recursive defined as follows:

$$f(x) \simeq \begin{cases} k_x & \text{if } x \leq n \\ k_n & \text{otherwise} \end{cases}$$

It is clear that X is the range of f. If X is *infinite* and χ_X is the characteristic function:

1. $f(0) \simeq \mu y.(\chi_X(y) = 0)$
2. $f(n+1) \simeq \mu y.(\chi_X(y) = 0 \wedge f(n) < y)$

Note that $Range(f) = X$.

Lemma 1. *The following are equivalent:*

1. *Either $A = \emptyset$ or $A = Range(f)$ for some f total recursive.*
2. *$A = Dom(\psi)$, for ψ partial recursive.*
3. *$A = Range(\phi)$, for ϕ partial recursive.*

Proof. 1. $\Rightarrow$ 2. if $A = \emptyset$, then $A = Dom(\phi)$, where ϕ is the function everywhere divergent. If $A \neq \emptyset$, $A = Range(f)$, then define ψ using the following instructions:

1. generate $Range(f)$

2. when y appears, put $\psi(y) \simeq y$

So we get $Dom(\psi) = Range(f) = A$.
2. $\Rightarrow$ 3. Let $A = Dom(\psi)$, let moreover $\phi(x) \simeq x + 0 \cdot \psi(x)$; note that $A = Range(\phi)$.
3. $\Rightarrow$ 1. Let $A = Range(\phi)$, let therefore:

$$f(\langle x, t \rangle) \simeq \begin{cases} z & \text{if } \phi(x) \downarrow \text{ in at most } t \text{ steps with output } z \\ a \in A & \text{otherwise} \end{cases}$$

Then $A = Range(f)$. QED

Remark 2. *Note that if we had omitted the bound t we would not have guaranteed the total character of f; the reason is that we can not recursively decide whether $\phi(x) \downarrow$ (unsolvability of the "halting problem"), but on the contrary we can decide whether $\phi(x) \downarrow$ in at most t steps. Moreover, in 1. the function f can be taken primitive recursive: suppose that $A \neq \emptyset$; if $f(x) = \phi_e(x) = y$, then $\exists s T(e, x, s) \wedge U(s) = y$. So, define $g(z) = U((z)_1)$, if $T(e, (z)_0, (z)_1)$, and an element $a \in A$ otherwise. Show that the range of g is A.*

There is another interesting characterization of the computably enumerable sets which will come in handy in sections 2.4 and 8.3.

Theorem 15. *The following are equivalent:*

1. *A is computably enumerable.*

2. *There is a total computable binary function $f(x, s)$ (i.e. with values $0, 1$) such that for every x, $f(x, 0) = 0$, there is at most one s such that $f(x, s+1) \neq f(x, s)$, and $\lim_s f(x, s) = \chi_A(x)$ =characteristic function of A.*

3. *There is a computable sequence of finite sets $A_s, s \in \mathbb{N}$, such that for all s, $A_s \subseteq A_{s+1}$, and $A = \cup_s A_s$.*

Lemma 2. *X is recursive iff X and $\overline{X}$ are both computably enumerable*

Proof. $\Rightarrow$ Let $\chi_X(x)$ be te characteristic funcion of X (by hypothesis computable) and let us consider the following:

$$\phi(x) \simeq \begin{cases} 0 & \text{if } \chi_X(x) = 0 \\ \uparrow & \text{otherwise} \end{cases}$$

$$\psi(x) \simeq \begin{cases} 1 & \text{if } \chi_X(x) = 1 \\ \uparrow & \text{otherwise} \end{cases}$$

and observe that $X = Dom(\psi)$ and $\overline{X} = Dom(\phi)$.

$\Leftarrow$ If X and $\overline{X}$ are computably enumerable not empty, are then generated respectively by f, g total recursive, i.e. $X = \{f(0), f(1), f(2)...\}$, $\overline{X} = \{g(0), g(1), g(2), ...\}$. To find out whether a number belongs to X just generate both sets. QED

It is worth pointing out the difference with this result:

Lemma 3. *The following statements are equivalent:*

1. *X is recursive.*
2. *Either $X = \emptyset$ or X is the range of a non decreasing recursive function f (i.e. if $a > b$ then $f(b) \geq f(a)$)*

Proof. $\Rightarrow$ Let $X \neq \emptyset$ recursive; let a its smallest element and let $f(0) = a$,

$$f(n+1) \simeq \begin{cases} n+1 & \text{if } n+1 \in X \\ f(n) & \text{otherwise} \end{cases}$$

Note that f is non-decreasing.

$\Leftarrow$ Let X infinite and range of f non decreasing recursive; to know if $x \in X$, the test is as follows: Search the minimum n such that $f(n) > x$. We will have that $x \in X$ iff $x \in \{f(0), ..., f(n)\}$ QED

Definition 8. *With notation W_e we mean the computably enumerable set $Dom(\phi_e)$. Hence every computably enumerable set can be written in this form.*

Moreover, from the normal form theorem:

$$W_e = \{\langle x_0, ..., x_{n-1}\rangle | \exists y T(e, x_0, ..., x_{n-1}, y)\}$$

It follows that X is computably enumerable iff there exist a recursive relation R such that:

$$X = \{\langle x_0, ..., x_{n-1}\rangle | \exists v R(x_0, ..., x_{n-1}, v)\}$$

A direction follows from the above; for the other direction, suppose:

$$X = \{\langle x_0, ..., x_{n-1}\rangle | \exists v R(x_0, ..., x_{n-1}, v)\}$$

Then we take $\psi(x_0, ..., x_{n-1}) \simeq \mu z.R(x_0, ..., x_{n-1}, z)$. Note that $X = Dom(\psi)$.

We now propose again, in this different context, the fundamental result about the unsolvability of the halting problem, already highlighted in the chapter on Turing machines.

Theorem 16. *The problem "$x \in W_x$" is undecidable.*

Proof. Let us suppose that the following function is computable:

$$f(x) = \begin{cases} 1 & \text{if } \phi_x(x)\downarrow \\ 0 & \phi_x(x)\uparrow \end{cases}$$

Then take:

$$g(x) = \begin{cases} 0 & \text{if } f(x) = 0 \\ \uparrow & f(x) = 1 \end{cases}$$

Observe that g is partial recursive; let $g = \phi_e$, for some e. Hence $e \in W_e$ iff $\phi_e(e)\downarrow$ iff $g(e) = 0$ iff $e \notin W_e$ (contradiction). Hence f cannot be computable. It follows that also the following is incomputable:

$$h(x,y) = \begin{cases} 1 & \text{if } \phi_x(y)\downarrow \\ 0 & \text{otherwise} \end{cases}$$

If it were, then it would also be $f(x) = h(x,x)$. Hence also the problem "$y \in W_x$" is unsolvable. QED

Corollary 1. *The following ("diagonal set" or "halting set") is computably enumerable but not recursive:*

$$K = \{n \in \mathbb{N} | \phi_n(n)\downarrow\} = \{n \in \mathbb{N} | n \in W_n\}$$

Proof. Let indeed $\psi(x) = \phi_x(x)$, and note that it converges, namely is defined, on x iff $x \in K$, i.e. $x \in Dom(\psi)$ iff $x \in K$; hence $K = Dom(\psi)$ and therefore K is computably enumerable But is not recursive: indeed its complement $\overline{K}$ is not computably enumerable Note that $x \in \overline{K}$ iff $\phi_x(x)\uparrow$ iff $x \notin W_x$. Since each computably enumerable set is of the form W_x, this means that $\overline{K}$ differs from each of them in at least one element. Hence is different from all computably enumerable sets. QED

Example. There is no effective way of deciding, given x if ϕ_x is total. It is observed that the set $Tot = \{x | \phi_x \; totale\}$ is not computably enumerable For if it were, there would be some f recursive of which would be the range:

$$\phi_{f(0)}, \phi_{f(1)}, \phi_{f(2)}, ...$$

But then we could take $g(x) = \phi_{f(x)}(x) + 1$ noting that $g \neq \phi_{f(x)}$, for all x. For if we had $g = \phi_{f(e)}$, for some e, then $g(e) = \phi_{f(e)}(e) = \phi_{f(e)}(e) + 1$

Index set. Let $\mathcal{F}$ a set of partial recursive functions; the set $\mathcal{F}^*$ of all the indices of the functions in $\mathcal{F}$, assuming that if $x \in \mathcal{F}^*$ and $\phi_x \simeq \phi_y$ then $y \in \mathcal{F}^*$, is called "index set". Note that the set Tot of the previous example is an index set, but K is not; take indeed $f(n) =$ index of $\{n\}$. For the fixed point theorem there exists e such that $W_{f(e)} = W_e$, whereby $e \in W_e$ iff $e \in W_{f(e)}$ iff $e \in \{e\}$. If now we take a different index j of $\{e\}$, then $j \notin W_j = \{e\}$, i.e. $\phi_j(j)\uparrow$ and therefore $j \notin K$. Note that an index set $\mathcal{F}$ contains all possible "programs" to calculate the functions contained in $\mathcal{F}$.

Theorem 17. (Rice theorem) *Each nontrivial property of programs is undecidable: if $\mathcal{F}$ is a class of partial recursive functions and $\mathcal{F}^*$ is its index set, then $\mathcal{F}^*$ is recursive iff either is empty, or coincides with the class of all partial recursive functions.*

Proof. Reduces K to $\mathcal{F}^*$; we exclude cases in which $\mathcal{F}^* = \emptyset$ and $\mathcal{F}^* = \mathbb{N}$; we exclude also the case everywhere undefined function $\phi_\emptyset$. Let therefore $a \in \mathcal{F}^*$ and let:

$$\psi(x,y) = \begin{cases} \phi_a(y) & \text{if } x \in K \\ \uparrow & \text{otherwise} \end{cases}$$

by parametrization we obtain $\psi(x,y) \simeq \phi_{h(x)}(y)$. Then:

1. If $x \in K$, then $\psi(x,y) = \phi_a(y)$, from which $\phi_a \simeq \phi_{h(x)}$ and $h(x) \in \mathcal{F}^*$.
2. If $x \notin K$, then $\psi(x,y) \uparrow$, from which $\phi_\emptyset \simeq \phi_{h(x)}$ and $h(x) \notin \mathcal{F}^*$

If therefore $\mathcal{F}^*$ would be recursive, also K would be. QED

An *index-set* is intended to code properties of functions, not of programs, i.e. not depending on the particular algorithm used for computing that functions; but Rice's theorem says that all nontrivial properties of index-sets are undecidable.

Lastly, an important characterisation of computably enumerable sets is that connected to the negative solution of Hilbert's X problem. A Diophantine equation is a polynomial with integer coefficients for which we seek integer solutions. In general they have the form: $f(x_0, ..., x_m) = 0$, or $f(x_0, ..., x_m) = g(x_0, ..., x_m)$ where f, g are polynomials with integer coefficients.

2.3. Hilbert's tenth problem

The tenth on the list of mathematical problems that Hilbert posed in 1900 asks: "given a Diophantine equation with any number of unknowns and integer coefficients, define a process that could end in a finite number of operations, to determine whether the equation has integers solutions" . We consider *families* of Diophantine equations $D(a_0, ..., a_k, x_0, ..., x_m) = 0$, where:

1. the variables $a_0, ..., a_k$ are called *parameters*.
2. the variables $x_0, ..., x_m$ are called *unknowns*.

By setting the parameters we obtain an equation; on the basis of the parameters, the equation may or may not have solutions for the unknowns (that is to say values which replaced the unknowns make it true). A set of the form:

$$X = \{\langle a_0, ..., a_k\rangle | \exists x_0 ... \exists x_m (D(a_0, ..., a_k, x_0, ..., x_m) = 0)\}$$

i.e. a set consisting of $k+1$-tuples of *natural numbers* $\langle a_0, ..., a_k\rangle$ for which there exist solution vectors for the unknowns, $x_0, ..., x_m$, is called *Diophantine*. Indeed, the expression "integers solutions" , il Logic books, is sometimes replaced with "solutions in natural numbers". This is due to a Lagrange's theorem which states that every non-negative integer is the sum of four squares of integers. Considers, therefore $f(x_0, ..., x_m) = 0$ and suppose that we are looking for solutions in the naturals. Then we take $f(a_0^2+b_0^2+c_0^2+d_0^2, ..., a_m^2+b_m^2+c_m^2+d_m^2) = 0$. From Lagrange's four-square theorem, this is guaranteed to be possible. Substituting $a_0^2 + b_0^2 + c_0^2 + d_0^2, ..., a_m^2 + b_m^2 + c_m^2 + d_m^2$ in place of $x_0, ..., x_m$ into $f(x_0, ..., x_m)$, we obtain a Diophantine equation in $4m$ variables, all of which are integers.

Theorem 18. *The problem of determining the existence or nonexistence of solutions to a Diophantine equation in natural numbers is reducible to the problem of determining the existence or nonexistence of solutions to a Diophantine equation with integer values, and the opposite is also true.*

In 1934 Godel shows that the undecidable statement of his 1931 theorem can be expressed in the form: $Q_0 x_0 ... Q_n x_n (f(x_0, ..., x_n) = 0)$ where f is a polynomial with integer coefficients and $Q_0, ..., Q_n$ are quantifiers. subsequently Gödel shows that every statement of the form $\forall x \phi(x)$ with $\phi(x)$ primitive recursive is equivalent to one of the form:

$$\forall x_0 ... \forall x_m \exists y_0 ... \exists y_n (f(x_0, ..., x_m, y_0, ..., y_n) = 0)$$

In 1950 Martin Davis conjecture that a set is computably enumerable iff it is Diophantine. It also shows that every computably enumerable set X has the form:

$$X = \{\langle a_0, ..., a_n\rangle | \exists z \forall y \leq z \exists x_0 ... \exists x_k f(a_0, ..., a_n, y, z, x_0, ..., x_k) = 0\}$$

It will take another two decades to eliminate the bounded quantifier "$\forall y \leq z$". In 1952 Julia Robinson shows that there is a polynomial f such that $a^b = c$ if and only if $\exists z_0 ... \exists z_m (f(a, b, c, z_0, ..., z_m) = 0)$ under a sufficient condition that a Diophantine set of pairs $X = \{\langle u, v\rangle | \exists y_0, ..., \exists y_k g(u, v, y_0, ..., y_n) = 0\}$ exists, such that:

1. if $\langle u, v\rangle \in X$, then $u < v^v$
2. for all k exists $\langle u, v\rangle \in X$ such that $u > v^k$.

If we call (JR) this hypothesis, the result is that if (JR) is true, then exponentiation is Diophantine. In 1959 Davis, Putnam and Robinson showed that each computably enumerable set is *exponential* Diophantine. Hence (JR) implies that every computably enumerable set is Diophantine. Remember that the exponential diophantine equations are $f(x_0, ..., x_m, y_0, ..., y_n) = 0$ where the f is an exponential polynomial, i.e. constructed by addition, multiplication and exponentiation, e.g.

$$(x+1)^{y+2} + x^3 = y^{(x+1)^x} + y^4$$

Lastly, in 1970 Y. Matiyasevich find the required equation (JR) demonstrating that the Fibonacci sequence $F_0 = 1$, $F_1 = 1$ and $F_n = F_{n-1} + F_{n-2}$ is definable by a set of ten Diophantine equations and that the set $X = \{\langle u, v\rangle | u = F_{2v} \wedge v > 2\}$ satisfies (JR).

Theorem 19. (Matiyasevich 1970, Robinson 1950, Davis 1949, Putnam 1959) *Each computably enumerable set is Diophantine.*

Note that for the theorem of normal form $P(x_0, ..., x_k)$ is computably enumerable iff there exist a recursive relation $R(x_0, ..., x_k)$ such that:

$$P(x_0, ..., x_k) \leftrightarrow \exists y R(x_0, ..., x_k)$$

Hence also the converse of Matiyasevich theorem holds and it follows:

Corollary 2. *Computably enumerable sets coincide with the Diophantine sets.*

Corollary 3. *Hilbert's tenth problem has a negative solution.*

Proof. The halting set K can be written in Diophantine form:

$$K = \{x | \exists y_0 ... \exists y_n (f(y_0, ..., y_n, x) = 0)\}$$

But K is not computable, then there is no algorithm for deciding whether the equations:

$$f(y_0, ..., y_n, 0) = 0, f(y_0, ..., y_n, 1) = 0, f(y_0, ..., y_n, 2) = 0...$$

have solutions in naturals: if there was a solution to the tenth Hilbert problem, such an algorithm would exist, so we could also decide K. QED

We believe it is useful to indicate some recent developments in the research on this issue. The MRDP theorem is provable in the *Elementary Aritmetic*, namely the theory obtained from $\mathsf{I}\Delta_0$ (the theory obtained from Peano Arithmetic restricting the induction to the bounded formulas) by adding $\forall x \exists y 2^x = y$ and therefore in this theory and all its extensions, each Σ_1 formula is equivalent to a diophantine one of the form:

$$\exists x (t(x, y) = s(x, y))$$

where t, s are terms of the language of *Peano Arithmetic.* By the well known essential undecidability results, for all consistent extensions of Q, the set of Σ_1 provable sentences and the set of $\Sigma_1$sentences consistent with the theory are both undecidable. Hence also the set of Diophantine formulas *provable* in *extensions of Elementary Arithmetic* and the set of those *consistent* with these theories are undecidable too. Let us call now D_T the set of Diophantine formulas consistent with the theory T. Or, in other word the set of Diophantine equations *solvable in some models* of T. We now ask: when D_T is decidable? See Jeřábek (2016) for some positive and negative answers.

2.4. Creative sets and productive sets

According to Gödel, truth in the standard model cannot coincide with provability, because, as he states in a unsent letter to Yossef Balas (see Dawson (1997), p. 61):

> It follows from the correct solution of the semantic paradoxes that 'truth' of the propositions of a language cannot be expressed in the same language, while provability (being an arithmetical relation) can.

However, he adds:

> in consequence of the philosophical prejudices of our time [...] a concept of mathematical truth as opposed to demonstrability was viewed with greatest suspicion and widely rejected as meaningless.

Truth cannot be expressed because otherwise the liar paradox would be reproducible: a version of Tarski's theorem before Tarski (see Krajewski (2004) for a general analysis on the relationship between Gödel and Tarski). Once we know that provability is definable and truth is not, then assuming that provable sentences are true, i.e. that $Pr \subseteq Tr$, we get $Pr \subset Tr$ and we conclude that there must be some undecidable sentence, only we have no concrete example (see Grattan-Guinness (1979)). So, why Gödel didn't publish the indirect proof of incompleteness? The explanation is seen in the fact that Gödel was unusually cautious. He feared that relying on the concept of truth would compromise the possibility of acceptance of his results from the scientific world, because it would cause suspicion on the part of Hilbert, and of scientific and philosophical context dominated by finitists, logical positivists, formalists. In a letter to Hao Wang (see Wang (1974), p. 6), Gödel expresses his doubts:

> formalists considered formal demonstrability to be an analysis of the concept of mathematical truth and, therefore, were of course not in a position to distinguish the two.

Mathematical truth is just provability: this was the opinion of most matematicians in that time. Due to his cautious attitude, in his famous paper of 1931 Gödel mentions "truth" only in the introductory section, and the proof itself is made without the notion of truth. Truth was suspicious, provability wasn't.

The aim of this section is to account for Gödel's original intuition with the means of modern computability theory, showing the different complexity of the theorems of a theory of formal arithmetic with certain properties, and of the set of true propositions of the language of this theory. For this purpose, we refine the notion of computably enumerable set through the concept of *creative set*, introduced by Emil Post:

> The conclusion is unescapable that even for such a fixed, well defined body of mathematical propositions, mathematical thinking is, and must remain, essentially creative (Post (1944), 295).

Hence this terminology emphasises a consequence of Gödel's incompleteness theorems, namely the proof of an inherent creativity of the human mind.

Creative sets are those computably enumerable sets whose complement fails to be computably enumerable in a rather strong way. In some sense are those which can be shown effectively to be incomputable. Dekker's notion of *productivness*, that we will use in this version of incompleteness theorem, was based on the eralier notion of *creativeness*. Post gave versions of Gödel's theorem based on his concept of creative set. In some way Post, denying the mechanizability of human reason, anticipated the argument that human mind cannot be reduced to any set of computational rules:

> The logical process is essentially creative. This conclusion ... makes of the mathematician much more than a kind of clever being who can do quickly what a machine could do ultimately. We see that a machine would never give a complete logic; for once the machine is made we could prove a theorem it does not prove (Post (1941), 55)

To characterize the notion of *creative set*, it will be necessary to further explore the concept of *reducibility*.

Definition 9. (Many-one riducibility) *Let $X, Y \subseteq \mathbb{N}$:*

1. *$X \leq_m Y$, i.e. X is "many-one" reducible to Y, means that $x \in X$ iff $f(x) \in Y$, for some f total computable. In case f is injective we speak of $1-$reducibility (we write $X \leq_1 Y$).*
2. *We say that X is m-complete, if it is computably enumerable and for all Y computably enumerable , $Y \leq_m X$.*
3. *$X =_m Y$ iff $X \leq_m Y$ and $Y \leq_m X$.*

Observe that if $X \leq_m Y$ and Y is computable, then also X is computable: indeed, suppose that $X \leq_m Y$ through f; let therefore $\chi_X(z) = \chi_Y(f(z))$, where χ_A is the characteristic function of A.

Theorem 20. (Post 1944) *X is computably enumerable iff $X \leq_m K$, where K is the halting set.*

Proof. $\Rightarrow$ Let X computably enumerable and let:

$$\psi(x,y) = \begin{cases} 1 & \text{if } x \in X \\ \uparrow & \text{otherwise} \end{cases}$$

By parametrization , there exists f such that $\phi_{f(x)}(y) = \psi(x,y)$. Thus

$$W_{f(x)} = \begin{cases} \mathbb{N} & \text{if } x \in X \\ \emptyset & \text{otherwise} \end{cases}$$

Note that:

1. if $x \in X$, then $W_{f(x)} = \mathbb{N}$, whereby $f(x) \in W_{f(x)}$ and therefore $f(x) \in K$

2. if $f(x) \in K$ then $f(x) \in W_{f(x)}$, and thus $W_{f(x)} \neq \emptyset$ and finally $x \in X$

Hence $X \leq_m K$.

$\Leftarrow$ If $X \leq_m K$, then X is computably enumerable; if $X \leq_m Y$ and Y is computably enumerable, then also X is. Let indeed $g : X \leq_m Y$ and let Y computably enumerable; then $Y = Dom(\phi)$. Hence let $h(x) = \phi(g(x))$ whereby $x \in X$ iff $g(x) \in Dom(\phi)$ iff $\phi(g(x)) \downarrow$ iff $x \in Dom(h)$, namely:

$$X = Dom(\phi(g(x)) = Dom(h(x))$$

Hence X is the domain of a partial computable function. QED

Definition 10. *X is* productive*, if there exists a total computable function f such that:*

$$W_e \subseteq X \Rightarrow (f(e) \in X \setminus W_e)$$

The f is called productive function*. Note that if X is productive , then is not computably enumerable (otherwise $X = W_e$ and we would have a contradiction).*

We say that X is creative *, if it is computably enumerable and its complement $\overline{X}$ is productive . For instance, the halting set K is creative , because is computably enumerable and its complement is productive :*

$$W_x \subseteq \overline{K} \Rightarrow id(x) \in \overline{K} \setminus W_x$$

where $id(x) = x$. Suppose that this does not hold and that $id(x) = x \in W_x$, namely $\phi_x(x) \downarrow$ and therefore $x \in K$ (contradiction, since $W_x \cap K = \emptyset$)

Hence a *productive set* can be "effectively" distinguished from any given computably-enumerable set. Gödel's incompleteness theorem implies that every attempt to effectively enumerate the truths of arithmetic is bound to fail: in any attempt to enumerate truth, either some falsehood is included or some truth is missed and in this case the construction permits to effectively produce the missing truth. The above definition expresses this phenomenon as a general, recursion theoretic property of sets[2].

Theorem 21. *If X is productive e $X \leq_m Y$, then Y is productive .*

Proof. Let ψ the productive function of X and let $f : X \leq_m Y$. It holds that $W_{g(x)} = f^{-1}[W_x] = \{z \in X | f(z) \in W_x\}$, $W_x \subseteq Y$, namely $f^{-1}[W_x]$ is uniformly computably enumerable in x. Hence $f(\psi(g(x)))$ is the productive function for Y; indeed:

$$W_e \subseteq Y \Rightarrow W_{g(e)} \subseteq X$$

$$\Rightarrow \psi(g(e)) \downarrow \wedge \psi(g(e)) \in X \setminus W_{g(e)}$$

But $\psi(g(e)) \notin W_{g(e)}$ implies $\psi(g(e)) \notin \{z \in X | f(z) \in W_e\}$, i.e. $f(\psi(g(e)) \notin W_e$ and then:

$$f(\psi(g(e)) \downarrow \wedge f(\psi(g(e)) \in Y \setminus W_e$$

QED

Theorem 22. *Every productive set contains an infinite computably enumerable set.*

Proof. Let X be productive and let be the productive function. Enumerate an infinite subset as follows:

1. let e_0 such that $W_{e_0} = \emptyset$; since $W_{e_0} \subseteq X$, we will have $f(e_0) \in X \setminus W_{e_0}$, namely $f(e_0) \in X$. Put $y_0 = f(e_0)$.

2. Suppose we have defined $\{y_0, ..., y_n\} \subseteq X$; let e_{n+1} such that $\{y_0, ..., y_n\} = W_{e_{n+1}} \subseteq X$; put $y_{n+1} = f(e_{n+1}) \in X \setminus W_{e_{n+1}}$ and therefore $y_{n+1} \neq y_0, ..., y_n$.

It is observed that there is an h such that $W_x \cup \{f(x)\} = W_{h(x)}$. For consider the function:

$$f_{h(x)}(y) \simeq \theta(x, y) = \begin{cases} 1 & \text{if either } y \in W_x \text{ or } y = f(x) \\ \uparrow & \text{otherwise} \end{cases}$$

Let $W_{e_{n+1}} = W_{e_n} \cup \{y_n\} = W_{e_n} \cup \{f(e_n)\} = W_{h(e_n)}$. The sequence of indices e_n is therefore given computablely: $e_{n+1} = h(e_n)$ and therefore the sequence $y_n = f(e_n)$. It follows that $B = \{y_0, y_1, y_3 ...\} = \{e_0, f(h(e_0)), f(h(h(e_0))), ...\}$ is the range of a total computable function. QED

An infinite set is called *immune* if it does not contain infinite recursively enumerable sets. A set is *simple* if it is recursively (or computably) enumerable and its complement has this property. Post proved the existence of such sets. Simple sets are in intermediate between recursive and creative sets. Indeed, since a productive set contains an infinite recursively enumerable subset, a simple set is neither recursive nor creative. Now we show the equivalence between the two notions of *creativeness* and *m-completeness*.

Theorem 23. *If X is m-complete, then is creative .*

[2] In some textbooks the productive function is given as a total function, while in others it is given as *partial*. We follow in this exposition Odifreddi (1989-1999).

Proof. Let $g : K \leq_m X$ and let $W_{h(x)} = \{z | g(z) \in W_x\}$, with X computably enumerable. Note that if $g(z) \in W_x \subseteq \overline{X}$, then $z \in \overline{K}$. Hence, if $W_x \subseteq \overline{X}$, then $W_{h(x)} \subseteq \overline{K}$. In particular $h(x) \notin W_{h(x)}$ and then $h(x) \in \overline{K}$. In conclusion:

$$W_x \subseteq \overline{X} \Rightarrow h(x) \in \overline{K} \setminus W_{h(x)}$$

$$\Rightarrow g(h(x)) \in \overline{X} \setminus W_x$$

Note that $g(h(x)) \in W_x \Rightarrow h(x) \in W_{h(x)}$. QED

Theorem 24. (Myhill 1955) *If X is creative, then X is $m-$complete.*

Proof. Let X creative ; then $\overline{X}$ is productive . Let h the productive function and let Y computably enumerable; let us define, for m, n fixed, the following algorithm: given r as input, look for this n in Y:

1. if n appears, compute $h(m)$,
2. if $h(m) = r$, output $= 0$.

Otherwise the algorithm does not give output. Because it depends uniformly from m and n, we can express it as $\phi_{f(m,n)}$ for some f total computable. Hence we have:

$$W_{f(m,n)} = \begin{cases} \{h(m)\} & \text{if } n \in Y \\ \emptyset & \text{otherwise} \end{cases}$$

Now we use this version of the second recursion theorem: for every f there exist a function ν_f such that $\phi_{\nu_f(x)} \simeq \phi_{f(\nu_f(x),x)}$. Let therefore ν_f be such that $W_{\nu_f(x)} = W_{f(\nu_f(x),x)}$, and then, returning to our case:

$$W_{\nu_f(n)} = W_{f(\nu_f(n),n)} = \begin{cases} \{h(\nu_f(n))\} & \text{if } n \in Y \\ \emptyset & \text{otherwise} \end{cases}$$

Hence:

1. $n \in Y \Rightarrow h\nu_f(n) \in W_{\nu_f(n)} \Rightarrow W_{\nu_f(n)} \not\subseteq \overline{X} \Rightarrow h\nu_f(n) \in X$
2. $n \notin Y \Rightarrow W_{\nu_f(n)} = \emptyset \Rightarrow h\nu_f(n) \in \overline{X}$

Hence $n \in Y$ iff $h\nu_f(n) \in X$, and therefore $h\nu_f : Y \leq_m X$. QED

Let us now return to logic, with the aim of arriving at the proof of versions of the incompleteness theorem that emphasise Gödel's original insights we mentioned at the beginning of this section on the different complexity of the set of true statements and the set of theorems, but saying something more precise about the complexity of these sets. We consider the classes of Δ_0 and Σ_1 formulas of the language $\mathcal{L} = \{+, \cdot, S, 0, \leq\}$ defined at p. 13. The following properties hold already in Robinson's theory Q:

1. If ϕ is un Δ_0-sentence, then:
   (a) if $\mathbb{N} \models \phi$ then $\mathsf{Q} \vdash \phi$
   (b) If $\mathbb{N} \models \neg\phi$ then $\mathsf{Q} \vdash \neg\phi$
2. If ϕ is a Σ_1-sentence, then:
   (a) if $\mathbb{N} \models \phi$ then $\mathsf{Q} \vdash \phi$

(b) Here the "negative part" is not valid. Indeed, Gödel's undecidable sentence is just a Π_1-sentence of the form $\neg\exists x\phi(x)$.

In its constructive (syntactic) version of the first incompleteness theorem, Gödel introduces the concept of ω-consistency, a property stronger than consistency, and proves that if T is an extesion of Robinson's arithmetic, then there exists a sentence ϕ of its language such that, if T is consistent, then ϕ is not provable in T; if it is ω-consistent, then also $\neg\phi$ is unprovable. The ω-consistency is the property that if $\mathsf{T} \vdash \phi(\overline{n})$ for all n, then $\mathsf{T} \nvdash \exists x\neg\phi(x)$. However we will see that in Rosser's version the ω-consistency can be dispensed, and replaced by mere consistency. We will see later with all the details that all computable relations have their "counterpart" formalized in the language $\mathcal{L}$, such that if the relation holds, then that counterpart is true. In fact, we can say more: if the relation holds, its counterpart is provable in Q and if the relation does not hold, then its counterpart is refutable. More formally, we will see that in all extension of Robinson's Q, all recursive (or computable) relations are *representable* (or *binumerable*) and all recursively enumerable (or computably enumerable) relations are *weakly representable* (or *numerable*). Those representing formulas are in Σ_1. Let therefore $\tau(x, y, z)$ be the "counterpart" of Kleene's $T(x, y, z)$ in the language $\mathcal{L}$ and let T be an extension Σ_1-sound of Q; then the following holds:

1. $n \in \overline{K}$ iff for all $m \in \mathbb{N}$, $\mathsf{T}(n, n, m)$ is false, iff $\mathbb{N} \models \tau(\overline{n}, \overline{n}, \overline{m})$, for all $m \in \mathbb{N}$, iff $\ulcorner\neg\exists y\tau(\overline{n}, \overline{n}, y)\urcorner \in Th(\mathbb{N})$.

2. $n \in \overline{K}$ iff for all $m \in \mathbb{N}$, we have that $T(n, n, m)$ is false and therefore by representability $\mathsf{T} \vdash \neg\tau(\overline{n}, \overline{n}, \overline{k})$ for all k and by Σ_1-soundness (or equivalently, the 1-consistency, see at p. 108) $\mathsf{T} \nvdash \exists y\tau(\overline{n}, \overline{n}, y)$, from which $\ulcorner\exists y\tau(\overline{n}, \overline{n}, y)\urcorner \in \overline{Thm_{\mathsf{T}}}$ (the set-theoretical complement of the set of theorems).

Theorem 25. *Let* T *an axiomatizable extension and* Σ_1-*sound of* Q*; then:*

1. *the set* Thm_{T} *is creative ;*

2. *the set* $Th(\mathbb{N})$ *is productive (hence not axiomatizable).*

Proof. Let $f(n) = \ulcorner\exists y\tau(\overline{n}, \overline{n}, y)\urcorner$ and $g(n) = \ulcorner\neg\exists y\tau(\overline{n}, \overline{n}, y)\urcorner$. What we have before verified is that:

1. $n \in \overline{K}$ iff $f(n) \in \overline{Thm_{\mathsf{T}}}$;

2. $n \in \overline{K}$ iff $g(n) \in Th(\mathbb{N})$

Recall that $\overline{K}$ is productive, and that is true in general that if a set is productive and is m-riducible to another set, also this will be productive. But $f : \overline{K} \leq_m \overline{Thm_{\mathsf{T}}}$ and $g : \overline{K} \leq_m Th(\mathbb{N})$; therefore $Th(\mathbb{N})$ will be productive. In addition, if T is axiomatizable then Thm_{T} is computably enumerable, and if its complement is productive, it will be creative. QED

So the two sets: the set of true statements and the set of theorems cannot coincide since they are of different complexities. Using these tools of computability theory again, we can also give a theorem of essential incompleteness and essential undecidability in Rosser's style (see Smullyan (1961), pp. 47-55). Now we work on pairs of sets:

1. $X, Y \subseteq \mathbb{N}$ are called *computablly inseparable* if there is no computable set U such that $X \subseteq U$ and $U \cap Y = \emptyset$, namely, that separates them.

2. $X, Y \subseteq \mathbb{N}$ are called *effectively* inseparable, if there exists a partial computable function ψ such that, if $X \subseteq W_u$, $Y \subseteq W_y$ and $W_u \cap W_v = \emptyset$, then $\psi(u, v) \downarrow \in \overline{W_u \cup W_v)}$

It holds that $X, Y \subseteq \mathbb{N}$ are effectively inseparable, then they are computably inseparable. Also it is true that $X, Y \subseteq \mathbb{N}$ are effectively inseparable, then they are creative. Such sets exists: for instance, the following sets $A = \{x|\phi_x(x) \simeq 0\}$ and $B = \{x|\phi_x(x) \simeq 1\}$ are effectively inseparable. We define indeed a $\chi(u, v, z)$ which lists simultaneously W_u and W_v and gives output 1, if z appears first in W_u, while giving output 0, if z appears first in W_v. This done, we place $\chi(u, v, z) \simeq \phi_{h(u,v)}z$. Suppose that $A \subseteq W_u$ and $B \subseteq W_v$, where $W_u \cap W_v = \emptyset$. Then

$$h(u, v) \in W_u \Rightarrow \chi(u, v, h(u, v)) \simeq 1$$
$$\Rightarrow \phi_{h(u,v)}h(u, v) \simeq 1$$
$$\Rightarrow h(u, v) \in B$$

Absurd, since $B \cap W_u = \emptyset$. Analogous, if $h(u, v) \in W_v$.

Theorem 26. (Rosser's essential undecidability) *Each consistent and axiomatizable extension* T *of Robinson's arithmetic has an undecidable sentence.*

Proof. Let A, B effectively inseparable computably enumerable. Hence $A = W_a$, $B = W_b$; recall a theorem, due to Gödel (which will be discussed in greater detail later on), saying that computable relations are *representable* (or *binumerable*) in T, namely, in our specific case, that we can find a formula $\tau(x, y, z)$ that represents Kleene's predicate $T(x, y, z)$, i.e. such that:

1. if $T(a, n, m)$ holds then $\mathsf{T} \vdash \tau(\overline{a}, \overline{n}, \overline{m})$
2. if $T(a, n, m)$ does not hold, then $\mathsf{T} \vdash \neg\tau(\overline{a}, \overline{n}, \overline{m})$

Now let $(A \prec B)(\overline{n})$ be the formula:

$$\exists z(\tau(\overline{a}, \overline{n}, z) \wedge \forall y \leq z \neg\tau(\overline{b}, \overline{n}, y))$$

The following applies:

1. if $n \in A$, then for some m, $T(a, n, m)$ is true and therefore it follows that $\mathsf{T} \vdash \tau(\overline{a}, \overline{n}, \overline{m})$ by binumerability. But A and B are disjoint, hence $n \notin B$ and therefore the relation $T(b, n, r)$ is false for all $r \leq m$ and for all $s \leq m$, $\mathsf{T} \vdash x = \overline{s} \rightarrow \neg\tau(\overline{b}, \overline{n}, \overline{s})$ and hence:

$$\mathsf{T} \vdash \bigvee_{s \leq m} x = \overline{s} \rightarrow \neg\tau(\overline{b}, \overline{n}, \overline{s})$$

Now let us consider that $Q \vdash x \leq \overline{m} \leftrightarrow x = \overline{0} \vee ... \vee x = \overline{m}$. It follows $\mathsf{T} \vdash x \leq \overline{m} \rightarrow \neg\tau(\overline{b}, \overline{n}, x)$. Hence we have proved in T:

$$\exists y(\tau(\overline{a}, \overline{n}, y) \wedge \forall z \leq y \neg\tau(\overline{b}, \overline{n}, z))$$

2. Analogously, if $n \in B$, then there is an r such that $T(b, n, r)$ and therefore $\mathsf{T} \vdash \tau(\overline{b}, \overline{n}, \overline{r})$. Since A and B are disjoint we also have that $T(a, n, m)$ is false for all m, from which follows $\mathsf{T} \vdash \neg\tau(\overline{a}, \overline{n}, \overline{m})$. But in Q is provable $y \leq \overline{r} \vee \overline{r} < y$. If $y \leq \overline{r}$ then $\mathsf{T} \vdash \neg\tau(\overline{a}, \overline{n}, y)$; if $\overline{r} < y$, then $\mathsf{T} \vdash \exists z < y\tau(\overline{b}, \overline{n}, z)$. We have proved $\mathsf{T} \vdash \neg(A \prec B)(\overline{n})$, namely:

$$\forall y(\neg\tau(\overline{a}, \overline{n}, y) \vee \exists z < y\tau(\overline{b}, \overline{n}, z))$$

Let $\mathring{A} = \{n|\mathsf{T} \vdash (A \prec B)(\overline{n})\}$ and $\mathring{B} = \{n|\mathsf{T} \vdash \neg(A \prec B)(\overline{n})\}$.

Note that $\mathring{A}$ and $\mathring{B}$ are *computably enumerable* sets: actually, for any θ and axiomatizable theory T, the relation: "there is a proof in T of θ", can be expressed by a Σ_1-formula. Moreover, if T is consistent, then these are disjoint sets. Let therefore $\mathring{A} = W_{\dot{a}}$ e $\mathring{B} = W_{\dot{b}}$. Since A, B are effectively inseparable we will have a function $h(\dot{a}, \dot{b}) \in \overline{W_{\dot{a}} \cup W_{\dot{b}}}$, i.e. $h(\dot{a}, \dot{b}) \notin \mathring{A}$ and $h(\dot{a}, \dot{b}) \notin \mathring{B}$ and therefore $T \nvdash (A \prec B)(\overline{h(\dot{a}, \dot{b})})$ and $T \nvdash \neg(A \prec B)(\overline{h(\dot{a}, \dot{b})})$.

QED

Using the terminology of Smullyan (1961) we can say that such a se sentence $(A \prec B)(\overline{h(\dot{a},\dot{b})})$ *separates A and B in* T. Smullyan calls "Rosser theory" a theory in which every pair of computably enumerable sets is separable in it and all axiomatizable theories that consistently extend Robinson Arithmetic actually meet this requirement. Above, in fact we have shown that this property implies essential undecidability. This further property applies to these theories T.

Corollary 4. *The sets* Thm_{T} *and* Ref_{T} *of (Gödel numbers of) provable sentences and of refutable sentences of* T *are effectively inseparable.*

Proof. Recall that if (A, B) and (X, Y) are disjoint pairs of computably enumerable sets and (A, B) is effectively inseparable and either $(A, B) \subseteq (X, Y)$ or $(A, B) \leq_m (X, Y)$, then also (X, Y) is effectively inseparable, where $(A, B) \leq_m (X, Y)$ means that for some computable f, $f[A] \subseteq X$ and $f[B] \subseteq Y$. Indeed, let $h(x, y)$ the "productive" function for (A, B). Recall that in general there is a g such that $W_{g(x)} = f^{-1}[W_x]$ and define:

$$h^*(x, y) = f(h(g(x), g(y)))$$

Let therefore $X \subseteq W_i$ and $Y \subseteq W_j$ for disjoint computably enumerable sets W_i, W_j. Hence $A \subseteq f^{-1}[W_i]$ and $B \subseteq f^{-1}[W_j]$ where $f^{-1}[W_j] = W_{g(j)}$, $f^{-1}[W_i] = W_{g(j)}$ are disjoint. Thus, $h(g(i), g(j)) \notin W_{g(i)} \cup W_{g(j)}$, from which $h^*(i, j) \notin W_i \cup W_j$.

Now, we have that the above effectively inseparable A, B satisfy $A \subseteq \mathring{A}$ and $B \subseteq \mathring{B}$. Moreover $(\mathring{A}, \mathring{B}) \leq_m (Thm_{\mathsf{T}}, Ref_{\mathsf{T}})$ via the function $f(n) = \ulcorner(A \prec B(\overline{n}))\urcorner$. QED

Some further refinement is possible. For instance, Bernardi (1981) shows that if $\mathsf{T} \nvdash \phi \leftrightarrow \psi$, then the equivalence classes $[\phi] = \{\sigma | \mathsf{T} \vdash \phi \leftrightarrow \sigma\}$ and $[\psi] = \{\sigma | \mathsf{T} \vdash \psi \leftrightarrow \sigma\}$ are effectively inseparable. Indeed, if $[\phi] \subseteq W_i$ and $[\psi] \subseteq W_j$ and $W_i \cap W_j = \emptyset$, take a formula $\alpha(x)$ which separates W_i and W_j, i.e. such that $\mathsf{T} \vdash \alpha(\overline{n})$ if $n \in W_i$ and $\mathsf{T} \vdash \neg\alpha(\overline{n})$ if $n \in W_j$. The take the fixed point:

$$\tau \leftrightarrow ((\alpha(\overline{\ulcorner\tau\urcorner}) \wedge \psi) \vee (\neg\alpha(\overline{\ulcorner\tau\urcorner}) \wedge \phi))$$

Now observe that if $\ulcorner\tau\urcorner \in W_i$, then $\mathsf{T} \vdash \alpha(\overline{\ulcorner\tau\urcorner})$ and by the logic this implies $\mathsf{T} \vdash \tau \leftrightarrow \psi$. However $\ulcorner\psi\urcorner \in W_j$, a contradiction. A specular argument applies if $\ulcorner\tau\urcorner \in W_j$.

From the effective inseparability of the sets of theorems and of refutable formulas of T clearly follows the essential undecidability of T. Indeed, suppose by contradiction that a theory T (consistent and r.e.) is effectively inseparable, but not essentially undecidable. Hence let S a axiomatizable consistent extension of it and suppose that ψ is the "productive function" for the pair $(Thm_{\mathsf{T}}, Ref_{\mathsf{T}})$. Let $W_i = Thm_{\mathsf{S}}$ and W_j its complement. Hence $Thm_{\mathsf{T}} \subseteq W_i$ and $Ref_{\mathsf{T}} \subseteq W_j$. It follows that $\psi(i, j) \notin W_i \cup W_j = \mathbb{N}$, which is a contradiction (for an in-depth analysis of these topics, see, for example, Cheng (2023)). The connection between the effective inseparability of theorems and inseparable statements and the incompleteness theorem can also be highlighted by resorting to the notion of effective extensibility, equivalent equivalent to effective inseparability, introduced in Boykan Pour-El (1968). Let us consider the axioms of a computably enumerable theory as a set W_e of index e. A theory T is said to be *effectively extensible* if there exists a computable function f such that, if e is an index of a computably enumerable extension S of T, then $f(e) \notin Thm_{\mathsf{S}} \cup Ref_{\mathsf{S}}$. The connection between the *effective inseparability* of theorems and refutable sentence, on the one hand, and the incompleteness theorem, on the other, can also be highlighted by resorting to the equivalent notion of *effective extensibility*, introduced in Boykan Pour-El (1968). Let us consider the axioms of a computably enumerable theory as a set W_e of index e. A theory T is said to be effectively extensible if there exists a computable function f such that, if e is an index of a computably enumerable extension S of T, then $f(e) \notin Thm_{\mathsf{S}} \cup Ref_{\mathsf{S}}$.

The issue of the different complexity between the set of true statements and the set of provable statements in certain theories can be further investigated by considering a more general notion of reducibility. Many modern computer processes are *online* interactive processes, in the sense that they interact with the environment or consult external databases (e.g. the web). In 1939 Turing

sketched a formalization of this idea, through the description of an *oracle machine* ("o-machine") and in so doing invented relativisation and, in essence, the Turing jump (see Cooper (2004) and Soare (2009) for an historical overview). This idea was later considerably developed in Post (1944) and Post (1948). An oracle is a kind of black box (not necessarily a program) able to solve some problems and to produce an answer to question like: "is the element a in X?" These machines represent an attempt to extend the power of ordinary machines and to overcome the incomputable, since they can compute something that ordinary Turing machines cannot compute. The oracle machine is a relativised model very useful in order to compare and classify degrees of undecidability of problems, and to define the *Arithmetical Hierarchy*.

We introduce a more general concept of reducibility, based on a model of machine that allows the consultation of an *oracle* during the course of computation. A limitation of the notion of m-reducibility that requires a correction can be derived from this example. If $B \leq_m A$, it seems natural to ask that A encompasses the information contained in B. In some sense A contains also the information about $\overline{A}$, yet there is no computably enumerable *not recursive* set such that $\overline{A} \leq_m A$. Suppose, on the contrary, that such a set exists, i.e. that there exist an f such that $n \in \overline{A}$ iff $f(n) \in A$. If A is computably enumerable, then $A = W_e$; hence $\phi_e(f(n)) \downarrow$ iff $f(n) \in A$ iff $n \in \overline{A}$; therefore $\overline{A} = Dom(\phi_e(f(x))$ and $\overline{A}$ would be computably enumerable and therefore A would be recursive, against the hypothesis.

To introduce this new model, let us consider for example our first very basic model of a Turing machine, where we made these conventions.

1. Input convention: To input n, place $n + 1$ consecutive 1's on the tape.
2. Output convention: If a computation halts — which only happens when there is no applicable quadruple in the program — output the number $f(n)$ of 1's left printed on the tape.

We can think of a machine with oracle for A as equipped, in addition of the input tape, of a further read-only tape, which contains the characteristic function $\chi_A(x)$ of A – substantially a binariy string, and the oracle tape head begins on the cell containing $\chi_A(0)$.

The oracle machine is based on instructions of the type $q_i \alpha q_j q_e$ ("You are in the state q_i reading α, count the number n of 1 on the input tape and ask the oracle if $n \in A$; if the answer is yes, go in the state q_j, otherwise, go in the state q_e").

Definition 11. *Let $B, A \subseteq \mathbb{N}$; we say that B is computable from A if we can answer the question "$n \in B$?" by means of an algorithm that may have available a finite number of answers to questions about the membership of A, i.e. to questions of this form:*

$$n_0 \in A?, ..., n_k \in A?$$

We write $B \leq_T A$ ("A computes B").

Note that if $B \leq_m A$, then $B \leq_T A$. Since any description of an oracle Turing machines is finite, it is possible to effectively encode them with natural numbers. We write, using uppercase (to emphasise that they are functional, rather than functions) Greek letters, $\Phi^A(x)$ to denote the computation of the $e-th$ oracle Turing machine with oracle A on input x. The m-reducibility can be seen as a Turing-reducibility in which we are allowed to ask A just once "$f(n) \in A$?". Note that now $\overline{A} \leq_T A$: indeed, let $\Phi^A(n)$ the function that outputs 1 if $n \in A$, and 0 if $n \notin A$, noting that $\chi_{\overline{A}} \simeq \Phi^A$.

Definition 12. *We say that:*

1. *ψ is $A-recursive$, iff there exist an e such that $\psi \simeq \Phi_e^A$*
2. *X is recursive in Y, i.e. $X \leq_T Y$, iff the characteristic function χ_X is $Y-recursive$, i.e. there is an e such that $\chi_X(y) \simeq \Phi_e^Y(y)$. For simplicity we will write $X(y) \simeq \Phi^Y(y)$, identifying X with its characteristic function.*
3. *X is computably enumerable in Y, if X is empty, or $X = Cod(f)$ and $f \simeq \Phi_e^Y$, for some e (equivalently, $X = W_e^Y$, for some e).*

4. The degree of a set A is the equivalence class $deg(A) = \{X | X =_T A\}$.

We can now formulate a *relativised* version of the Church-Turing thesis which has this shape:

> *Post-Turing thesis*: B is effectively *computable from* A iff B is computable by a Turing machine *with oracle in* A.

Degrees are ordered by the relation $deg(A) \preccurlyeq deg(B)$ iff $A \leq_T B$. Let $< \mathbb{D}, \preccurlyeq >$ the set of degrees with this relation. In particular, there is a minimum degree (that of recursive sets $deg(\emptyset)$, denoted by $\mathsf{0}$, but there is no maximum degree, $< \mathbb{D}, \preccurlyeq >$ is not linearly ordered and it is an upper semilattice: that is to say, given two elements, exists the supremum, but there are pairs of degrees that do not have the infimum.

Definition 13. *The* jump *A' of A is defined as follows:*

1. $A' = \{x | \Phi_x^A(x) \downarrow\} = K^A$.
2. $A^{n+1} = (A^n)'$
3. $A^\omega = \{\langle m, n\rangle | m \in A^n\}$

Let $\mathsf{a}^n = deg(A^n)$; the level $\mathsf{0}'$ is therefore that of K, of creative sets as the theorems of axiomatizable theories of formal arithmetic (e.g. *Peano arithmetic*). We will see that the set of the true statements of these theories is $\mathsf{0}^\omega$. Given two degrees $\mathsf{a} = deg(A)$ and $\mathsf{b} = deg(B)$, always exists $sup\{\mathsf{a}, \mathsf{b}\}$ and is defined as $\mathsf{a} \vee \mathsf{b} = deg(A \oplus B)$, where $A \oplus B = \{2x | x \in A\} \cup \{2x+1 | x \in B\}$. This does not apply for the infimum.

In fact, for a result of Kleene, Post and Spector, the sequence $\mathsf{0}, \mathsf{0}', \mathsf{0}'', ...$ has an "exact pair", that is, a couple of degrees a and b such that:

1. for all $n \in \mathbb{N}$, $\mathsf{0^n} \preccurlyeq a$ and $\mathsf{0^n} \preccurlyeq b$

2. for all d, $\mathsf{d} \preccurlyeq \mathsf{a}$ and $\mathsf{d} \preccurlyeq \mathsf{b}$ implies that exists $n \in \mathbb{N}$, $\mathsf{d} \preccurlyeq \mathsf{0^n}$.

It follows that a and b can not have an infimum.

In Post (1944) the Polish-American logician posed the problem od determining if there are sets A, B computably enumerable incomparable with respect to $\leq_T$, i.e. $A \not\leq_T B$ and $B \not\leq_T A$. After in 1954 Kleene and Post had proved that there exist sets A, B (not necessarily computably enumerable) incomparable and such that $A, B \leq_T \emptyset'$, in the mid-1950s the following theorem was finally proved.

Theorem 27. (Friedberg-Muchnik 1956-1957) *There are sets A, B computably enumerable incomparable (hence $\emptyset \leq_T A, B \leq_T \emptyset' = K$, and being A, B incomparable, actually we have $<_T$).*

Proof. (see Cooper (2003), pp. 238-41 for a proof). QED

In what follows, we must investigate the relationships between *computability* and *definability* (in the standard model). For this reason, we must first dwell on the *Arithmetical Hierarchy*, more than we have already anticipated. Let us consider formulas of the language of a first order theory of formal arithmetic. We start with the following classification[3].

Arithmetical hierarchy of formulas. Let us return briefly to the hierarchy of formulas introduced at p. 13, pointing out only that some authors use a different, *extended*, definition of the arithmetical hierarchy, in which bounded quantifiers are considered immaterial when counting the complexity of a formula: for example, Σ_{n+1} is the set of formulas obtained by prepending an arbitrary block of existential quantifiers *and bounded universal* quantifiers to Π_n -formulas.

A relation R is *arithmetical* if there is a formula θ of the language of Peano (and of Robinson) arithmetic which defines it in the standard model, i.e. such that:

$$\mathbb{N} \models \theta(\overline{n_0}, ..., \overline{n_k}) \text{ if and only if } \langle n_0, ..., n_k\rangle \in R$$

[3] The correct notation would be Σ_n^0, Π_n^0, where the superscript 0 means "first order"; as we deal only with these formulas, we omit this symbol for simplicity.

where $\overline{n_j}$ is the term which denotes the number n_j. Accordingly, a relation R is Σ_n-definable, if it is definable by a formula of complexity Σ_n and so on. However, we would like to warn the reader against possible misunderstandings, because in the scientific literature there are different conventions with regard to the level $\Delta_0 = \Sigma_0 = \Pi_0$ in the hierarchy of sets:

1. *In computability theory*: at this level there are just the computable relations. In Odifreddi (1989-1999) pp. 363-73 this convention is adopted: extend the language with symbols for every recursive relation; the intended model is the standard model expanded with all recursive relations. Being the computable relations first-order definable in the language of arithmetic, we may suppose that, for each computable relation, the language contains a relation symbol.

2. *In formalized in arithmetic*: at the zero-level there are the relations definable by formulas in which the quantifiers that are allowed to appear are only *bounded* quantifiers. Hence, what there is actually depends on the language.

Some closure properties allow to simplify the definition. Since in the standard model $\langle \mathbb{N}, +, \cdot, S, 0, < \rangle$ holds the collection (or replacement) scheme:

$$\forall x \leq t \exists y \theta(x, y) \leftrightarrow \exists z \forall x \leq t \exists y \leq z \theta(x, y)$$

saying that a bounded quantifier can be 'pushed inside' an unbounded quantifier, the above definitions (with or without bounded quantifiers interpsersed) are equivalent *in the model*. If the theory in which we work is strong enough to prove it, the equivalence holds in *the theory* too. We can also Δ_0-define the *pairing relation* $\langle x, y \rangle = z$ in the language of Peano arithmetic as $2z = (x + y + z)(x + y) + 2x$. With this, we can contract two quantifiers of the same sort into one, e.g. $\exists x \exists y \theta(x, y)$ becomes $\exists v \forall x \leq v \forall y \leq v (2v = (x + y + z)(x + y) + 2x \rightarrow \theta(x, y))$. In conclusion, in definability, the alternations of quantifiers can be be understood as alternations of *blocks* of universal or existential quantifiers and each block of quantifiers of the same species can be contracted to only one quantifier of that species. The relativised arithmetic hierarchy Σ_n^A, Π_n^A is defined as above, except that the matrix R of a formula in prenex form $Q_0 x_0 ... Q_n x_n R(x_0, ..., x_n)$ instead of being recursive, is $A-$recursive, i.e. $R \leq_T A$. More exactly:

1. $\Sigma_0^A = \Pi_0^A = \Delta_0^A =$ sets recursive in A.
2. $\Sigma_{n+1}^A =$sets definable by formulas of the form $\exists x R(x, y)$ with $R \in \Pi_n^A$.
3. $\Pi_{n+1}^A =$sets definable by formulas of the form $\forall x R(x, y)$ with $R \in \Sigma_n^A$.
4. $\Delta_{n+1}^A = \Sigma_{n+1}^A \cap \Pi_{n+1}^A$

Many results "relativise" by substituting the concept of *recursiveness* with its version relativised to a given set. For instance $B \in \Sigma_1^A$ if and only if $B = W_e^A$.

We now come to an important result for the purpose of this section, namely Post's theorem. Though Post did not publish this theorem with a proof, Kleene (1952) credits Post with the idea.

Theorem 28. (Post's Theorem) *The following relations hold:*

1. $B \in \Sigma_{n+1}$ *iff* B *is computably enumerable in some* $A \in \Sigma_n$.

2. $\emptyset^{n+1}$ *is* Σ_{n+1}*-complete, namely is itself* Σ_{n+1} *and for all* $A \in \Sigma_{n+1}$, $A \leq_m \emptyset^{n+1}$.

3. $B \in \Sigma_{n+1}$ *iff* B *is computably enumerable in* $\emptyset^n$.

4. $A \in \Delta_{n+1}$ *iff* $A \leq_T \emptyset^n$.

5. *Hence, in particular:*

   (a) $A \in \Delta_1$ *iff* $A \leq_T \emptyset$.

   (b) $A \in \Delta_2$ *iff* $A \leq_T \emptyset'$

Proof. The proof requires some preliminary lemma, which we will see now. First let's make these simple observations:

1. If A is computably enumerable in B and $B \leq_T C$, then A is computably enumerable in C. Indeed, since $B \leq_T C$ iff $B(x) \simeq \Phi_e^C(x)$, then, if $A = W_a^B$, we can consider C as another oracle that we can consult to know if $n \in B$ and then $A = W_j^C$.

2. B is computably enumerable in A iff B is Σ_1^A. This is a relativised version of a well-known result.

3. B is computably enumerable in A iff B is computably enumerable in $\overline{A}$. It follows from the first point, and from the fact that $\overline{A} \leq_T A$.

QED

Lemma 4. *B is computably enumerable in A, iff $B \leq_m A'$.*

Proof. $\Leftarrow$ If $f : B \leq_m A'$, then $x \in B$ iff $f(x) \in A'$ iff $\Phi^A_{f(x)}(f(x)) \downarrow \simeq \psi^A(x)$. Ergo $B = Dom(\psi^A(x))$. $\Rightarrow$ Let $B = W_e^A$; let us remember that if $B = W_e^{\emptyset}$ then $B \leq_m \emptyset' = K$ (completeness of di K). More generally, relativizing, $B = W_e^A$ implies $B \leq_m A'$. QED

Lemma 5. *$B \leq_T A$ iff $B, \overline{B}$ are computably enumerable in A.*

Proof. $\Leftarrow$ If $B = W_e^A$ and $\overline{B} = W_i^A$, take the algorithm $\xi(e, i, n)$ that runs simultaneously $\Phi_e^A(n)$ e $\Phi_i^A(n)$. One of the two halts and answers the question if $n \in B$, hence $B \leq_T A$. $\Rightarrow$ Exercise. QED

Definition 14. *. $\phi_{e,s}(x) = y$ iff $x, y, e < s$ and y is the output of $\phi_e(x)$ obtained in less than s steps.*

Theorem 29. *$B \in \Sigma_{n+1}$ iff B is computably enumerable in some set Σ_n iff B is computably enumerable in some set Π_n.*

Proof. We follow Cooper (2003) pp. 154-57 in proposing rather a sketch of the proof to highlight the key points. $\Rightarrow$ $x \in B$ iff $\exists y A(x, y)$ is true, for some $A \in \Pi_n$ (notation: $B \in \Sigma_1^A$), namely is computably enumerable in A; but $A \leq_T \overline{A}$ and therefore B is computably enumerable also in $\overline{A}$ and $\overline{A} \in \Sigma_n$. $\Leftarrow$ If $B = W_e^A$ for $A \in \Sigma_n$, then $x \in B$ iff:

1. "there exists an s.
2. there exist finite *positive* answers $y_0 \in A, ..., y_m \in A$,
3. there exist finite *negative* answers $x_0 \in \overline{A}, ..., x_k \in \overline{A}$,

that allow us to determine whether $x \in W_{e,s}^A$".

We remark that 2. is Σ_n, 3. is Π_n, e.g. for $n = 3$ the conjunction of 2. and 3. has the form $\exists y_0 ... \exists y_m [\exists \forall \exists \wedge ... \wedge \exists \forall \exists] \wedge \exists x_0 ... \exists x_k [\forall \exists \forall \wedge ... \wedge \forall \exists \forall]$ that is equivalent to $\exists \forall \exists \forall$.

Lastly, "$x \in W_{e,s}^A$" is a computable relation and the entire expression in quotes is Σ_{n+1} QED

Theorem 30. *For all $n > 0$, $\emptyset^n$ is Σ_n-complete.*

Proof. For $n = 1$ obvious, because $\emptyset' = K$; if $n > 1$ suppose by induction hypothesis that $\emptyset^n$ is Σ_n-complete; note that $e \in \emptyset^{n+1}$ iff $\Phi_e^{\emptyset^n}(e) \downarrow$ iff $e \in W_e^{\emptyset^n}$, namely $\emptyset^{n+1}$ is computably enumerable in $\emptyset^n$, that by hypothesis is Σ_n complete. Hence from the previous theorem it follows that $\emptyset^{n+1}$ is Σ_{n+1}. In addition, it is complete: in fact suppose that $B \in \Sigma_{n+1}$, hence is computably enumerable in Σ_n and by the induction hypothesis $\emptyset^n$ is Σ_n-complete; it follow that B is computably enumerable in $\emptyset^n$; but for the above lemmas we have that this is true iff $B \leq_m (\emptyset^n)' = \emptyset^{n+1}$. QED

Theorem 31. *$B \in \Sigma_{n+1}$ iff B is computably enumerable in $\emptyset^n$.*

Proof. $\Leftarrow$ If B is computably enumerable in $\emptyset^n$, then, being $\emptyset^n \in \Sigma_n$, B is computably enumerable in a set Σ_n. But this means that $B \in \Sigma_{n+1}$. $\Rightarrow$ if $B \in \Sigma_{n+1}$, then B is computably enumerable in some $A \in \Sigma_n$; but $\emptyset^n$ is Σ_n-complete and therefore $A \leq_T \emptyset^n$ e B is computably enumerable in $\emptyset^n$. QED

Theorem 32. *$B \in \Delta_{n+1}$ iff $B \leq_T \emptyset^n$.*

Proof. $B \in \Delta_{n+1}$ iff $B \in \Sigma_{n+1}$ and $B \in \Pi_{n+1}$, namely $\overline{B}, B \in \Sigma_{n+1}$, iff $\overline{B}, B$ are computably enumerable in $\emptyset^n$, iff $B \leq_T \emptyset^n$. QED

We give another indirect version of Gödel theorem, from which the distance, in terms of complexity, between the set of true statements and the set of theorems of an axiomatisable theory becomes even more evident, considering that the set of theorems is placed at the level of the first jump of the empty set.

Theorem 33. *$\emptyset^\omega \equiv_T Th(\mathbb{N})$.*

Proof. 1. $\emptyset^\omega \leq_T Th(\mathbb{N})$. Remember that A' is computably enumerable in A, in particular $\emptyset^{n+1} = W_e^{\emptyset^n}$. For E. Post theorem we have then $\emptyset^{n+1}$ is Σ_{n+1}-definable:

$$m \in \emptyset^{n+1} \Leftrightarrow \mathbb{N} \models \overbrace{\exists x_0 \forall x_1 \exists x_2 ...}^{n+1-times} R(x_0, ..., x_n, m)$$

Take therefore $h(m) = \ulcorner \overbrace{\exists x_0 \forall x_1 \exists x_2 ...}^{n+1-times} R(x_0, ..., x_n, m) \urcorner$. Clearly $\langle m, n \rangle \in \emptyset^\omega$ iff $m \in \emptyset^n$ iff $h(m) \in Th(\mathbb{N})$, i.e. $\emptyset^\omega \leq_T Th(\mathbb{N})$.

1. $\emptyset^\omega \geq_T Th(\mathbb{N})$. Consider a formula in the prenex form $Q_0 x_0, ..., Q_n x_n \theta$, where each Q_j is a quantifier. We transform it in a Σ_{n+1}-formula in this way:

$$\exists y Q_0 x_0, ..., Q_n x_n (\theta \wedge y = y \wedge z = z)$$

Let now $B = \{c | \exists y Q_0 x_0, ..., Q_n x_n (\theta \wedge y = y \wedge c = c)\}$. Note that:

(a) If $Q_0 x_0, ..., Q_n x_n \theta$ is true, then $B = \mathbb{N}$

(b) If $Q_0 x_0, ..., Q_n x_n \theta$ is false, then $B = \emptyset$

We give a sketch of the proof (see Rogers (1987) p. 318 for further details). Since $B \in \Sigma_{n+1}$, then for Post's results it is computably enumerable in $\emptyset^n$, namely $B = W_e^{\emptyset^n}$. Moreover we know that if B is computably enumerable in A, then $B \leq_T A'$: more exactly , we can uniformly find an index e of B as a set computably enumerable in $\emptyset^n$, and from this an index $f(e)$ such that $\phi_{f(e)} : B \leq_m A'$. Hence, if $B = W_e^{\emptyset^n}$, then $\phi_{f(e)} : B \leq_m \emptyset^{n+1}$. Lastly, $\mathbb{N} \models Q_0 x_0, ..., Q_n x_n \theta$ iff $0 \in B$ iff $\phi_{f(e)}(0) \in \emptyset^{n+1}$ iff $\langle \phi_{f(e)}(0), n+1 \rangle \in \emptyset^\omega$, namely $Th(\mathbb{N}) \leq_T \emptyset^\omega$.

QED

2.5. Guide for further studies: trial-and-error machines

The positions of Alan Turing and Kurt Gödel appear somewhat paradigmatic in the debate about the relationships between computability and the mind. While reiterating his unconditional admiration for the work of the English mathematician, Gödel attributed a "philosophical error" to Turing, which in his view consisted in the belief that mental procedures cannot go beyond mechanical procedures. Turing's argument rested - according to Gödel - on the assumption that a finite mind is only capable of a finite number of distinguishable states. If we admit an infinity of mental states - says Turing - some of them will be arbitrarily close and therefore confused (see Turing (1936), p. 250), whereas on the contrary, according to Gödel, "the mind, in its use, is not static, but in continuous development":

> Therefore, although at each stage of the mind's development the number of its possible states is finite, there is no reason why this number should not converge to infinity in the course of its development (Letter to Hao Wang, in Wang (1974), p. 325).

A closer examination, however, reveals how Gödel's criticism of Turing is misleading, especially if one takes into account his post-war writing on mind (see Copeland and Shagrin (2013) for a thorough discussion), where Turing began to deepen the idea of *learning*:

> What we want is a machine that can learn from experience. The possibility of letting the machine alter its own instructions provides the mechanism for this. One can imagine that after the machine had been operating for some time, the instructions would have altered out of all recognition (Turing (1947), p. 393).

He conceived a Multi-Machine theory of mind, or the transformation of one Turing machine into another. A machine with the ability to learn is able to modify its table of instructions, transforming itself into a different Turing machine. The notion of mind change was later used by Putnam in developing his notion of "trial and error predicates". Trial-and-error machines where introduced in Putnam (1965) and in Gold (1967), and based on the idea of computability in the limit, i.e. a type of computation performed by an ideal model which proceeds by changing its opinion a finite number of times about the membership of a number to a set, but stabilized to the limit (hence going so far beyond the classical boundaries of computability) with the aim to represent a cognitive phenomenon like language learning and actually these writings had a strong influence on the development of the formal study of the process of gaining information through observation (the *Formal Learning Theory*, see Osherson, Stob and Weinstein (1986) and Kelly (2023)). Some researchers in cognitive science and philosophy of mind, even go so far as to claim that humans are automata of this sort, namely trial-and-error machines (see e.g. Kugel (1986)). The trial-and-error model transcends the Church-Turing thesis; indeed a similar machine solves the *halting-problem*: let for example U be a machine that, when receives as input the code e of another machine ϕ_e and a number n, returns immediately 0 (to say that $\phi_e(n)\uparrow$), then it starts to compute $\phi_e(n)$. If later, at a computation step $\phi_e(n)\downarrow$, U then change its minds and writes 1. Sets calculated from these kind of machines turn out to be the Δ_2, and not only the Δ_1, (i.e. the *recursive* ones). However, it must be emphasised that this is a purely ideal model, since we have no way to know at any given time whether the latest output is the correct output.

In the 1970s, Magari (1974) and Jeroslow (1975) proposed two further formal counterparts of the concept of "trial and error theory"; in particular, Magari's so-called *dialectical sets* form a strict subclass of the Δ_2 sets. Thinking of Gödel's limitative results, Magari's purpose, in the light of Lakatos (1976) dialectical reconstruction of history of mathematics, influenced by Popper's fallibilism, was in particular that of introducing a kind of formal systems, consisting of two actions: on the one hand removing contradictions when they arise, by removing some axioms, and on the other hand adding axioms until they do not give rise to contradictions. In more recent years, this idea has been developed in a series of works (see Amidei, Pianigiani, San Mauro, Simi and Sorbi (2016) and the other articles by these authors, cited in the bibliography).

Here we just want to demonstrate a major achievement in this area, as a starting point for the study of this topic. A set A will be identified by its characteristic function, that is with the infinite binary sequence $A(0), A(1), A(2), ...$, where $A(n) = 1$ iff $n \in A$. We denote $A \upharpoonright y$ the restriction of the characteristic function of A to the initial segment of elements $z < y$, $A(0), A(1), ..., A(y-1)$ To say that a set A is computably enumerable is to say that it has a computable approximation (see p. 15), namely that there exists a computable function f such that $A(x) = \lim_{s\to\infty} f(x,s)$, $f(n,0) = 0$ and for at most a s, we have that $f(n,s) \neq f(n,s+1)$, i.e. the function, changes its mind about n at most once. Note that if we put $f(n,s) = A_s(n)$ this means that that $A_s \subseteq A_{s+1}$ (monotonicity of the sequence). Now we *generalise* this definition, admitting that f may change its mind a *finite number of times* (thus abandoning the *monotonicity*). Let's say that $\{A_s\}_s$ is a Δ_2-approximation of A, if $A = lim_{s\to\infty} A_s$, that is approximated by a computable sequence, where an element can enter and exit a certain finite number of times. Call *modulus of convergence* for such a sequence, a function $m(x)$ such that, for all x and all $s \geq m(x)$, $A_{\upharpoonright x+1} = A_{s\upharpoonright x+1}$. That is

to say, after the stage $m(x)$ the initial segment of A until x does not change any more. In what follows we shall consider in particular the *minimum* modulus:

$$m(x) = \mu(s).\forall t \geq s(A_{\restriction x+1} = A_{t\restriction x+1})$$

Note that if $A = lim_{s\to\infty}A_s$, then $A \leq_T m$; actually $A(x) = A_{m(x)}(x)$. On the other hand, if we consider the computably enumerable set (called *the set of changes*):

$$B = \{\langle x, s\rangle | \exists t > s(A_s(x) \neq A_t(x))\}$$

then clearly $m \leq_T B$. In fact, also $m \leq_T B$ holds. So ultimately $m =_T B$, namely, the minimum modulus has Turing degree computably enumerable.

Lemma 6. (Shoenfield's Limit Lemma 1959) *A is computable in the limit iff $A \leq_T \emptyset'$ (iff, by Post's results, $A \in \Delta_2$).*

Proof. $\Leftarrow$ suppose that $A = \Phi^{\emptyset'}$; let g a function with values $0 - 1$ such that $g(n, s) = 1$ iff $\Phi_s^{\emptyset'}(n) = 1$. Let $z = \mu x.\Phi^{\emptyset'\restriction x}(n)\downarrow$ the so-called *use* of the computation (i.e. the length of the minimal initial segment of the oracle sufficient for computing the function on n). Recall that the *Use Principle* says: if z is the use of $\Phi^A(n)$ and B is such that $B \restriction z = A \restriction z$, then $\Phi^A(n) = \Phi^B(n)$. But $\emptyset'$ is computably enumerable and therefore can be approximated by a sequence $\emptyset'_s \subseteq \emptyset'_{s+1}$. Let therefore s be a sufficiently large stage, for which $\emptyset' \restriction z = \emptyset'_s \restriction z$. For the "Use Principle", this means that $g(n, t) = \Phi^{\emptyset'}(n) = \Phi^{\emptyset'_t}(n)$, for all $t \geq s$. Take $A(x) = \lim_t g(x, t)$.

$\Rightarrow$ Suppose that $A(x) = \lim_t g(x, t)$, where $g(x, 0) = 0$. Notes that this means that, if I count the number k of changes of opinion by g, the function returns value 1 only when k is odd. Consider now the set B of pairs $\langle n, k\rangle$ such that the numbers s of stages in which $g(n, s) \neq g(n, s + 1)$ is greater or equal to k and note that it is computably enumerable (therefore $B \leq_T \emptyset'$). Consider a Φ^B working as follows. On input n, look for the minimum k such that $\langle n, k\rangle \notin B$: if you find it, output 0 in case k is even, and output 1 in case is odd. Note that $A(x) = \Phi^B(x)$ and therefore $A(x) = \Phi^{\emptyset'}(x)$. QED

3. Church's formal system of lambda-calculus

3.1. Models of computation: introduction to β-reduction

A. Church in 1936 introduced a formal system based on the operations of function abstraction and application, called *the lambda calculus* and defined the notion of computable function in this system. Church's original goal was to construct a formal system for the foundations of mathematics based on functions together with a set of logical notions. When this system was discovered to be inconsistent, Church then separated out the consistent subsystem that is now called lambda calculus and concentrated on it. The language of untyped λ-calculus consists of an infinite set of variables, the abstraction operator λ, parentheses, the application operator $\cdot$ (usually$(t \cdot s)$ is written (ts)). If there is also a set of constants, the calculus is *applied*. We will deal with calculus without constants, called *pure*.

Terms. The variables are terms; if t, s are terms, also (ts) is a term. If t is a term and x is a variable, then $(\lambda x.t)$ is a term. We shorten $(((...(ts)r)...)p)$ with $tsr...p$, and we shorten $(\lambda x(\lambda y(...(\lambda z.(t)))))$ with $\lambda xy...z.t$. Moreover, to avoid confusion with other types of equality, we still use the symbol $=$ to denote the meta-linguistic syntactic identity. Following Barendregt (1984) p. 26 we make this convention:

1. Terms so-called $\alpha-$equivalent, i.e. such that $\lambda x.t =_\alpha \lambda y.t[y/x]$ are identified.

2. We will always assume without loss of generality that in a certain context (proof, definition ecc.) bounded variables have been renamed to be distinct, and distinct from free variables.

This convention is called *Barendregt's variables convention*. Thanks to it, the substitution operation can be defined quite simply as follows:

1. $x[s/x] = s$
2. $y[s/x] = y \;\; x \neq y$
3. $(tr)[s/x] = t[s/x]r[s/x]$
4. $(\lambda x.t)[s/x] = \lambda x.t$
5. $(\lambda y.t)[s/x] = \lambda y.(t[s/x])$

Lemma 7. (Substitution lemma) *Let $x \neq y$ and x is not among the free variables of s. Then:*

$$t[r/x][s/y] = t[s/y][r[s/y]/x]$$

Proof. Induction on t. For instance, if $t = \lambda z.V$, where z is a new variable, for the above convention; from the definition we have:

$$(\lambda z.V)[r/x][s/y] = \lambda z.(V[r/x][s/y]) =$$

$$= \lambda z.(V[s/y][r[s/y]/x]) \;\; (inductive \;\; hypot.) =$$

$$= (\lambda z.V)[s/y][r[s/y]/x]$$

Exercise. Verify the other cases. If t is a variable (a) if $t = x$, the we have $r[s/y] = r[s/y]$; (b) if $t = y$, then we get $s = s$ and (c) if $t = z$ variable different from y and from y we get $z = z$. QED

Duccio Pianigiani, University of Siena, Italy, duccio.pianigiani@unisi.it, 0000-0001-9441-7226

Referee List (DOI 10.36253/fup_referee_list)
FUP Best Practice in Scholarly Publishing (DOI 10.36253/fup_best_practice)

Book References DOI 10.36253/979-12-215-0778-2.references

For *redex* we mean a term like $(\lambda x.t)s$; it can be simplified by repalcing all occurence x in t, with s, that we denote $t[s/x]$ (its *contractum*); the fundamental operation of the λ-calculus, called β-contraction, is the following:

$$(\lambda x.t)s \to_\beta t[s/x]$$

We write $t \to_1 s$, meaning that s is obtained by contracting a single redex of t. We say that t *converts in a term* s, if the latter results from a finite number of $1-$contractions and inverse $1-$contractions (retractions) from t. A term is in *normal form*, if it cannot be further reduced, because it does not contain any *redex*. shall we say that the "reduction", that we denote with $\Rightarrow$, constitutes the reflexive and transitive closure of $\to_1$, while the "conversion" $=$ is the reflexive, transitive and symmetric closure. We get an *extensional* calculus, by adding the η-rule $\lambda x.Ux = U$, for $x \notin FVar(U)$. Observe that the equality between functions is extensional:

$$\forall x(f(x) = g(x)) \to f = g$$

But if $f = y$ and $g = \lambda x.yx$, we get $fU = gU$. However $f \neq g$. The η-rule allows to overcome the obstacle.

The following is easily shown.

Lemma 8. *The following holds:*

1. *if* $t \Rightarrow s$, *then* $r[t/x] \Rightarrow r[s/x]$.
2. *if* $t = s$ *then* $t[r/x] = s[r/x]$.
3. *if* $t = s$ *then* $r[t/x] = r[s/x]$.
4. *if* $t \to_1 s$ *then* $t[r/x] \to_1 s[r/x]$

The original theory introduced by Church (called λ_I-calculus) was different from that we have illustrated, called λ_{K}-calculus. A term of the λ_{I}-calculus has the property that in all its subterms of the form $\lambda x.s$, the variable x must occur free in s at least once. Hence $\mathsf{K} = \lambda xy.x$ is not a well-formed term of this language. To highlight the main difference, consider the term $\mathsf{KI}\Omega$ of the λ_K-calculus, where $\mathsf{I} = \lambda x.x$ and $\Omega = (\lambda x.xx)(\lambda x.xx)$. Depending on the order in which we perform the reduction, the term $\mathsf{KI}\Omega$ can be reduced either to normal form I, or produce an infinite reduction $\mathsf{KI}\Omega \to \mathsf{KI}\Omega \to$ ..... This cannot happen in Church's original calculus, where a term t has normal form iff each its subterm also has normal form.

Confluence. A binary relation $\prec$ is called *confluent* iff:

$$\forall xyz\exists w(x \prec y \wedge x \prec z \Rightarrow y \prec w \wedge z \prec w)$$

We claim that the relation $\Rightarrow$ is confluent and we show this beginning, as is usually done, by demonstrating that it is fulfilled by the relation of *parallel reduction* due to Tait and Martin-Löf (see for instance Barendregt (1984) pp. 60-3), but in the elegant simplified version of Takahashi (1995). We cannot start with the one-reduction because it is known not to be confluent.

Definition 15. (Parallel reduction)

1. $x \to_p x$
2. $\dfrac{t \to_p s}{\lambda x.t \to_p \lambda x.s}$
3. $\dfrac{t \to_p s \quad r \to_p w}{tr \to_p sw}$
4. $\dfrac{t \to_p s \quad r \to_p w}{(\lambda x.t)r \to_p s[w/x]}$

In presence of η-rule, add:

$$\frac{t \to_p s}{\lambda x.tx \to_p s}$$

We show that there exists a term t^T, dependent from t, but not from s, such that if $t \to_p s$, then $s \to_p t^T$ and follows the confluence for $\to_p$. Since $\Rightarrow$ is the transitive closure of $\to_p$, it follows the confluence also for this relation.

Definition 16. *The term t^T is inductively defined as follows:*

1. $x^T = x$
2. $(\lambda x.t)^T = \lambda x.(t)^T$
3. $(rs)^T = r^T s^T \ \ (rs \ \ non \ \ redex)$
4. $((\lambda x.t)s)^T = t^T[s^T/x]$

Theorem 34. *If $t \to_p s$, then $s \to_p t^T$.*

Proof. Induction on the complexity of t.

1. If $t = x$, i.e. $x \to_p s$, then $s = x = t^T$.
2. If $t = \lambda x.r$ and $\lambda x.r \to_p s$, then take $s = \lambda x.u$, where $r \to_p u$ and apply the induction hypothesis, obtaining $u \to_p r^T$, from which $\lambda x.u \to_p \lambda x.r^T$; take finally $t^T = \lambda x.r^T$.
3. If $t = ru$ and $ru \Rightarrow s$ (ru non redex), this means that $s = r'u'$ and $r \to_p r'$, $u \to_p u'$. By induction hypothesis there are r^T and u^T such that $r'u' \to_p t^T$, where $t^T = r^T u^T$.
4. If $t = (\lambda x.r)u \to_p s$ we will have these possibilities: either $s = (\lambda x.r')u'$, or $s = r'[u'/x]$. In both cases $r \to_p r'$ and $u \to_p u'$. By induction hypothesis $r' \to_p r^T$ and $u' \to_p u^T$. In the first case take $t^T = r^T[u^T/x]$. Clearly $s \to_p t^T$. Similarly in the second case.

QED

Corollary 5. *If a normal form for t exists, this is unique.*

Proof. If there were two n_1, n_2, where $t \Rightarrow n_1$ and $t \Rightarrow n_2$, for Church-Rosser there will be a term s to which both n_1 and n_2 converge: being in normal form we will have s, n_1 and n_2 are the same term. QED

In the type-free calculus we have *fixed points* operators. A fixed point operator is a term Fix such that, for all term s, Fix$s = s$(Fix$s$). For instance:

$$\mathcal{Y} = \lambda f.(\lambda x.f(xx))(\lambda x.f(xx))$$

is a fixed point. Observe that, if $\mathcal{W} = \lambda x.F(xx)$, we have :

$$\mathcal{Y}F \Rightarrow \mathcal{W}\mathcal{W} \Rightarrow F(\mathcal{W}\mathcal{W}) = F(\mathcal{Y}F)$$

For instance, if we want to build a term F such that $FXY = FYXF$, we can argue as follows: we need an F such that $F = (\lambda fxy.fxyf)F$; therefore we define $F = \mathcal{Y}(\lambda fxy.fxyf)$. In general we have the following.

Theorem 35. (Fixed point theorem) *Given a term $\mathcal{C}[f, x]$, with free variables f and x, there is a term F such that for all T, $FT = \mathcal{C}[F, T]$*

Proof. As we have seen, take $F = \mathcal{Y}(\lambda fx.\mathcal{C}[f, x]))$.

To ensure that those terms, defines the required function, we reason as follows: let $W = \lambda fxy.fxyf$ (and then $\mathcal{Y}W = W(\mathcal{Y}W)$); then $F = \mathcal{Y}(W) = W(\mathcal{Y}W) = (\lambda fxy.fxyf)(\mathcal{Y}W) = (\lambda fxy.fxyf)F$. QED

Exercise. Let $A = \lambda xy.y(xxy)$ and $\Theta = AA$; show that Θ is a fixed point operator. It is observed that, while it is not valid $\mathcal{Y} \Rightarrow f(\mathcal{Y}f)$, It is valid instead $\Theta \Rightarrow f(\Theta f)$.

Remark 3. *A fundamental difference between the calculus without types and the calculus with types is that in the former the conversion is undecidable, whereas in the latter (as satisfying normalisation) it is decidable.*

The lambda calculus can be presented as an equational theory; we introduce therefore the theory $\widetilde{\lambda}$, defined by the following axioms and rules:

1. $\lambda x.U = \lambda y.U[y/x]$, $y \notin FVar(U)$
2. $(\lambda x.U)V = U[V/x]$
3. $U = U$
4. $\dfrac{U = V}{ZU = ZV}$
5. $\dfrac{U = V}{UZ = VZ}$
6. $\dfrac{U = V \quad V = Z}{U = Z}$
7. $\dfrac{U = V}{V = U}$
8. $\dfrac{U = V}{\lambda x.U = \lambda x.V}$

So, we want to show taht the theory $\widetilde{\lambda}$ is undecidable. Recall that two sets X, Y are called *recursively separable* if there is a recursive set $\mathcal{C}$ such that $X \subseteq \mathcal{C}$ and $Y \cap \mathcal{C} = \emptyset$. A pair of sets of terms is called recursively separable iff the corresponding sets of Gödel numbers are recursively separable. We say that a set of λ-terms X is *closed* with respect to β- conversion, iff $t \in X$ *and* $t = r$ *implies* $r \in X$.

Second fixed point theorem. We write, to simplify the notation, $X^{\#}$, meaning by this $\overline{\ulcorner X \urcorner}$, the numeral of Gödel number of X and we observe that these functions are definable by terms of our calculus $APP(M^{\#}N^{\#}) = (MN)^{\#}$, $NUM(M^{\#}) = (M^{\#})^{\#}$.

Theorem 36. *Let* $W = \lambda x.F(APPx(NUMx))$ *and* $\widetilde{X} = WW^{\#}$*; then*

$$\widetilde{X} = WW^{\#} = F(APPW^{\#}(NUM(W^{\#})) = F(WW^{\#})^{\#} = F\widetilde{X}^{\#}$$

Hence: $\forall F \exists X : FX^{\#} = X$

Proof. Immediate. QED

Theorem 37. (Scott and Curry) *Let* $X, Y \neq \emptyset$ *sets of terms, closed with respect to* β*-conversion. Hence* X, Y *are not recursively separable.*

Proof. Let $t \in X$, $r \in Y$. Let $\mathcal{C}$ a recursive set such that $X \subseteq \mathcal{C}$ e $Y \cap \mathcal{C} = \emptyset$. Let F the term that defines the characteristic function of $\mathcal{C}$, and then $F^{\overline{\ulcorner u \urcorner}} = \overline{0}$ if $u \in \mathcal{C}$ and $F^{\overline{\ulcorner u \urcorner}} = \overline{1}$ if $u \notin \mathcal{C}$. Let finally:

$$\mathcal{G} = \lambda x.\delta(ZERO(Fx))rt$$

where $ZERO\overline{n}$ is the test for zero (see below at p. 65) which returns $\lambda xy.x$, if $n = 0$, and $\lambda xy.y$, if $n > 0$. Observe that, if $u \in \mathcal{C}$, then $\mathcal{G}^{\overline{\ulcorner u \urcorner}} = r$; if instead $u \notin \mathcal{C}$, then $\mathcal{G}^{\overline{\ulcorner u \urcorner}} = t$. But for the previous fixed point theorem, there exists X such that $\mathcal{G}^{\overline{\ulcorner X \urcorner}} = X$. Hence we finally have:

* $X \in \mathcal{C} \Rightarrow X = \mathcal{G}^{\overline{\ulcorner X \urcorner}} = r \Rightarrow X \notin \mathcal{C}$, because $r \in Y$.
* $X \notin \mathcal{C} \Rightarrow X = \mathcal{G}^{\overline{\ulcorner X \urcorner}} = t \Rightarrow X \in \mathcal{C}$, because $t \in X$. QED

Corollary 6. *If* X *is closed for the relation "=", then it is not recursive.*

Proof. In previous theorem take $Y = \overline{X}$ (the complement of X); it follows that if X is closed for β-conversion, then X is not recursive. QED

Corollary 7. *The set* X_{nf} *of terms that have a normal form is recursively enumerabile, but not recursive.*

Proof. Recall that a set is computably enumerable iff can be defined by a formula of the form $\exists x P$, for P recursive. Hence X_{nf} is recursively enumerabile because M has a normal form iff $\exists U \exists p$("U is in normal form and p is a proof of $M = U$"). But it is not recursive (*Exercise*). This is one of the first examples of computably enumerable but not recursive sets. QED

Corollary 8. *The relation "=" is undecidable.*

Proof. Take $X = \{r | \vdash r = (\lambda x.x)\}$. It is closed for β-conversion and therefore not recursive. QED

It follows the undecidability of the *Entscheidungsproblem*: When terms are given Gödel numbers, then "=" corresponds to a relation between natural numbers, i.e. once goedelized, terms become numbers and "=" become a relation between numbers; the numbers can also be represented in a very simple predicative language, for example $0 = x, 1 = f(x), 2 = f(f(x))$.... Axioms of the equational theory of $\lambda - calculus$ can thus be translated into formulas $\alpha_0, ..., \alpha_n$ of the predicate calculus, in such a way that $(\alpha_0 \wedge ... \wedge \alpha_n) \rightarrow E(\overline{m}, \overline{n})$ is provable in the predicate calculus, if $m = \ulcorner s \urcorner$, $n = \ulcorner t \urcorner$ e $t = s$. But if we could decide this formula, we could also decide "=". Indeed, if we could decide all questions of provability in pure predicate logic, then we could decide whether arbitrary lambda-terms are convertible.

Theorem 38. (Böhm's separability theorem 1968) *It t and s are $\beta\eta$-normal forms, and u,v are arbitrary terms, the it is possible to construct terms $r_0, ..., r_n$ and find variables $x_0, ..., x_k$ such that $\lambda x_0 ... \lambda x_k . t r_0 ... r_n = u$ and $\lambda x_0 ... \lambda x_k . s r_0 ... r_n = v$.*

Proof. (See Krivine (1993), p. 88) QED

A consequence is that if $t = s$ is not provable and it is added to the calculus, the calculus crashes.

Let us now begin to assess the expressive power of this calculus. There are many ways in which we can represent numbers, using appropriate terms ("numerals"), in λ-calculus. For instance, these are the *Church numerals*:

$$\overline{n} =_{def} \lambda xy. \overbrace{x(x(x(...(xy)...)}^{n-times}$$

with these numerals, we can define the basic arithmetic operations as follows:

1. $(sum) \quad \oplus = \lambda xyzu.xz(yzu)$
2. $(multiplication) \quad \otimes = \lambda xyz.x(yz)$
3. $(exponentiation) \quad EXP = \lambda xy.yx$

If we abbreviate $\overbrace{x(x(x(...(xy)...)))}^{n-times}$ with $x^n y$, we see for example that:

$$\overline{n}uz = (\lambda wv.w^n v)uz = (\lambda v.u^n v)z = u^n z$$

from which follows:

$$\oplus \overline{n}\overline{m} = (\lambda xyzu.xz(yzu))\overline{n}\overline{m} = \lambda zu.\overline{n}z(\overline{m}zu) =$$
$$\lambda xyzu.z^n(z^m u) = \lambda zu.(z^{m+n}u) = \overline{m+n}$$

The reader can verify by exercise that with respect to Church numerals, the followings are respectively the predecessor function and the test for zero:

1. $P = \lambda xfy.x(\lambda pq.q(pf))(\mathsf{K}y)\mathsf{I}$, $\mathsf{K} = \lambda xy.x$ and $\mathsf{I} = \lambda x.x$.
2. $Zero = \lambda x.x(\lambda y.\mathsf{K}^*)\mathsf{K}$, where $\mathsf{K}^* = \lambda xy.y$.
3. $\top = \mathsf{K}$, $\bot = \mathsf{K}^*$.

That of Church is not the only system of numerals. More abstractly a system of numerals is a sequence $N = n_0, n_1, n_2...$ of closed terms such that there exist terms S and Zero for which it $Sn_k = n_{k+1}$, $Zero\ n_0 = \top$, $Zero\ n_{k+1} = \bot$ (where $\top$ and $\bot$ are truth-values, see below).

Definition 17. *A partial recursive function ψ is definable with respect to N, iff there exists a term F of the lambda claculus such that for all $r_0, ..., r_k$,*

$$F n_{r_0} ... n_{r_0} = n_{\psi(r_0,...,r_k)}$$

We say that N is adequate, if with respect to it we can define all partial recursive functions; equivalently: if we can define the predecessor $P(n_{k+1}) = n_k$.

Here we choose the approach to the numbers and functions of the classical Barendregt (1984).

1. *Booleans.* $\top = \lambda xy.x = \mathsf{K}$, $\bot = \lambda xy.y = \mathsf{K}^*$
2. *Discriminator.* An operator δXPQ such that:

$$\delta XPQ = \begin{cases} P & \text{if } X = \top \\ Q & \text{if } X = \bot \end{cases}$$

   Take $\delta = \lambda xyz.xyz$. Actually $\top PQ = XPQ = P$, $\bot PQ = XPQ = Q$.
3. *Pair and projections.* $\langle P, Q\rangle = \lambda x.xPQ$, $\pi_0 = \lambda xy.x = \mathsf{K}$, $\pi_1 = \lambda xy.y = \mathsf{K}^*$.
   Clearly $\langle P, Q\rangle\pi_0 = P$ and $\langle P, Q\rangle\pi_1 = Q$
4. *Barendregt's numerals.* $\overline{0} = \lambda x.x$, $\overline{n+1} = \langle\bot, \overline{n}\rangle$. For instance $\overline{3} = \langle\bot, \overline{2}\rangle = \langle\bot, \langle\bot, \overline{1}\rangle\rangle = \langle\bot, \langle\bot, \langle\bot, \overline{0}\rangle\rangle\rangle$
5. *Successor.* $S\overline{n} = (\lambda x.\langle\bot, x\rangle)\overline{n}$
6. *Predecessor.* $P\overline{n} = (\lambda x.x\bot)\overline{n}$
7. *Test for zero.* $ZERO = \lambda x.x\top$.

   Observe that

$$\mathsf{ZERO}\overline{0} = (\lambda x.x\top)\overline{0} = \top$$

$$\mathsf{ZERO}\overline{n+1} = (\lambda x.x\top)\langle\bot, \overline{n}\rangle = \bot$$

Recall that in general $\langle X\pi_0, X\pi_1\rangle = X$ does not hold; pair is not "surjective" (C.Mann 1972). We now obtain the *total* recursive functions in this way:

1. *Initial functions.* (*projections*) $U_i = \lambda x_0, ..., x_p.x_i$ (*successore*) $\lambda x.\langle\bot, x\rangle$ (*zero*) $\lambda x.\overline{0}$
2. *Closure under composition.* Let $g, h_0, .., h_m$ be defined by terms $G, H_0, ..., H_m$; we define composition as followe:

$$F = \lambda\overrightarrow{x}.G(H_0\overrightarrow{x})...(H_m\overrightarrow{x})$$

3. *Closure under primitive recursion.* Suppose that:

   (a) $f(0, n) = g(n)$,

   (b) $f(k+1, n) = h(f(k, n), k, n)$

   and let h, g be defined respectively by H, G.
   By *Fixed point theorem*, given a term $\mathcal{C}[f, x]$, there is a term F such that for all T, $FT = \mathcal{C}[F, T]$. As we have seen, we can take $F = \mathcal{Y}(\lambda fx.\mathcal{C}[f, x]))$. Hence, applying this theorem, we have that we can define F such that:

$$Fx\overrightarrow{y} = \delta(\mathsf{ZERO}x)(G\overrightarrow{y})(H(F(Px)\overrightarrow{y})(Px)\overrightarrow{y})$$

4. *Minimalization.* Let $f(n) = \mu m(g(n, m) = 0)$ and let g be defined by G. Let then:

$$Hxy = \delta(\mathsf{ZERO}(Gxy)y(Hx(Sy))$$

   Let finally $F = \lambda x.Hx\overline{0}$. Observe that $F\overline{n} = \overline{0}$, if $G\overline{n}\overline{0} = \overline{0}$ and $F\overline{n} = H\overline{n}\overline{1}$ otherwise; $F\overline{n} = \overline{1}$, if $G\overline{n}\overline{1} = \overline{0}$ and $F\overline{n} = H\overline{n}\overline{2}$ otherwise, etc. if moreover $G\overline{n}\overline{2} = \overline{0}$, then:

$$H\overline{n}\overline{0} = H\overline{n}\overline{1} = H\overline{n}\overline{2} = \delta(\mathsf{ZERO}(G\overline{n}\overline{2})\overline{2}(H\overline{n}\overline{2}) = \overline{2}$$

Let us now address the problem of of representing *partial* functions. We will say that ψ is $\lambda-$definible with respect to a set of terms X, if there exists a term F of the $\lambda - calculus$ such that:

1. if $\psi(n)\downarrow$, then $F\overline{n} = \overline{\psi(n)}$
2. If $\psi(n)\uparrow$, then $F\overline{n} \in X$

The elements of X constitutes the formalization of the concept of "meaningless" terms. But what set X we have to choose? Various purposes are admitted. Church took the terms without normal form, for instance $\Omega = (\lambda x.xx)(\lambda x.xx)$. Barendregt take the unsolvable terms: a closed term M is *solvable*, iff for some n and terms $T_0, ...T_n$, $MT_0, ..., T_n = (\lambda x.x)$. There are terms without normal form but with *head normal-form* $\lambda x_0...\lambda x_n.yM_0...M_n$. A term does not convey any information, if not even possess this normal form. We will see that a term is solvable if and only if it has a head normal form. In this regard, there are general results.

Theorem 39. (Statman 1970) *A set of closed terms X is said "co-Visser set", if $\overline{X}$ is computably enumerable and it is closed for $=$, i.e. if $t \in \overline{X}$ and $t = s$, then $s \in \overline{X}$. If X is a "co-Visser set", then all partial recursive functions can be defined using it as a representative of undefined terms.*

Proposals of Church and Barendregt satisfy this theorem (see Barendregt (1992)). The two different choices by these logicians are actually related:

1. in λ_I calculus: meaningless=unsolvable= no normal form.

2. In λ_K calculus: meaningless=unsolvable= no head normal form.

3.2. Solvability and head normal forms

Let us begin by defining more formally the notion of "solvable term".

Definition 18. *A closed term t is called* solvable *iff there exists a number n and terms $s_0, ..., s_n$ such that $ts_0, ..., s_n = \mathsf{I}$. An arbitrary term $t(x_0, ..., x_k)$ is solvable iff its closure $\lambda x_0...x_k.t$ is solvable.*

For instance, if $\mathcal{Y}$ is the above-mentioned fixed point operator, then it is solvable: $\mathcal{Y}(\mathsf{KI}) = (\mathsf{KI})(\mathcal{Y}(\mathsf{KI})) = \mathsf{I}$.

Lemma 9. *Let t be a term: t is solvable iff there exists a closed substitution instance t^* of it and closed terms $s_0, ..., s_n$ such that $t^*s_0, ..., s_n = \mathsf{I}$.*

Proof. $\Rightarrow$ Consider the closure $\lambda x_0...\lambda x_k.t$, that, by hypothesis, is solvable, i.e. for some $s_0, ..., s_n$, $(\lambda x_0...\lambda x_k.t)s_0, ..., s_n$ (assume $n > k$, perhaps by adding copies of I), Hence $t[s_0/x_0, ..., s_k/x_k]s_{k+1}...s_n = \mathsf{I}$, Put $t^* = t[s_0/x_0, ..., s_k/x_k]$.
$\Leftarrow$ If $t^* = t[s_0/x_0, ..., s_k/x_k]$ is solvable, then clearly also $\lambda x_0...\lambda x_k.t$ is solvable. QED

Lemma 10. *t is solvable iff $\lambda x_0.t$ is solvable.*

Proof. (Sketch) Let $\lambda x_0...\lambda x_k.t$ be the closure of t. By previous lemma t is solvable iff there exist closed terms $s_0, ..., s_k, s_{k+1}, ..., s_n$ such that:

$$t[s_0/x_0, ..., s_k/x_k]s_{k+1}...s_n = \mathsf{I}$$

iff

$$((\lambda x_0.t)s_0)[s_1/x_1, ..., s_k/x_k]s_{k+1}, ..., s_n = \mathsf{I}$$

iff

$$(\lambda x_0.t)[s_1/x_1, ..., s_k/x_k]s_0 s_{k+1}, ..., s_n = \mathsf{I}$$

iff $\lambda x_0.t$ is solvable. QED

Corollary 9. *t is unsolvable iff for all s, we have that ts, $t[s/x]$, $\lambda x.t$ are unsolvable.*

One can prove by an easy induction that each term is either of the form

1. $\lambda x_0 ... \lambda x_n . y t_0 ... t_k$, or of the form
2. $\lambda x_0 ... \lambda x_n . (\lambda y . s) r t_0 ... t_n$

where the sequences $x_0, ..., x_n$ and $t_0, ..., t_k$ might be empty. The redex $(\lambda y.s)r$ is called *head redex.* A contraction of a term t to a term u that eliminates a head redex is called *head reduction* and denoted $t \to_h u$. If in a reduction sequence $t_0 \to_h t_1 \to_h t_2 \to ...$ there is a term t_n of the form $\lambda x_0 ... \lambda_k . y s_0 ... s_m$ (called *head normal form* h.n.f.) such a reduction sequence terminates at t_n. If $t = s$ and s is in h.n.f. then we say that t *has* an h.n.f. It holds that t *has* an h.n.f. iff its head reduction path of terminates.

Lemma 11. *$\lambda x.t$ has an h.n.f. iff t has an h.n.f.*

Proof. Clearly if $\lambda x.t \Rightarrow s$ then s has the form $\lambda x.u$, where $t \Rightarrow u$. QED

Lemma 12. *If $t[s/x]$ has an h.n.f. also t has an h.n.f.*

Proof. Suppose on the contrary that t has no head normal form. It is not hard to show that the following substitution holds: if $t \to_h s$, then $t[r/x] \to_h s[r/x]$, hence $t[s/x]$ has an infinite head-reduction path and therefore has no h.n.f. itself. QED

Theorem 40. *If ts has an h.n.f. then t has an h.n.f too.*

Proof. Let $t \to_h t_0 \to_h t_1 \to_h t_2 ...$ be the head reduction of t and consider these cases:

1. If no term of this sequence has form $\lambda x.u$, the an head-reduction of ts has the form $ts \to_h t_0 s \to_h t_1 s \to_h t_2 s ...$ and by hypothesis it is finite; hence also $t \to_h t_0 \to_h t_1 \to_h t_2 ...$ is finite and then t has an h.n.f.
2. If some term of this sequence has form $\lambda x.u$, let t_k the first of such terms. Hence the head reduction has the form:

$$ts \to_h t_0 s \to_h t_1 s \to_h t_2 s ... \to_h t_{k-1} s \to_h (\lambda x.u)s \to_h u[s/x] \to_h ...$$

However, if ts has an h.n.f. also $u[s/x]$ has an h.n.f., as well as u and $\lambda x.u$ and t, by previous lemmas. QED

Theorem 41. *A term t has an h.n.f. iff its head reduction path terminates.*

Proof. $\Rightarrow$ Recall that the Church-Rosser theorem holds also for β-equality (conversion), by "composing the boxes": if $t = s$ then there exists r such that $t \Rightarrow r$ and $s \Rightarrow r$. Let us suppose that $t = \lambda x.yU$. Hence, by Church-Rosser theorem there exists z such that both $t \Rightarrow z$ and $\lambda x.yU \Rightarrow z$. Hence z must be of the form $\lambda x.yV$ where $U \Rightarrow V$ and by the *Standardization Theorem* at p. 70 there will be a standard reduction $t = t_0 \to^{\delta_0} t_1 \to^{\delta_1} t_2 \to ... \to \lambda x.yV$. Now, if all redexes δ_j contracted in this reduction are head redexes, then this is a terminating head reduction. Otherwise, let δ_i the first *internal* redex reduced at step $t_i \to^{\delta_i} t_{i+1}$. Then t_i must be in head normal form, because otherwise the head redex would remain: but the reduction of t_i to $\lambda x.yV$ in this case would not be standard. Then $t \to_h ... \to_h t_i$ is a terminating head reduction.

$\Leftarrow$ Trivial.

QED

Theorem 42. (Wadsworth 1971) *The term t is solvable iff it has an h.n.f.*

Proof. By previous results being solvable is true for a term as well as for its closure. The same equivalence holds for the property of having a head normal form. Hence we can assume that t is closed. $\Rightarrow$ If $ts_0 ... s_n = \mathsf{I}$ then I is its h.n.f. Hence by previous lemmas t also has itself an h.n.f. $\Leftarrow$ If t has an head normal form $\lambda x_0 ... \lambda x_n . y t_0 ... t_k$, let $y = x_i$. Observe that:

$$(\lambda x_0 ... \lambda x_n . y t_0 ... t_k) \overbrace{(\mathsf{K}^{k+1}\mathsf{I})(\mathsf{K}^{k+1}\mathsf{I})...(\mathsf{K}^{k+1}\mathsf{I})}^{n+1-times}$$

reduces to $(\mathsf{K}^{k+1}\mathsf{I})t_0^* ... t_k^*$ and therefore to I ($\mathsf{K}^m\mathsf{I}$ abbreviates $\mathsf{K}(\mathsf{K}(\mathsf{K}(...(\mathsf{KI})...)))$ m-times). QED

Corollary 10. *Unsolvable terms have no normal form.*

Proof. If t is in normal form, then has also an head normal form and therfore is solvable. QED

Theorem 43. *A term t is unsolvable iff it is hereditarily without normal form, i.e. for all substitution t^* and all $s_0, ..., s_n$, $t^*s_0...s_n$ ha no normal form.*

Proof. $\Rightarrow$ If $t^*s_0...s_n$ has a normal form, then it is solvable. Therefore there exists a closed substitution instance $(t^*s_0...s_n)^*$ of it and terms $r_0, ..., r_k$ such that $(t^*s_0...s_n)^*r_0, ..., r_k = \mathsf{I}$. Thus $(t^{**}s_0^*...s_n^*)r_0, ..., r_k = \mathsf{I}$ and therefore t is solvable. $\Leftarrow$ If t is solvable, then $(t^*s_0...s_n) = \mathsf{I}$ for some $s_0...s_n$. QED

Definition 19. *A partial recursive function ϕ is lambda-definable, iff there exists a term F such that for all natural numbers n:*

$$F\overline{n} = \begin{cases} \overline{\phi(n)} & \text{if } \phi(n) \downarrow \\ no\ h.n.f & \text{otherwise} \end{cases}$$

Lemma 13. (Solvability of numerals) $\overline{m}\mathsf{KII} = \mathsf{I}$.

Proof. Recall that $\overline{0} = \mathsf{I}$ and $\overline{n+1} = \langle \mathsf{K}^*, \overline{n} \rangle$. Hence $\overline{0}\mathsf{KII} = \mathsf{IKII} = \mathsf{I}$.
Moreover $\overline{n+1}\mathsf{KII} = \langle \mathsf{K}^*, \overline{n} \rangle \mathsf{KII} = \mathsf{KK}^*\overline{n}\mathsf{II} = \mathsf{K}^*\mathsf{II} = \mathsf{I}$. QED

Lemma 14. *If F defines a partial function ϕ, then:*

1. *$F\overline{n}\mathsf{KII} = \mathsf{I}$, if $\phi(n) \downarrow$.*
2. *$F\overline{n}\mathsf{KII}$ is unsolvable, otherwise.*

Proof. If $\phi(n) \downarrow$, then $F\overline{n} = \overline{\phi(n)}$; but $\overline{\phi(n)}\mathsf{KII} = \mathsf{I}$. If $\phi(n) \uparrow$, then $F\overline{n}$ is unsolvable. Then it $F\overline{n}s_0...s_n$ is unsolvable for all $s_0, ..., s_n$. QED

Theorem 44. (Closure under composition) *Suppose:*

$$\phi(n) \simeq \chi(\psi_0(n), ..., \psi_m(n))$$

and suppose that $G, H_0, ..., H_m$ respectively define $\chi, \psi_0, ..., \psi_m$. Then:

$$F = \lambda x(H_0 x\mathsf{KII})...(H_m x\mathsf{KII})(G(H_0 x)...(H_m x))$$

defines ϕ.

Proof. If some $\psi_i(n) \uparrow$, then $H_i\overline{n}\mathsf{KII}$ (the "jamming factor") is unsolvable, by the previous lemma, and therefore the whole term F is unsolvable. In case all $\psi_i(n) \downarrow$, then all $H_i\overline{n}\mathsf{KII} = \overline{\psi_i(n)}\mathsf{KII} = \mathsf{I}$ and therefore $F\overline{n} = (G(H_0\overline{n}...H_m\overline{n}))$. QED

To prove closure under minimisation, we must now focus on reduction strategies. We will show in particular that leftmost reductions are *normalizing*, that is, when t has a normal form, then there exists n such that applying n-times the strategy to it leads to the normal form.

Definition 20. (Leftmost reduction) *Call the lambda of a redex $(\lambda x.t)s$ the first occurrence of λ in it. Let δ_0, δ_1 be two occurrences of redex in t. We say that δ_0 is to the left of δ_1, if the λ of δ_0 is to the left of the λ of δ_1. A redex occurrence si called the leftmost redex of a term, if it is to the left of all other redexes.*

Example. In this case δ_0 is the *leftmost*:

$$\underbrace{(\lambda x.\overbrace{(\lambda z.z)x}^{\delta_1}\, x)u}_{\delta_0}(\lambda u.V)$$

To sum up, we write:

1. $t \to_l s$ if s is obtained by contraction of the leftmost redex.
2. $t \to_h s$ if s is obtained by contraction of the head redex.
3. $t \to_i s$ if s is obtained by contraction of an internal redex (we will write $t \to_{pi} s$ for a contraction, both parallel and internal).

The Standardization theorem. Let us call a number p the *position* of a redex δ_i in the reduction step $t_i \to^{\delta_i} t_{i+1}$, if the first symbol of δ_i in t_i is the $p-th$, starting from the left. We denote $p = p_i$. Then the reduction

$$t_0 \to^{\delta_0} t_1 \to^{\delta_1} t_2 \to^{\delta_2} t_3 ....$$

is standard if $p_0 \leq p_1 \leq p_2 \leq ...$, i.e. the sequence of redexes contracted in the reduction moves from left to right. In other words, in this reduction sequence, once a redex is contracted, all redexes whose first symbol is at its left become "frozen". For example, this is a standard reduction:

$$\lambda a.\overbrace{(\lambda b.(\lambda c.c)bb)d}^{1} \to \lambda a.\overbrace{(\lambda c.c)d}^{2}\, d \to \lambda a.dd$$

On the contrary, this is not a standard reduction:

$$\lambda a.(\lambda b.\overbrace{(\lambda c.c)b}^{1}\, b)d \to \lambda a.\overbrace{(\lambda b.bb)d}^{2} \to \lambda a.dd$$

In particular, leftmost reductions $\to_\ell$ and head reductions $\to_h$ are standard.

Theorem 45. (Standardization theorem) *If $t \Rightarrow s$, then there is a standard reduction from t to s.*

We omit the proof, which is performed with the same techniques we will use later in this paragraph (see Sørensen, Urzyczyin (2006) pp. 13-7 and especially Takahashi (1995) p. 124).

Lemma 15. *If there is a standard reduction of t to s and s is in normal form, then all its reductions are leftmost.*

Proof. Suppose that t is reduced to s in standard way, but, by contradiction, at some step a leftmost redex $(\lambda x.p)q$ is not reduced. Hence, since the reduction is standard, this redex is frozen (perhaps p is reduced to p' and q is reduced to q', but $(\lambda x.p')q'$ is never reduced). Hence s cannot be in normal form. QED

Definition 21. *Let us write $t \rhd s$ iff $t = t_0 \to_h t_1 \to_h ... \to_h t_n \to_{pi} s$ and $t_i \to_p s$ for all $i \leq n$.*

Lemma 16. *The following hold:*

1. *If $t \rhd s$ then $\lambda x.t \rhd \lambda x.s$*
2. *If $t \rhd s$ and $r \to_p q$, then $tr \rhd sq$.*
3. *If $t \rhd s$ and $r \rhd q$, then $t[r/z] \rhd s[q/z]$.*

Proof. 1. Recall that $t \rhd s$ iff $t = t_0 \to_h t_1 \to_h ... \to_h t_n \to_{pi} s$ and $t_i \to_p s$ for all $i \leq n$. But $t_n \to_{pi} s$ implies $t_n \to_p s$ and by definition this implies $\lambda x.t_n \to_{pi} \lambda x.s$. Moreover, if $t_i \to_h t_{i+1}$, then $\lambda x.t_i \to_h \lambda x.t_{i+1}$.

2. Now let $t = t_0 \to_h t_1 \to_h ... \to_h t_n \to_{pi} s$ and $t_i \to_p s$ for all $i \leq n$. If at least one t_k is an abstraction, take the first with this property. Hence $t_k r \to_{pi} sq$ and $tr = t_0 r \to_h t_1 r \to_h ... \to_h t_k r \to_{pi} sq$, where each $t_i r \to_p sq$. If no t_k is an abstraction, keep $k = n$.

3. Lastly let us first assume that $t \to_{pi} s$ and $r \rhd q$. We want to show that $t[r/x] \rhd s[q/x]$ by considering two subcases:

   3.1 if $t = \lambda z.(\lambda y.P)Qr_0...r_n$ and $s = \lambda z.(\lambda y.P')Q'r'_0...r'_n$, where for each $i \leq n$, $r_i \to_p r'_i$, $P \to_p P'$, $Q \to_p Q'$. But $t[r/x] \to_{pi} s[q/x]$ (recall that substitution holds for $\to_p$ and that if $r \rhd q$, in particular $r \to_{pi} q$).

   3.2 If $t = \lambda z.yr_0...r_n$, then $s = \lambda z.yr'_0...r'_n$.

   (a) If $x \neq y$, then $t[r/x] \to_{pi} s[q/x]$.

   (b) If $x = y$, notice that $r \to_p q$ might be a contraction and we don't want a head reduction. Then argue by applying substitution for parallel reduction, (2) and (1):

$$t[r/x] = \lambda z.rr_0[r/x]...r_n[r/x] \rhd \lambda z.qr'_0[q/x]...r'_n[q/x] = s[q/x]$$

More precisely:

$$\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\vdots}{r_0 \to_{pi} r'_0 \quad r \to_p q}}{r_0[r/x] \to_{pi} r'_0[q/x] \qquad r \rhd q}}{rr_0[r/x] \rhd qr'_0[q/x] \qquad r_1[r/x] \to_{pi} r'_1[q/x]}}{rr_0[r/x]r_1[r/x] \rhd qr'_0[q/x]r'_1[q/x]}}{\dfrac{\dfrac{\vdots}{rr_0[r/x]...r_n[r/x] \rhd qr'_0[q/x]...r'_n[q/x]}}{\lambda z.rr_0[r/x]...r_n[r/x] \rhd \lambda z.qr'_0[q/x]...r'_n[q/x]}}$$

Hence $t[r/x] \rhd s[q/x]$. Now let us consider the general case $t \rhd s$. Notice now that in the general case we have:

$$\underbrace{t[r/x] = t_0[r/x] \to_h t_1[r/x] \to_h ... \to_h}_{substitution\ for\ h-reduction} \overbrace{t_m[r/x] \to_h ... \to_h t_n[r/x] \to_{pi} s[q/x]}^{points\ 3.1\ and\ 3.2}$$

Observe that for each i, still by substitution in parallel reduction, $t_i[r/x] \to_p s[q/x]$. Hence $t[r/x] \rhd s[q/x]$. QED

Lemma 17. (Factorization) *If $t \to_p s$ then $t = t_0 \to_h t_1 \to_h t_2 \to_h ... \to_h r \to_{pi} s$, for some r.*

Proof. Show that $t \to_p s$ implies $t \rhd s$ (notice that this is a *stronger* statement) by induction on the structure of t, using properties 1,2,3 above:

- If $t \to_p t$, nothing to prove.
- If $t = vr \to_p uq$, and by IH $v \rhd u$, then apply 2.
- If $t = \lambda x.v \to_p \lambda x.u = s$, where $v \to_p u$, then by IH $v \rhd u$ and apply 1.
- If $t = (\lambda x.v)r \to_p u[q/x] = s$ and $v \to_p u$ and $r \to_p q$, then by IH obtain $v \rhd u$ and $r \rhd q$. Now apply 3.

QED

Lemma 18. (Inversion) *If $t \to_{pi} r \to_h s$, for some r, then $t \to_h t'_0 \to_h t'_1 \to_h t'_2 \to_h ... \to_h v \to_{pi} s$, for some v.*

Proof. We must have $t = \lambda z.(\lambda x.P)Qr_0...r_n$, since the second step is a head reduction, $r = \lambda z.(\lambda x.P')Q'r'_0...r'_n$ where $P \to_p P'$, $Q \to_p Q'$ and for all $i \leq n$, $r_i \to_p r'_i$. This means that $s = \lambda z.P'[Q'/x]r'_0...r'_n$. Now put $v = \lambda z.P[Q/x]r_0...r_n$. Clearly this is obtained from t by head-reduction, and in turn is reducible to s by parallel reduction (P could be a redex) and therefore, since $v \to_p s$, the previous lemma apply and the parallel reduction can be factorized in two blocks: first a sequence of head reductions, then a parallel internal reduction. QED

Theorem 46. *If t has a normal form s, then there is a leftmost reduction from t to s.*

Proof. By induction on s. Thanks to the previous lemma we can decompose each step $t_i \to_p t_{i+1}$ of the parallel reduction of t to s in this way: $t_i \to_h ...r \to_{pi} t_{i+1}$ for some r. Hence we have a reduction of this form:

$$t = \overbrace{t_0 \to_h ...r_0 \to_{pi}}^{①}\overbrace{t_1 \to_h ...r_1 \to_{pi}}^{②}\overbrace{t_2 \to_h ...r_2}^{③}... \to_{pi} s$$

But by the previous result the head reduction can be brought at the beginning, so that $t = t_0 \to_h ... \to_h r_k \to_{pi} ... \to_{pi} s$. Hence $r_k = \lambda x.yP_0...P_n$ and $s = \lambda x.yP'_0...P'_n$ where P'_i is the normal form of P_i and by induction hypothesis is obtained by leftmost-reduction. But head-reduction too is a kind of leftmost-reduction. Hence t reduces to s by leftmost-reduction. QED

Now, it is important to point out that similar result holds also for *quasi-leftmost* reductions:

$$t_0 \to^{d_0} t_1 \to^{d_1} t_2 \to^{d_2} ...$$

where for all n there exists an $m \geq n$ such that the redex d_m is leftmost in t_m (where d_i is the redex eliminated at step i).

Theorem 47. *If t has an infinite quasi leftmost reduction, then t has no normal form.*

So we are ready to demonstrate the closure under minimalisation (Barendregt (1984) pp. 178-83).

Definition 22. *Let $\Theta = (\lambda xy.y(xxy))(\lambda xy.y(xxy))$ be the Turing fixed point operator; then let $H = \Theta(\lambda uz.If\ Pz,\ then\ z,\ else\ uz^+)$ and $\mu P = H\overline{0}$.*

The above fixed-point operator has the property that $\Theta X \Rightarrow X(\Theta X)$. Recall that "if x then y, else z" is simply $\lambda xyz.xyz$. Note that $H\overline{n}$ means "If $P\overline{n}$ then $\overline{n}$, else $H\overline{n+1}$". Hence μP is "If $P\overline{0}$ then $\overline{0}$ else $H\overline{1}$".

Theorem 48. *Let P be a term such that $P\overline{n} \Rightarrow \bot$, for all n, where $\bot = K^*$. Then:*

1. *μP has no normal form.*
2. *μP is unsolvable.*

Proof. 1. $H\overline{n} \Rightarrow If\ P\overline{n},\ then\ \overline{n},\ else\ H\overline{n+1}$. Now, if $P\overline{n} \Rightarrow \bot$, for all n, we have the infinite sequence of reduction $H\overline{0} \Rightarrow H\overline{1} \Rightarrow H\overline{2} \Rightarrow ....$ In the passage $H\overline{n} \Rightarrow H\overline{n+1}$ at least one head redex is contracted. Hence μP has an infinite quasi-leftmost reduction. Hence it has no normal form.

2. Let $(\mu P)^*$ a closed substitution instance of (μP). This has the form $\mu(P^*)$ and by 1. has no normal form. Hence it is hereditarily without normal form and therefore μP is unsolvable.

QED

Theorem 49. (Closure under minimalization) *Let $\phi(\overline{n}) \simeq \mu m(\chi(n,m) = 0)$ and suppose that G defines a total function χ. Take $P = \lambda y.\mathsf{Zero}(Gxy))$. Hence $F = \lambda x.\mu(\lambda y.\mathsf{Zero}(Gxy))$ defines ϕ.*

Proof. Consider two cases:

1. if $\phi(n)\downarrow$, then this means that for some m (say, the minimum), $\chi(n,m)=0$. Hence $\phi(n)=m$ and therefore $F\overline{n}=\overline{\phi(n)}=\overline{m}$
2. if $\phi(n)\uparrow$, then for all m, $\mathsf{Zero}(G\overline{nm})\Rightarrow\bot$ and therefore $F\overline{n}=\mu(\lambda y.Zero(G\overline{n}y))$ is unsolvable.

To see that all partial recursive functions are representable, recall now this results of characterization: the class of the partial recursive functions is the least class of partial numeric functions containing all primitive recursive functions and closed under composition and this schema of unbounded minimalization:

> Suppose $g(x,y)$ is a *total* function. Then f is defined by minimization from g if and only if
>
> $$f(x)=\mu b(g(x,b)=0)$$
>
> or $f(x)\uparrow$ if there is no such b.

(The total recursive functions are obtained by replacing this schema by the regular minimization, where g is *regular* iff $\mu b(g(x,b)=0)\downarrow$). QED

3.3. The simply typed lambda calculus $\lambda_{\times,\to}$

First developed by Bertrand Russell in the early 1900s to avoid paradoxes, typed calculus was later develped by Church and Curry. In the approach *à la* Church all terms are furnished with type information: if a term is typeable, it has a unique type. In the approach *à la* Curry, types are assigned to existing untyped terms, using typing rules. Thinking terms as algorithms/programs, type as specification, these two approaches reflect two different paradigms in programming languages. We consider here as constructors $\to$ and $\times$. In this case, among the terms, in addition to abstraction and application, we will also have pairs and projections.

Definition 23. *1. (Set of types $\mathbb{T}$)*

(a) A set of base types $\tau_0,\tau_1,\tau_2...$

(b) if σ, τ are types, also $\sigma\to\tau$, $\sigma\times\tau$ are types

(c) these are the ony types.

2. (Typed terms in Church's style) The set of typed terms $\bigcup\{\lambda_\sigma|\sigma\in\mathbb{T}\}$ is defined as follows:

(a) $\sigma\in\mathbb{T},x\in Var\Rightarrow x^\sigma\in\lambda_\sigma$

(b) $t\in\lambda_{\sigma\to\tau},s\in\lambda_\sigma\Rightarrow(ts)\in\lambda_\tau$

(c) $t\in\lambda_\tau\Rightarrow(\lambda x^\sigma.t)\in\lambda_{\sigma\to\tau}$

(d) $t\in\lambda_\sigma,s\in\lambda_\tau\Rightarrow\langle t,s\rangle\in\lambda_{\sigma\times\tau}$

(e) if $t\in\lambda_{\sigma\times\tau}$, then $\pi_0 t$ and $\pi_1 t$ are respectively in λ_σ and in λ_τ.

As already remarked, sometimes the calculus *à la* Church is introduced by means of rules of assignement of types in a similar way to Curry's style. Instead of assuming that the set of variables is partitioned into disjoint sets indexed by types one uses *environments* Γ to declare types of free variables, and says that a judgement $\Gamma\vdash t:\sigma$ holds[1], if it is derivable from a certain set of typing rules.

1. *Environments.* Are set of the form $\Gamma=\{x_0:\tau_0,...,x_n:\tau_n\}$

[1] Rules are often given 'in sequential form'. Below we present them in the form of rules of natural deduction in the Gentzen-Prawitz style, so that the relation to Proof Theory is immediately clear. The reader accustomed to natural deduction in the style of Gentzen-Prawitz may think of this expression as denoting a derivation t of σ from hypotheses not discharged in Γ.

2. *Judgements.* $\Gamma \vdash t : \sigma$ whose meaning is: t is of type σ in the environment Γ.

We say that a term t is *typable*, if there are Γ and σ such that $\Gamma \vdash t : \sigma$. Typical problems of study are the followings:

1. *Type checking.* Given Γ, t and σ, decide whether $\Gamma \vdash t : \sigma$.

2. *Typability.* To decide whether a term is typable.

3. *Type inhabitation.* To decide, for a given type τ, whether there exists a closed term t such that $\vdash t : \tau$.

The first two problems are decidable in polynomial time; the last is PSPACE complete (see Statman (1979))). Type derivations are built up from assumptions of the form $x : \sigma$ by using the following inference rules.

$$\frac{t : \sigma \quad s : \tau}{\langle t, s\rangle : \sigma \times \tau} \qquad \frac{t : \sigma \times \tau}{\pi_0(t) : \sigma} \qquad \frac{t : \sigma \times \tau}{\pi_1(t) : \tau}$$

$$\frac{\begin{array}{c}[x : \sigma] \\ \vdots \\ t : \tau\end{array}}{\lambda x^{\sigma}.t : \sigma \to \tau} \qquad \frac{t : \sigma \to \tau \quad s : \sigma}{(ts) : \tau}$$

Example. Notice that the types assigned to combinators $\mathsf{K} = \lambda xy.x$ and $\mathsf{S} = \lambda xyz.xz(yz)$, as formulas, correspond to well known tautologies, also intuitionistically valid. Write "$t : \alpha$" to mean "$t \in \lambda_\alpha$".

$$\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{x : \alpha \to \beta \to \gamma \quad z : \alpha}{xz : \beta \to \gamma} \quad \dfrac{y : \alpha \to \beta \quad z : \alpha}{yz : \beta}}{xz(yz) : \gamma}}{\lambda z.xz(yz) : (\alpha \to \gamma)}}{\lambda yz.xz(yz) : (\alpha \to \beta) \to (\alpha \to \gamma)}}{\lambda xyz.xz(yz) : (\alpha \to \beta \to \gamma) \to (\alpha \to \beta) \to (\alpha \to \gamma)}$$

Analogously:

$$\dfrac{\dfrac{x : \alpha}{\lambda y.x : (\beta \to \alpha)}}{\lambda xy.x : \alpha \to (\beta \to \alpha)}$$

in analogy with natural deduction, think to an empty application of the introduction rule for implication. Recall that $\mathsf{K} = \lambda xy.x$ is not a term of λ_I-calculus and $\alpha \to (\beta \to \alpha)$ is not accepted in linear as well as relevant logic.

Exercise. Show that Church's numerals have types $(\alpha \to \alpha) \to (\alpha \to \alpha)$.

Proofs as programs. The expert reader will recognize the similarity between these rules and the Gentzen-Prawitz natural deduction propositional rules for the fragment $\wedge, \to$ of the intuitionistic logic. The Curry-Howard isomorphism states that for any derivation in intuitionistic logic there exists a typable λ-term (*à la* Church) and conversely. Actually le language can be exended with an empty type and further constructors in order to reach a full correspondence between typed lambda calculus and intuitionistic logic, where the lambda terms are an appropriate notation for intuitionistic proofs, according to the Brouwer- Heyting-Kolmogorov interpretation of intuitionistic operators. This is the basis of the so-called *proofs-as-programs* paradigm, or *Curry-Howard*

correspondence.

Lambda − calculus	Natural Deduction
$Variables$	$Assumptions$
$Terms$	$Proofs$
$Types$	$Formulas$
$Constructors$	$Connectives$
$Redex$	$Detours$
$Contraction\ \ step$	$Reduction\ \ step\ \ in\ \ normalization$

(1)

The correspondence between $\beta - contraction$ and elimination of *detours* in the normalization algorithm for Natural Deduction is evident:

$$\dfrac{\dfrac{\begin{array}{c}[x:\phi]\\ \vdots\\ t:\psi\end{array}}{\lambda x.t:\phi\to\psi}\,\mathcal{I} \qquad p:\phi}{(\lambda x.t)p:\psi}\,\mathcal{E}$$

Reduces to (where $\mathcal{D}^* = \mathcal{D}[p/x]$):

$$\begin{array}{c} p:\phi\\ \vdots\\ \mathcal{D}^*\\ \vdots\\ t[p/x]:\psi\end{array}$$

Unlike the case of untyped calculus, in the calculus with types *any* term can be reduced to normal form. By the confluence theorem (Church-Rosser) if the normal form exists, this is unique. A term is said to *strongly* normalizzabile when there is no *infinite* sequence of reductions that starts with it. Regarding the fragments of the typed lambda calculus considered above, both of these two results apply: *weak normalization*, i.e. there is at least one strategy reduction leading to the normal form, and *strong normalization*, which means that all reduction sequences are finite (so you get a normal form regardless of the order of application of the rules of contraction). Here we show the *strong* normalization of $\lambda_{\times,\to}$.

This proof is due to Tait and Girard (see Girard, Lafont and Taylor (1989) pp. 41-5 and Troelstra and Schwichtenberg (2000) pp. 157-59) and starts by introducing the notion of reducibility. The set Red_α ("reducible terms of type α") is defined inductively with respect to α as follows:

1. If α is atomic and t is a term of type α, then: $t \in \mathsf{Red}_\alpha$ if there is no infinite reduction sequence starting with t (i.e. t is strongly normalizable).
2. If $\alpha = \sigma \times \tau$ is t a term of type α, then $t \in \mathsf{Red}_\alpha$ if $\pi_0 t \in \mathsf{Red}_\sigma$ and $\pi_1 t \in \mathsf{Red}_\tau$
3. If $\alpha = \sigma \to \tau$ and t a term of type α, then $t \in \mathsf{Red}_\alpha$ if $\forall u \in \mathsf{Red}_\sigma$ we have $tu \in \mathsf{Red}_\tau$.

Now we introduce the notion of neutrality, due to Girard.

Definition 24. *A term t is neutral, if it is neither of the form $\langle u, v\rangle$, nor $\lambda x.t$.*

We show a preliminary result, based on König's lemma ("*An infinite tree finitely ramified, has at least an infinite branch*").

Lemma 19. *If t is strongly normalizable, then exists a number $n_t \in \mathbb{N}$ such that the length of all reduction sequences starting with t is less than n_t (the inverse is clear).*

Proof. Consider the tree $\mathcal{T}$ of all possible reductions of t. Since t has a finite number of subterms the tree is finitely ramified. But if t is strongly normalizable, the tree has no infinite branchs. Hence, by König, $\mathcal{T}$ is finite and therefore there exist an n_t as required. QED

Theorem 50. *The following holds:*

(a) if $t \in \mathsf{Red}_\alpha$, then t is strongly normalizable
(b) if $t \in \mathsf{Red}_\alpha$ and $t \Rightarrow s$, then $s \in \mathsf{Red}_\alpha$.
(c) if t is neutral *and if contracting a redex r of t we always obtain a term $t' \in \mathsf{Red}_\alpha$, then also $t \in \mathsf{Red}_\alpha$.*

Proof. Induction α. In the atomic case (a) follows by definition; as regards instead (b), if $t \in \mathsf{Red}_\alpha$, then t is strongly normalizable; hence, all s such that $t \Rightarrow s$ is strongly normalizable and therefore $s \in \mathsf{Red}_\alpha$; to show (c) we need the previous lemma: each reduction of t goes, by hypothesis, through an $s \in \mathsf{Red}_\alpha$ and such an s is by definition strongly normalizable; So each branch of the tree representing the possible reductions of t is finite. It follows that $n_t = max\{n_s | t \Rightarrow_1 s\} + 1$, and therefore that t is strongly normalizable.

As for the compelx cases:

If $\alpha = \sigma \times \tau$; as regards to (a), suppose that $t \in \mathsf{Red}_\alpha$; then by (IH), since $\pi_0 t \in \mathsf{Red}_\sigma$ and $\pi_1 t \in \mathsf{Red}_\tau$, we will have that $\pi_0 t$ e $\pi_1 t$ are strongly normalizable. Then exists in particular $n_t \leq n_{\pi_0 t}$ and so t is strongly normalizable.

As regards to (b), if $t \Rightarrow s$, then $\pi_0 t \Rightarrow \pi_0 s$ and $\pi_1 t \Rightarrow \pi_1 s$. Suppose that $t \in \mathsf{Red}_{\sigma\times\tau}$. By definition this means that $\pi_0 t \in \mathsf{Red}_\sigma$ and $\pi_1 t \in \mathsf{Red}_\tau$. By (IH) $\pi_0 s \in \mathsf{Red}_\sigma$ and $\pi_1 s \in \mathsf{Red}_\tau$ and therefore by definition $s \in \mathsf{Red}_{\sigma\times\tau}$. As regards to (c), if t is neutral and all s accessible from t in one step are reducible, then we consider $\pi_0 s$ obtained reducing from inside $\pi_0 t$: this is the only possible reduction, not being (for neutrality) $\pi_0 t$ a redex and therefore $t \neq \langle t_0, t_1\rangle$. By hypothesis $s \in \mathsf{Red}_{\sigma\times\tau}$ and therefore $\pi_0 s \in \mathsf{Red}_\sigma$. Hence we have that $\pi_0 t$ is neutral and all $\pi_0 s$ obtained from it in one step is reducible of type σ. Hence we apply (IH) with respect to (c), and conclude that also $\pi_0 t$ is reducible of type σ. The proof for $\pi_1 t$ is symmetric. Hence we have that $\pi_0 t$ is reducible of type σ and $\pi_1 t$ is reducible of type τ and therefore $t \in \mathsf{Red}_{\sigma\times\tau}$.

Let us see the most complex case, namely that in which $\alpha = \sigma \to \tau$. As regards to (a) , if $t \in \mathsf{Red}_{\sigma\to\tau}$ and x is of type σ, we use the inductive hypothesis (on the type) relative to (c): being x neutral and normal, we will have $x \in \mathsf{Red}_\sigma$. But if $t \in \mathsf{Red}_{\sigma\to\tau}$ and $x \in \mathsf{Red}_\sigma$, then by definition $tx \in \mathsf{Red}_\tau$. By applying (IH) on the type relative to (a), we conclude that there exist a n_{tx} finite as in the lemma; henche must exist also $n_t \leq n_{tx}$ and therefore t is strongly normalizable We now quickly see (b); if $t \Rightarrow s$ and t is reducible, take a term $r \in \mathsf{Red}_\sigma$, so that we have $tr \in \mathsf{Red}_\tau$ and $tr \Rightarrow sr$: from (IH) on the type concerning (b), we have that $sr \in \mathsf{Red}_\tau$ and therefore by definition $s \in \mathsf{Red}_{\sigma\to\tau}$.

The proof relative to (c) requires more attention: let t neutral and suppose that all s accessible from t in one step is reducible of type $\sigma \to \tau$; let $u \in \mathsf{Red}_\sigma$ and consider by (IH) on the type relative to (a), that u can be assumed strongly normalizable and therefore there exist n_u as in the lemma. Now we perform a subinduction on n_u: let us see the possible one-step reductions of tu.

1. if $tu \Rightarrow_1 su$, from initial hypothesis $s \in \mathsf{Red}_{\sigma\to\tau}$ and by definition $su \in \mathsf{Red}_\tau$.

2. if $tu \Rightarrow_1 tw$, where $u \Rightarrow_1 w$, then by (IH) on the type relative to (b), we have $w \in \mathsf{Red}_\sigma$; moreover $n_w < n_u$, and therefore by (IH) on n_w we have that $tw \in \mathsf{Red}_\tau$.

We remark that there are no other alternatives, since by neutrality of t, tu cannot be a redex and therefore $t \neq \lambda x.v$. Summarizing, tu is neutral and reduces in one step only to terms in Red_τ. By (IH) on the type relative to (c), we obtain that $tu \in \mathsf{Red}_\tau$ and by definition $t \in \mathsf{Red}_{\sigma\to\tau}$. QED

Now we consider non-neutral terms, and the following holds.

Lemma 20. *If t, s are reducible, the so is also $\langle t, s\rangle$.*

Proof. By (a), t and s are strongly normalizable. Then we use the numbers n_t and n_s and perform an induction on $n_t + n_s$, showing that $\pi_0\langle t, s\rangle$ and $\pi_1\langle t, s\rangle$ are reducible, from which follows that also $\langle t, s\rangle$ is reducible. Indeed, notice that $\pi_0\langle t, s\rangle$ contracts as follows:

1. $\pi_0\langle t, s\rangle \Rightarrow_1 t$ and t is by hypothesis reducible.

2. if $t \Rightarrow_1 r$, then from (b), r is reducible. Moreover $n_r < n_t$, hence we can apply (IH) to $n_r + n_s < n_t + n_s$ and conclude that $\pi_0\langle r, s\rangle$ is reducible.
3. if $s \Rightarrow_1 w$, analogously we obtain that $\pi_0\langle t, w\rangle \Rightarrow_1 t$ is reducible.

With the same criterium we analize $\pi_1\langle t, s\rangle$, concluding that $\pi_0\langle t, s\rangle$ and $\pi_1\langle t, s\rangle$ reduces in one step to a reducible term and therefore by (c) are reducible. Hence $\langle t, s\rangle$ is reducible. QED

Lemma 21. *If for all reducible $s \in \mathsf{Red}_\sigma$ we have that $t[s/x]$ is reducible, then $\lambda x.t$ is reducible.*

Proof. Induction on $n_t + n_s$; we show that $(\lambda x.t)s$ is reducible, if s is reducible. Observe that $(\lambda x.t)s$ is reducible in one step to the following terms:

1. $t[s/x]$, that by hypotesis is reducible.
2. $(\lambda x.r)s$, where $t \Rightarrow_1 r$. Given that t is reducible (from (b)), also r is, and $n_r < n_t$. By (IH) we have that $(\lambda x.r)s$ is reducible.
3. $(\lambda x.t)w$, where $s \Rightarrow_1 w$. Argue as in previous case.

Hence $(\lambda x.t)s$ reduces in one step only to reducible terms, and by (c) is itself reducible. Hence $\lambda x.t$ is reducible, if $t[s/x]$ is, for all s reducible. QED

We obtain the strong normalization result, proving that all the terms are reducible (and therefore, strongly normalizable!). To obtain this additional result, we show that $t[s_0/x_0, ..., s_n/x_n]$ is reducible, for $s_0, ..., s_n$ reducible. In particular, for $s_i = x_i$, we have that $t(x_0, ..., x_n)$ is reducible.

Lemma 22. *$t[s_0/x_0, ..., s_n/x_n]$ is reducible, for $s_0, ..., s_n$ reducible.*

Proof. Induction on the complexity of t:

1. If $t = x_i$, obvious, because $t[s_i/x_i] = s_i$ and s_i is reducible by hypothesis.
2. $t = \pi_0 w$ and by (IH) $w[s_i/x_i]$ is reducible, by definition also $\pi_0 w[s_i/x_i] = t[s_i/x_i]$ is reducible. (the case of π_1 is symmetric).
3. If $t = \langle r, s\rangle$ and by (IH) $r[s_i/x_i]$ $s[s_i/x_i]$ are reducible, then $\langle r[s_i/x_i], s[s_i/x_i]\rangle$ is reducible, from the above results.
4. If $t = rs$ (*Exercise*).
5. If $t = \lambda x.s$ and by (IH) $s[v/x, s_i/x_i]$ is reducible, for all v reducible, the previous emma states that $\lambda x.s[s_i/x_i]$ is reducible.

QED

Corollary 11. *All terms are strongly normalizzabile.*

Proof. Taking $s_i = x_i$ in the previous result, we have that all terms are reducible. QED

Corollary 12. *The relation $t = s$, in the typed calculus, is decidable.*

Proof. The normal forms of t and s are effectively computable.

QED

However the decision method is not elementary, i.e. there is no fixed k such that its complexity is on the order:

$$2^{2^{2^{\cdot^{\cdot^{\cdot^{2}}}}}}$$

iterated $k - times$, as shown in Statman (1979). This is the content of the following theorem (in this version taken from Nerode and Odifreddi (1990) pp. 249-50 and Troelstra and Schwichtenberg (2000) pp. 161-63).

Theorem 51. *For infinitely many terms t any normalizing procedure takes at least a superexponential number of steps.*

Proof. Let us define:

1. $Exp_0(n) = n$
2. $Exp_{k+1}(n) = n^{Exp_k(n)}$
3. $Superexp(n) = Exp_n(n) = n^{n^{\cdot^{\cdot^{\cdot^{n}}}}} \Big\} n+1-times$

Let us consider a Church's numeral $\overline{n}$ and recall that $\overline{nn} = \overline{n^n}$. Recall also that $\overline{n}fx = f^n x$ and therefore $\overbrace{\overline{n}...\overline{n}}^{n+1-times} \; x = f^{Superexp(n)}x$

Notice that at each reduction step a term at most squares its length, e.g. $t = (\lambda x.x^n r)s$ of length about $n + |r| + |s|$ reduces to $s^n r$ on the order $n \cdot |s| + |r| < |t|^2$. How many steps are needed for going from a term of length n^2 to a term of length $Superexp(n)$? Notice that at each step the length increases as follows:

$$n^2, n^{2^2}, n^{2^3} ... n^{2^k} ...$$

Hence to get at step k a value n^{2^k} of the order of $Superexp(n)$ such a k (the number of steps) must be of the order of $Exp_{n-2}(n)$. Notice that:

$$Exp_{n-2}(n) \geq Exp_{n-2}(n-2) = Superexp(n-2)$$

QED

The strong normalization and therefore the decidability, highligt the fact that the typed calculus is weaker than the untyped calculus. Actually, not all computable functions are definable in the typed version. The extended polynomials (generated by projections, constants, sg, $\overline{sg}$, $+$, $\times$, composition) are definable in this calculus. In Schwichtenberg (1976) it is proved that exactly the extended polynomials are definable in it (with numerals expressed *à la Church* and typed as above, where α is a base type). Hence primitive recursion cannot be defined (in fact not even the predecessor function).

A sensitive enrichment in power is represented by the system T of functionals of finite type, dating back to Gödel, that provides a higher type analogue of the notion of primitive recursive computation. Already in Zilsel's lecture of 1938, Gödel had proposed to extend the finitistic mathematics with higher type functionals. This extension comes to light in the context of a proof of consistency of arithmetic and is part of the so-called "Dialectica interpretation", from the name of the journal where appeared the article Gödel (1958) at the 70° of Paul Bernays. The author, however, was not fully satisfied and continued in work there during the next decade (see Gödel (1972) for a later revision). Gödel describes a hierarchy of systems which he calls 'finitary'. These include not only finitary number theory, which he calls *the lowest level of the hierarchy*, but systems which extend beyond this, involving functions of higher type. Following Bernays, he actually distinguished two components in the finitary attitude: the *constructive component*, for which we are allowed to speak of mathematical objects in so far as we can produce them by means of a construction, and the properly *finitistic component*, for which the objects of our statements and our constructions are *intuitive*, spatio-temporal arrangements of elements in a precise sense:

> What Hilbert means by *Anschauung* is substantially Kant's space-time intuition confined, however, to configurations of a finite number of discrete objects (Gödel (1972)).

He argued that the second, too restrictive requirement, must be dropped and certain abstract notions must be admitted in finitistic mathematics, as the primitive recursive functionals of finite types, that generalize the primitive recursive functions. This opens up the possibility of a constructive (although not finitistic) proof of the consistency of arithmetic; what is proposed, is the interpretation of intuitionistic arithmetic HA in a in a quantifier-free theory theory of *functional type* called sytem T (via the so-called Gentzen-Gödel interpretation, the same goes for PA). Already in fragments of arithmetic, it is provable that the consistency of T, implies the consistency of HA and PA. The modern definition of the system T, as a purely typed lambda

calculus, has been given later. This system also provided a point of departure for modern type theory (see e.g. Bishop (1970)). It is confluent and strongly normalizzabile, although, unlike the case of $\lambda_{\rightarrow,\times}$, the strong normalization theorem uses methods that transcend PA and not formalized in it.

The system has a base-type Int. As regards to terms t of type $\mathsf{Int} \rightarrow \mathsf{Int}$, they induce a function $f_t : \mathbb{N} \longrightarrow \mathbb{N}$ such that $t\overline{m} \Rightarrow^* \overline{n}$ if and only if $f_t(m) = n$. A function f for which there exist such an t_f, is called *definable in* T.

Theorem 52. *The class of functions definable in* T *by terms of type* $\mathsf{Int} \rightarrow \mathsf{Int}$ *coincides with those provably total in* PA.

Adding an operation of abstraction on types, as in Girard and Reynold system F, greatly increases the class of representable functions to the class of functions provably total in second order Peano arithmetic (See Girard, Lafont and Taylor (1989) pp. 61-3 and pp. 89-102). This system had a strong influence on the development of type theory.

3.4. The Curry-Howard-Lambek "computational trinitarianism"

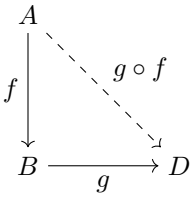
In this chapter, we will discuss the three-way isomorphism between type theory, intuitionistic Proof Theory and *category theory*, widely used in Computer Science, sometimes referred as the *computational trinitarianism*, all three theories being "manifestations of one divine notion of computation" (Harper (2011)). To show this correspondence, it will be necessary to introduce some basic concepts of category theory. This mathematical theory starts with the observation that many properties of mathematical systems can be unified and simplified by a presentation with diagrams of arrows, functors, natural transformations, adjointness etc. Many properties of mathematical constructions may be represented by *universal properties* of diagrams. It is therefore a sort of universal mathematical language, like set theory, that highligts surprising connections among different fields. Category theory started in 1945 with Eilenberg and Mac Lane and has many applications, e.g. in 1950s Grothendieck found applications to algebraic geometry. Eilenberg and Mac Lane actually claimed that category theory should have been considered a generalization of Klein's *Erlangen program*, that placed emphasis on the role of transformation groups and their invariants to classify geometries. According to some authors (see Awodey (1996)) category theory support a form of mathematical structuralism, based on the primitive concept of morphism and distinct from the set-theoretic structuralism of the mathematicians of the Bourbaki group. In 1960s Lawvere applied this theory to logic (see Lawvere (1970) and Lawvere (1969)) and subsequently, Lambek showed in the early 1970s that the proofs of intuitionistic propositional logic and the combinators of typed combinatory logic share a common equational theory which is the one of cartesian closed categories (see Lambek and Scott (1986)). The expression *Curry–Howard–Lambek correspondence* refers to the three way isomorphism between intuitionistic logic, typed lambda calculus and *cartesian closed categories*, with objects being interpreted as types or propositions and morphisms as terms or proofs, so that a morphism $f : \alpha \longrightarrow \beta$ is interpreted as a proof of β from the assumption α. Beyond the below most famous example of correspondence between *Intuitionistic Logic* and *Cartesian Closed Categories*, an interesting development and a suggestion for further study is in the direction of categorical semantic for *Linear Logic* and of the categorical counterpart to linearity. However in that case *Cartesian Closed Categories* are not an adequate a categorical counterpart anymore, but a better choice is that of the so-called symmetric monoidal closed categories (see Abramsky and Tzevelekos (2011)). In fact, this paradigm is the basis for two wide-ranging research programmes: Girard's *Linear Logic* and Martin-Löf's *Intuitionistic Type Theory*. Without attempting to provide an introduction to category theory, we will limit ourselves here to recalling some fundamental concepts that will be useful in the rest of this discussion.

Definition 25. *A category* $\mathcal{C}$ *consists of:*

1. *a collection* $Obj(\mathcal{C})$ *of objects,*
2. *for each objects* A, B *of* $Obj(\mathcal{C})$*, a collection* $\mathcal{C}(A, B)$ *(if this collection is a set, the category is called locally small) of morphisms* $f : A \rightarrow B$*, satisfying the following properties:*
   (a) *there is a distinguished morphism* $Id_A \in \mathcal{C}(A, A)$ *for any object* A *of* $\mathcal{C}$

(b) for each object A, B, D of $\mathcal{C}$ there is an associative mapping $\circ$ such that if $f : A \to B$ and $g : B \to D$, then $g \circ f : A \to D$ (arrows are composable, see figure below),

(c) in particular, if $f : A \to B$, then $Id_B \circ f = f = f \circ Id_A$.



All constructions are given as properties of morphisms between objects. Here some examples of categories:

Category	Objects	arrows
Set	*Sets*	*functions*
Top	*Topological spaces*	*continuous functions*
Vect	*Vectos spaces*	*linear transformations*
Pos	*Posets*	*monotone functions*

(2)

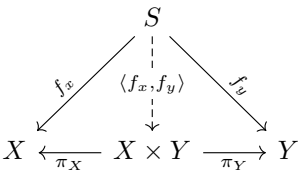
By generalising certain properties of functions (injectivity, surjectivity), we obtain the definition of the following properties of arrows: an arrow (or morphism) $f : A \to B$ is an *isomorphism*, if there is another arrow $g : B \to A$ such that $g \circ f = Id_A$ *and* $f \circ g = Id_B$. Since inverses are unique, we write $g = f^-$. Moreover we have an abstract characterizations of injectivity and surjectivity: an arrow $f : A \to B$ is *monic*, iff given $g, h : C \to A$ we have that $f \circ g = f \circ h$ implies $g = h$. An arrow $f : A \to B$ is instead *epic* , iff given $g, h : B \to D$ we have that $g \circ f = h \circ f$ implies $g = h$. Every isomorphism is both mono and epic. In *Set* the *epic* (right cancellable) arrows are the surjective functions and *monic* (left cancellable) are the injective arrows. Hence arrows both monic and epic are isomorphisms. In other categories this may not be true (il holds in general only that every *iso* is *monic* and *epic*). Rather informally, categories whose *objects* are sets and arrows are functions are called *concrete* and categories whose *class of objects* are (perhaps structured) sets and the class morphisms are (structure-preserving) in turn sets, are called *small*. In the terminology of category theory, *functors* are mapping between categories that map objects to objects in such a way that they preserve domain and codomain, identity arrows and composition, and as categories have morphisms between them –i.e. functors– analogously functors have morphisms between them too, the so-called *natural transformations*. The following are some very basic constructions:

1. *Initial objects*. An object 0 is *initial* in the category $\mathcal{C}$, if for any object A of the category, there is only an arrow $0 \to A$. For example, in the category of sets and functions, since there is only a subset $f \subseteq \emptyset \times A = \emptyset$, for each set A, also there is only one function $\emptyset \to A$, which is $\emptyset$, and therefore the empty set is the only initial object.

2. *Terminal objects*. An object 1 is *terminal* in the category $\mathcal{C}$, if for any object A of the category, there is only an arrow $A \to 1$. For example, in the category of sets and functions the terminal objects are all singletons $\{x\}$: given a set A the function $f(x) = a$, for all $x \in A$, is the only function $f : A \to \{a\}$. In the particular cases of *Mon* and *Grp* (respectively monoids and groups with homomorphisms as arrows), initial objects as well as terminal objects are of the form $\langle \{a\}, *, a \rangle$ with $a * a = a$. In general, in all categories any two initial objects are isomorphic and the same holds for terminal objects.

3. *Elements*. In categories with a terminal object 1, there may be several arrows $1 \to A$, that are called *global elements*, or points.

4. *Products*. A category has *binary products* if for all pairs of objects X, Y, there is an object $X \times Y$ in the category and projections π_X, π_Y such that for all objects S and arrows $f_x : S \to X$ and $f_y : S \to Y$ there is a unique map $\langle f_x, f_y \rangle$ such that the following diagram commutes:

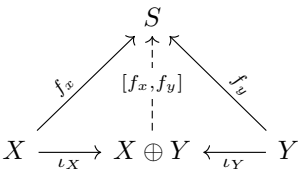


namely $\pi_X \circ \langle f_x, f_y \rangle = f_x$ and $\pi_Y \circ \langle f_x, f_y \rangle = f_y$ and this operation is called the *pairing* of f_x and f_y (dashed arrow means "there is a unique arrow"). A terminal object 1 is a *nullary* product, so, having binary and nullary product we also have unary product, since $A \times 1$ is isomorphic to A.

In the above we say that $\langle \pi_X, \pi_Y \rangle$ is *universal* among pairs arrows from some object, respectively to X and Y, in the sense that any other f_X and f_Y factor uniquely via $\langle f_X, f_Y \rangle$ through $\langle \pi_X, \pi_Y \rangle$. We can therefore say that products are defined by a *universal mapping property*. Other examples of universal mapping properties: terminal and initial objects are defined by how any other object has a unique arrow to or from them.

5. *Coproducts.*

For any category $\mathcal{C}$ the dual (or opposite) category is the category $\mathcal{C}^{op}$ where the objects are the same and the arrows are reversed in direction. The dual notion of *product*, is the *coproduct* with iniections ι_X, ι_Y. If X, Y are objects in a category, a coproduct is an object $X \oplus Y$ with arrows $\iota_X : X \longrightarrow X \oplus Y$ and $\iota_Y : Y \longrightarrow X \oplus Y$ such that for any triplet $X \xrightarrow{f_x} S \xleftarrow{f_y} Y$ we have a unique morphism $[f_x, f_y]$ such that $[f_x, f_y] \circ \iota_X = f_x$ and $[f_x, f_y] \circ \iota_Y = f_y$:

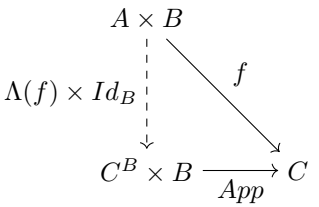


For example, in *Set* the coproduct is:

$$X \oplus Y = \{\langle a, 0\rangle | a \in X\} \cup \{\langle b, 1\rangle | b \in B\}$$

i.e. the disjoint union and the injections $\iota_X(a) = \langle a, 0\rangle$ and $\iota_Y(b) = \langle b, 1\rangle$ and $[f, g](a, 0) = f(a)$, $[f, g](a, 1) = g(a)$. We will not encounter this concept again, but it is important to emphasise one fact. We have decided to deal only with the typed lambda calculus with constructors $\times$ and $\rightarrow$, corresponding to the fragment of intuitionistic propositional calculus with conjunction and implication. We could have extended the Curry-Howard-Lambek isomorphism by adding a type constructor $+$, corresponding to disjunction, i.e. to the coproduct, dual to the product. The correspondence actually extends to coproducts in *bicartesian* closed category, rather than cartesian, namely in cartesian closed categories with the addition of finite coproducts.

6. *Exponentials.* Let us consider for example the category of sets and functions, and the function $f(x, y) : A \times B \rightarrow C$. Now fix $x = a$, so that we have a function $f(a, y) : B \rightarrow C$ i.e. an element of the set C^B of the functions from B to C. We associate to it a map $\Lambda(f) : A \rightarrow C^B$ defined as $a \mapsto f(a, y)$ and uniquely determined by the equation $\Lambda(f)(a)(b) = f(a, b)$. Actually any map $A \rightarrow C^B$ is of this form, coming from some f as above and $\Lambda(f)$ is called *the transpose of f*. Associated to it, there is a map called *evaluation* $App : C^B \times B \rightarrow C$ defined as $App(f, b) = f(b)$. In any category having binary products, an *exponential* for B and C, consists in an object C^B and an arrow App as above, such that the following *universal property* holds: given A and $f : A \times B \longrightarrow C$, there is a unique $\Lambda(f)$ such that $App \circ (\Lambda(f) \times Id_B) = f$.



where the product of morphisms, for $f : A \longrightarrow B$ and $g : C \longrightarrow D$, is the morphism $f \times g : A \times C \longrightarrow B \times D$ defined as $f \times g = \langle f \circ \pi_A, g \circ \pi_C \rangle$, where π_A and π_C are respectively the first and the second projection of $A \times C$. The process of transforming a function of two arguments into a function of one argument is known as *currying*.

The notion of products can be generalised to arbitrary arities. A category is said to have *all finite products* if it has a terminal object and products of any finite cardinality. Actually a terminal object can be seen as a *zero-ary* product: indeed, a zero-ary product is an object t without projections and for any object S there is a unique $S \longrightarrow t$; hence $t = 1$. A *unary* product of A is an object t with a unique projection $t \longrightarrow A$ such that for any $f : S \longrightarrow A$ there is a unique $u : S \longrightarrow t$ such that $\pi \circ u = f$. Take $t = A$ and $\pi = Id_A$. For a formal proof of this, see Mac Lane (1971), p. 73.

Definition 26. *A category is* cartesian closed (CCC) *if it has all finite products and exponentials (i.e. exponent object, evaluation and transposes such that the diagram in the figure above commutes).*

Examples of CCC. Consider a single poset $\langle P, \leq \rangle$ as a category in itself, where $a \to b$ iff $a \leq b$. A terminal object is a largest element 1, product is the infimum $a \wedge b$ and the exponential, usually written $a \Rightarrow b$ satisfies:

$$x \leq (a \Rightarrow b) \; iff \; x \wedge b \leq a$$

that is, $a \Rightarrow b$ is the largest element $x \in P$ such that $x \wedge b \leq a$. If the poset has also an initial object 0 and coproducts $a \vee b$, it is an *Heyting algebra*, i.e. a model of intuitionistic propositional logic, where $p \Rightarrow 0$ is called *pseudo-complement* of p and denoted $\neg p$. If moreover $\neg p \vee p = 1$, then it is a boolean algebra, i.e. a model of classical propositional logic. Another example of CCC is just the category *Pos* of posets and monotone functions. We consider:

1. the poset $P \times Q$ has pairs as its elements, orderd as follows $\langle a, b \rangle \leq \langle c, d \rangle$ iff $a \leq c$ and $b \leq d$. Show by exercise that if $f : X \to P$ and $g : X \to Q$ are monotone, also the projections are monotone.

2. The poset $Q^P = \{f | f : P \to Q, \; f \; monotone\}$ whose elements are ordered pointwise: $f \leq g$ iff for all a, $f(a) \leq g(a)$.

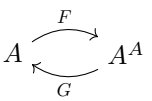
3. *App* and $\Lambda(f)$ defined as above are monotone. Let $p \leq q$ and $\langle f, p \rangle \leq \langle g, q \rangle$, we have:

   $$App(f, p) \leq f(p) \leq f(q) \leq g(q) = App(g, q)$$

   If $x \leq y$, then $\Lambda(f)(x)(p) = f(x, p) \leq f(y, p) = \Lambda(f)(y)(p)$

4. Terminal objects are singleton posets $\langle \{a\}, \{\langle a, a \rangle\} \rangle$. Note that for any poset P there is a unique map $P \to \{a\}$.

Reflexive objects. An object A of a CCC is called *reflexive*, if there are $F : A \to A^A$ and $G : A^A \to A$ such that $F \circ G = Id_{A^A}$



Scott (1980) proved that models of the untyped calculus may be defined as reflexive objects in cartesian closed categories. We address this topic later. Rather, we deal here with calculus *with types*.

Categorical semantic of simply typed lambda calculus. Based on the concepts introduced, we are able to establish a correspondence between Logic, Computability and Categories. We are going to define a *model* of the language of typed lambda-calculus in a cartesian closed category: let therefore $\mathcal{C}$ be a *cartesian closed category*; let us fix an interpretation $[\![\tau]\!]$ of base types τ as objects of the category. For the purposes of an interpretation in cartesian closed categories let us add a *unit type* 1 and a unit value a. The unit value is the only term of type unit (similar to *void* type in certain programming languages). We see that the models of the simply typed lambda calculus with product type and unit are exactly the cartesian closed categories. We say that it is the *internal language* of CCC's. First, we formulate a formal system for the calculus of the $\beta\eta$-conversion as a set of rules. *Judgements* are expressions of the form $\Gamma \vdash t : \alpha$, whose meaning is "t is a term of type α in the context Γ"; this relation is inductively defined by the following rules:

1. Rules that say that the conversion $=$ is an equivalence (i.e. reflexivity, transitivity and simmetry).

2. Unit type

$$\frac{\Gamma \vdash t : 1}{\Gamma \vdash t = a : 1.}$$

3. Product rules:

(a) $$\frac{\Gamma \vdash t = s : \alpha \quad \Gamma \vdash r = h : \beta}{\Gamma \vdash \langle t, r\rangle = \langle s, h\rangle : \alpha \times \beta}$$

(b) $$\frac{\Gamma \vdash s = t : \alpha \times \beta}{\Gamma \vdash \pi_0 s = \pi_0 t : \alpha} \qquad \frac{\Gamma \vdash s = t : \alpha \times \beta}{\Gamma \vdash \pi_1 s = \pi_1 t : \beta}$$

(c) $$\frac{\Gamma \vdash t : \alpha \quad \Gamma \vdash u : \beta}{\Gamma \vdash \pi_0\langle u, t\rangle = u : \alpha} \qquad \frac{\Gamma \vdash t : \alpha \quad \Gamma \vdash u : \beta}{\Gamma \vdash \pi_1\langle u, t\rangle = t : \beta}$$

4. $\eta-$extensional rules:

(a) $$\frac{\Gamma \vdash t : \alpha \to \beta}{\Gamma \vdash \lambda x : \alpha.tx = t : \alpha \to \beta}$$, where x is not among the free variables of t.

(b) $$\frac{\Gamma \vdash t : \alpha \times \beta}{\Gamma \vdash t = \langle \pi_0 t, \pi_1 t\rangle : \alpha \times \beta}$$

5. $\beta-$conversion and application rules:

(a) $$\frac{\Gamma, x : \alpha \vdash t = u : \beta}{\Gamma \vdash (\lambda x : \alpha.t) = (\lambda x : \alpha.u) : \alpha \to \beta}$$

(b) $$\frac{\Gamma \vdash s = t : \alpha \to \beta \quad \Gamma \vdash u = v : \alpha}{\Gamma \vdash su = tr : \beta}$$

(c) $$\frac{\Gamma, x : \alpha \quad \Gamma \vdash s : \alpha}{\Gamma \vdash (\lambda x : \alpha.t)s = t[s/x] : \beta}$$

These rules are valid under the below interpretation in arbitrary cartesian closed category. Fix therefore a CCC and suppose we are given an assignment of an object to each base type. Then we define by induction:

1. $[\![\alpha \to \beta]\!] = [\![\alpha]\!] \to [\![\beta]\!]$ (hence, a morphism between the objects that interpret α and β).

2. $[\![\alpha \times \beta]\!] = [\![\alpha]\!] \times [\![\beta]\!]$.

3. $[\![1]\!] = 1_C$ (where 1_C is a terminal object of the category)

A judgement $\Gamma \vdash t : \alpha$, where Γ has the form $x_0 : \alpha_0, ..., x_n : \alpha_n$ will be interpreted as a morphism:

$$[\![\alpha_0]\!] \times [\![\alpha_1]\!] \times ... \times [\![\alpha_n]\!] \to [\![\alpha]\!]$$

We now list the various terms formation rules (in sequential style) together with their interpretations.

1. *Variables.* $x_0 : \alpha_0, ..., x_n : \alpha_n \vdash x_i : \alpha_i$ will be intepreted in CCC as projections:

$$\pi_i : [\![\alpha_0]\!] \times [\![\alpha_1]\!] \times ... \times [\![\alpha_n]\!] \to [\![\alpha_i]\!]$$

2. *Unit.* $\Gamma \vdash a : 1$ is intepreted as the unique morphism from $[\![\Gamma]\!]$ to 1_C.
3. *Abstraction.*

$$\frac{\Gamma, x : \alpha \vdash t : \beta}{\Gamma \vdash \lambda x.t : \alpha \to \beta}$$

Let f the morphism that intepretes the premiss; then the conclusion is intepreted as the traspose $\Lambda(f)$, so that $App \circ (\Lambda(f) \times Id_{[\![\alpha]\!]}) = f$. Note that, if A, B, C are the objects that intepret respectively α, β, Γ, then:

$$C \times A \xrightarrow{\Lambda(f) \times Id_A} B^A \times A \xrightarrow{App} B$$

and that $f : C \times A \longrightarrow B$ and $\Lambda(f) : C \to B^A$.

4. *Application.*

$$\frac{\Gamma \vdash t : \alpha \to \beta \quad \Gamma \vdash s : \alpha}{\Gamma \vdash ts : \beta}$$

If $f : C \longrightarrow B^A$ and $g : C \longrightarrow A$ are the interpretations of the premisses, then the conclusion is intepreted as $App \circ \langle f, g \rangle$:

$$C \xrightarrow{\langle f,g \rangle} B^A \times A \xrightarrow{App} B$$

Recall that we are in CCC, and therefore if $h : C \longrightarrow P$ and $k : C \longrightarrow Q$, there exists a unique map $\langle h, k \rangle : C \longrightarrow P \times Q$, such that $\pi_0 \circ \langle h, k \rangle = h$ and $\pi_1 \circ \langle h, k \rangle = k$.

5. *Product.*

$$\frac{\Gamma \vdash t : \alpha \quad \Gamma \vdash s : \beta}{\Gamma \vdash \langle t, s \rangle : \alpha \times \beta}$$

If $f : C \longrightarrow A$ and $g : C \longrightarrow B$ are the interpretations of the premisses, take $\langle f, g \rangle : C \longrightarrow A \times B$ as the intepretation of conclusion.

6. *Projections.*

$$\frac{\Gamma \vdash t : \alpha \times \beta}{\Gamma \vdash \pi_0 t : \alpha}$$

If $f : C \longrightarrow A \times B$ intepretes the premiss, then $\pi_0 \circ f : C \longrightarrow A$ intepretes the conclusion:

$$C \xrightarrow{f} A \times B \xrightarrow{\pi_0} A$$

(Analogously for the second projection)

Substitution as composition of morphisms. Let us consider a context $x_0 : \alpha_0, ..., x_n : \alpha_n$, interpreted as the product $A_0 \times ... \times A_n$. A substitution of terms $t_0, ..., t_n$ for the variables $x_0, ..., x_n$, where for all $i \leq n$, $\Delta \vdash t_i : \alpha_i$ and these are interpreted as morphisms $g_i : C \longrightarrow A_i$, is a morphism:

$$C \xrightarrow{\langle g_0, ..., g_n \rangle} A_0 \times ... \times A_n$$

Lemma 23. *If $f : A_0 \times ... \times A_n \times A \longrightarrow B$ interprets $\Gamma, x : \alpha \vdash t : \beta$, and $g : A_0 \times ... \times A_n \longrightarrow A$ the interpretation of $\Gamma \vdash s : \alpha$; then the interpretation of $\Gamma \vdash t[s/x] : \beta$ is $f \circ \langle Id_{A_0 \times ... \times A_n}, g \rangle$:*

$$A_0 \times ... \times A_n \xrightarrow{\langle Id_{A_0 \times ... \times A_n}, g \rangle} A_0 \times ... \times A_n \times A \xrightarrow{f} B$$

Proof. By structural induction on t. QED

Theorem 53. (Soundness of the categorical semantic) *The interpretation is sound, namely, if $\Gamma \vdash t = s : \alpha$ is provable, then $[\![\Gamma \vdash t : \alpha]\!] = [\![\Gamma \vdash s : \alpha]\!]$.*

Proof. We will actually show the most significant cases, leaving the others as exercises.

(i) As for the unit case, the result follows from the fact that there is a unique morphism $[\![\Gamma]\!] \longrightarrow 1$.

(ii) The case 5.a is proved arguing by induction on the hight of derivation: suppose by inductive hypothesis the theorem holds for the premiss, i.e. that $[\![\Gamma, x : \alpha \vdash t : \beta]\!] = [\![\Gamma, x : \alpha \vdash u : \beta]\!]$; hence:

$$[\![\Gamma \vdash \lambda x : \alpha.u : \alpha \to \beta]\!] = \Lambda([\![\Gamma, x : \alpha \vdash u : \beta]\!]) =$$

$$= \Lambda([\![\Gamma, x : \alpha \vdash t : \beta]\!]) = [\![\Gamma \vdash \lambda x : \alpha.t : \alpha \to \beta]\!]$$

The case 5.b is proved in a similar way, reasoning again by induction. As for $\beta-$conversion 5.c, let $\Gamma = x_0 : \alpha_0, ..., x_n : \alpha_n$ and suppose that $\Gamma, x : \alpha \vdash t : \beta$ and $\Gamma \vdash s : \alpha$ are intepreted resp. by $f : A_0 \times ... \times A_n \times A \longrightarrow B$ and $g : A_0 \times ... \times A_n \longrightarrow A$. By definition:

$$[\![\Gamma \vdash (\lambda x : \alpha.t)s : \beta]\!] = App \circ \langle \Lambda(f), g \rangle$$

But this is equal to $App \circ (\Lambda(f) \times Id_A) \circ \langle Id_{A_0 \times ... \times A_n}, g \rangle$:

$$A_0 \times ... \times A_n \xrightarrow{\langle Id_{A_0 \times ... \times A_n}, g \rangle} A_0 \times ... \times A_n \times A \xrightarrow{\Lambda(f) \times Id_A} B^A \times A \xrightarrow{App} B$$

Since in CCC $App \circ (\Lambda(f) \times Id_A) = f$, lastly we get $f \circ \langle Id_{A_0 \times ... \times A_n}, g \rangle$, namely:

$$[\![\Gamma, x : \alpha \vdash t : \beta]\!] \circ \langle Id_{A_0 \times ... \times A_n}, g \rangle$$

But from the substitution lemma it follows that this is equal to $[\![\Gamma \vdash t[s/x] : \beta]\!]$. Hence $[\![\Gamma \vdash (\lambda x : \alpha.t)s : \beta]\!] = [\![\Gamma \vdash t[s/x] : \beta]\!]$.

(iii) As for $\eta - reduction$ 4.a, observe that:

$$[\![\Gamma \vdash \lambda x : \alpha.tx]\!] = \Lambda([\![\Gamma, x : \alpha \vdash tx]\!]) = \Lambda(App([\![\Gamma \vdash t]\!] \times Id_A) = [\![\Gamma \vdash t]\!]$$

As regards 4.b, just observe that:

$$[\![\Gamma \vdash \langle \pi_0 t, \pi_1 t \rangle : \alpha \times \beta]\!] = \langle \pi_0 \circ [\![\Gamma \vdash t : \alpha \times \beta]\!], \pi_1 \circ [\![\Gamma \vdash t : \alpha \times \beta]\!] \rangle =$$

$$= [\![\Gamma \vdash t : \alpha \times \beta]\!]$$

(iv) As regards to product case 3.c. note that:

$$[\![\Gamma \vdash \pi_0(\langle u, v \rangle) : \alpha]\!] = \pi_0 \circ [\![\Gamma \vdash \langle u, v \rangle : \alpha \times \beta]\!] =$$

$$= \pi_0 \circ \langle [\![\Gamma \vdash u : \alpha]\!], [\![\Gamma \vdash v : \beta]\!] \rangle = [\![\Gamma \vdash u : \alpha]\!]$$

(v) Analyse the other cases by exercise.

QED

With regard to *completeness*, we provide a sketch of the proof, which the reader can find in more detail in Abramsky and Tzevelekos (2011), 62-64, which reproduces and simplifies the one in ch.4.8 of Crole (1994). We just sketch the basic intuition. Our aim is to find a category in which true equalities correspond precisely to provable conversions between terms in the calculus. Such a categorical model is the *term model*, or *classifying category*, the "free category" generated by the syntax of the calculus:

1. the *objects* of this category $\mathbb{C}$ are types, plus a terminal object 1;

2. the *morphisms* $\alpha \longrightarrow \beta$ are *equivalence classes* of pairs variable-term x, t such that x, t and y, s are considered equivalent if both $x : \alpha \vdash t : \beta$ and $y : \alpha \vdash s : \beta$ are provable and t can be converted in $s[x/y]$. It is not hard to show that this actually is an equivalence. Special cases are considered, when the first element of the pair is empty: hence we say that $-, t$ and $-, s$ are equivalent if $\vdash t : \beta$ and $\vdash s : \beta$ are provable and t can be converted in s. We denote with $(x|t)$ the equivalence class of the couple x, t. Hence $\mathbb{C}(\alpha, \beta)$ will be the set of equivalence classes $(x|t)$ such that $x : \alpha \vdash t : \beta$ is provable. In particular $\mathbb{C}(1, \beta)$ is the set of equivalence classes $(-|t)$ such that $\vdash t : \beta$ is provable and $\mathbb{C}(\beta, 1)$ is the singleton of the terminal arrow $1_\alpha : \alpha \longrightarrow 1$.

3. Identities are $Id_\alpha = (x|x)$, if $\alpha \neq 1$, and $Id_\alpha = 1_\alpha$, if $\alpha = 1$.

4. The composition of arrows is defined as $(x|t) \circ (y|u) = (y|t[u/x])$, with the special cases of $(-|t) \circ 1_\alpha = (y|t)$, if $\alpha \neq 1$, and $(-|t) \circ 1_\alpha = (-|t)$, if $\alpha = 1$, and of $1_\beta \circ f = 1_\alpha$, where $f \in \mathbb{C}(\alpha, \beta)$.

Some facts:

1. Composition is associative. Indeed, for the substitution lemma at p. 61, the the following terms (a) and (b) are equal:

   (a) $(x|t) \circ ((y|u) \circ (z|y)) = (x|t) \circ (z|u[v/y]) = (z|t[u[v/y]/x])$

   (b) $((x|t) \circ (y|u)) \circ (z|v) = (y|t[u/x]) \circ (z|v) = (z|t[u/x][v/y])$

   So what we have defined is a *category*. It is actually a *cartesian closed* category:

2. The category $\mathbb{C}$ has binary products. We define projections π_i in $\alpha \xleftarrow{\pi_0} \alpha \times \beta \xrightarrow{\pi_1} \beta$ as $\pi_i = (x|\pi_i x)$ and, given a triple $\alpha \xleftarrow{(x|t)} \mu \xrightarrow{(x|u)} \beta$, we define the morphism $\langle (x|t), (x|u) \rangle : \mu \longrightarrow \alpha \times \beta$ as $(x|\langle t, u \rangle)$. Hence, for instance we have $\pi_1 \circ \langle (x|t), (x|u) \rangle = (u|\pi_1 y) \circ (x|\langle t, u \rangle) = (x|\pi_1, u\rangle) = (x|t)$. Readers can practise demonstrating uniqueness.

3. The category $\mathbb{C}$ has exponentials too. Let us define App as $(x|(\pi_0 x)(\pi_1 x))$ and, given $(x|t) : \sigma \times \alpha \longrightarrow \beta$, let us define $\Lambda((x|t)) = (x_0|\lambda x_1.t[\langle x_0, x_1 \rangle / x])$, so that:

$$
\begin{aligned}
App \circ \Lambda((x|t)) \times Id &= App \circ \langle \Lambda((x|t) \circ \pi_0, Id \circ \pi_1 \rangle \\
&= App \circ \langle (x_0|\lambda x_1.t[\langle x_0, x_1 \rangle / x]) \circ (y|\pi_0 y), (y|\pi_1 y) \rangle \\
&= App \circ \langle (y|\lambda x_1.t[\langle \pi_1 y, x_1 \rangle / x]), (y|\pi_1 y) \rangle \\
&= (z|(\pi_0 z)(\pi_1 z)) \circ (y|\langle \lambda x_1.t[\langle \pi_1 y, x_1 \rangle / x], \pi_1 y \rangle) \\
&= \langle y, (\pi_0 X)(\pi_1 X) \rangle = (y|(\lambda x_1.t[\langle \pi_0 y, x_1 \rangle / x](\pi_1 y) \\
&= (u|t[\langle \pi_0 y, \pi_1 y \rangle / x]) = (y|t[y/x]) = (x|t)
\end{aligned}
\tag{3}
$$

   where $X = \langle \lambda x_1.t[\langle \pi_0 y, x_1 \rangle / x], \pi_1 u \rangle$, and $\Lambda((x|t)) : \sigma \longrightarrow \beta^\alpha$; moreover $App : \beta^\alpha \times \alpha \longrightarrow \beta$ and $Id = Id_\alpha$. Note that we have used $\eta-$reduction at the penultimate step. Uniqueness and other cases left to the reader.

Now, using the interpretation of lambda terms in categories that we used in the soundness theorem, we can easily verify that:

$$[\![\Gamma \vdash t : \tau]\!] = (x|t[\pi_0/x_0, ..., \pi_n/x_n])$$

where x does not occur in $\Gamma = \{x_0 : \tau_0, ..., x_n : \tau_n\}$ and $x : \tau_0 \times ... \times \tau_n$. From this follows that if $[\![\Gamma \vdash t : \alpha]\!] = [\![\Gamma \vdash s : \alpha]\!]$ holds in it, then $\Gamma \vdash t = s : \alpha$ is provable and together with soundness, this implies that in this model true equalities coincide with provable conversions.

3.5. Categorical models for untyped lambda calculus

Various semantics for the untyped lambda calculus have been studied. Of these, in particular, the *Scott continuous semantics* is based on the class of reflexive objects in the cartesian closed category CPO whose objects are complete partial orders and morphisms are Scott continuous functions. This category actually has reflexive objects. More in general, a categorical model of the untyped λ-calculus is a reflexive object of a Cartesian closed category. Since in untyped lambda calculus self-application xx is admitted, terms can be seen both as objects as well as functions, so in the study of models we will address the so-called domain equation $D \cong (D \to D)$. Clearly, by Cantor's theorem $(D \to D)$ cannot be the whole space of functions from D to D (the only case in which holds is when D is a singleton). We will illustrate now Scott's method for the concrete cartesian closed category CPO of complete partial orders. However the method works for all concrete cartesian closed categories.

Definition 27. *Let $\langle P, \leq \rangle$ be a poset. A subset $X \subseteq P$ is directed iff for all $x, y \in X$ there exists $z \in X$ such that $x \leq z$ and $y \leq z$. The poset is a cpo iff:*

1. *there is a bottom element $\bot$, i.e. for all $x \in P$, $\bot \leq x$.*

2. *If $X \subseteq P$ is directed, then the supremum $\sqcup X \in P$ exists.*

Definition 28. (Scott topology) *The Scott topology on a cpo $\langle P, \leq \rangle$ is defined as follows. A subset $\mathcal{O} \subseteq P$ is open, iff:*

1. *$x \in \mathcal{O}$ and $x \leq y$, then $y \in \mathcal{O}$ (closure upwards).*

2. *If X is directed and $\sqcup X \in \mathcal{O}$, then $X \cap \mathcal{O} \neq \emptyset$ (inaccessibility by directed joins).*

Since according to this definition the empty set and P are open, and opens are closed under arbitrary unions and if $\mathcal{O}, \mathcal{O}'$ are opens, then $\mathcal{O} \cap \mathcal{O}'$ is open, we have that this definition actually determines a topology in the usual sense. Moreover the sets $U_x = \{z | z \not\leq x\}$ are clearly opens.

Continuous maps.

Theorem 54. . *Let D, D' be two cpo's and $f : D \to D'$. Hence f is continuous in the usual sense (i.e. $f^{-1}(Y) = \{x \in D | f(x) \in Y\}$ is open, when Y is open), iff for all directed $X \subseteq D$, $f(\sqcup X) = \sqcup f(X)$, where $f(X) = \{f(x) | x \in X\}$.*

Proof. . $\Rightarrow$ Let f be continuous. We see that it is also monotonic: actually, if $x \leq y$, but $f(x) \not\leq f(y)$, then $f(x) \in U_{f(y)} = \{z | z \not\leq f(y)\}$. Since $U_{f(y)}$ is open and f is continuous, by definition $f^{-1}(U_{f(y)})$ is open too, and x belong to it. Hence by closure upwards, also y belong to it and $f(y) \in U_{f(y)}$ (contradiction). So f is monotonic. It follows that $x \in X$, $f(x) \leq f(\sqcup X)$ and therefore $\sqcup_{x \in X} f(x) \leq f(\sqcup X)$. Now suppose that the other way round does not hold. Hence $f(\sqcup X) \in U_{\sqcup f(X)}$ by definition, and therefore $\sqcup X \in f^{-1}(U_{\sqcup f(X)})$. Now, since X is directed and $f^{-1}(U_{\sqcup f(X)})$ is open, then $X \cap f^{-1}(U_{\sqcup f(X)})$ is not empty. Take x in this intersection. It follows that $f(x) \in U_{\sqcup f(X)}$ (contradiction). $\Leftarrow$ Let $f(\sqcup X) = \sqcup f(X)$. Since $x \leq y$ implies $y = x \sqcup y$, we have $f(y) = f(x \sqcup y)$ and by hypothesis $f(x \sqcup y) = f(x) \sqcup f(y)$, from which monotonicity, $f(x) \leq f(y)$. Observe that if $X \subseteq D$ is directed and $\sqcup X \in f^{-1}(\mathcal{O})$, then (since by hypothesis $f(\sqcup X) = \sqcup f(X)$) we have $\sqcup f(X) \in \mathcal{O}$. But $f(X)$ is directed and therefore by definition $f(X) \cap \mathcal{O} \neq \emptyset$, from which

follows $X \cap f^{-1}(\mathcal{O}) \neq \emptyset$. But $f^{-1}(\mathcal{O})$ is also upwards closed. It follows that $f^{-1}(\mathcal{O})$ is open and therefore f is continuous.

Remarks. Continuous maps in cpo's are always monotone.

We now show that the category CPO of cpo's and continuous maps is cartesian closed. To this purpose, it is necessary to show some facts, but first we make two further remarks. QED

Now we show that CPO, the category of cpo's and continuous maps as arrows, is cartesian closed. We need these results:

Theorem 55. *The set $[D \to D']$ of continuous maps from D to D' is a cpo, where:*

1. *it is ordered pointwise:*
$$f \leq g \; iff \; \forall d \in D(f(d) \leq g(d))$$
2. *$\bot = D \times \{\bot_{D'}\}$ is the function $x \mapsto \bot_{D'}$,*
3. *If $X \subseteq [D \to D']$ is directed, then $\sqcup X$ is the map $x \mapsto \sqcup\{f(x) | f \in X\}$.*

Proof. We just show that in 3. if $X \subseteq [D \to D']$ is directed, then $\sqcup X$ is continuous. Note that if X is directed, then for all $f, g \in X$ there is an h such that for all $a \in D$, $f(a) \leq h(a)$ and $g(a) \leq h(a)$. Let $f(p) = \sqcup\{f_i(p) | f_i \in X\} = \sqcup X(p)$. Hence:

$$f(\sqcup_j p_j) = \sqcup_i f_i(\sqcup_j p_j) = \sqcup_i \sqcup_j f_i(p_j) = \sqcup_j \sqcup_i f_i(p_j) = \sqcup_j f(p_j)$$

Theorem 56. *If D, D' are cpo's, then $D \times D'$, is a cpo, where:*

1. *the bottom is $\langle \bot, \bot \rangle$*
2. *$\langle a, b \rangle \leq \langle c, d \rangle$ iff $a \leq c$ and $b \leq d$,*
3. *if $X \subseteq D \times D'$ is directed, then $\sqcup X = \langle \sqcup X_0, \sqcup X_1 \rangle$, where:*
   (a) $X_0 = \{x \in D | \langle x, a \rangle \in X, \; for \; some \; a \in D'\}$
   (b) $X_1 = \{x \in D' | \langle a, x \rangle \in X, \; for \; some \; a \in D\}$

Theorem 57. *Let D, D', B be cpo's. Hence $f : (D \times D') \to B$ is continuous iff it is continuous in both arguments separately, i.e. iff both $d \mapsto f(d, d')$ and $d' \mapsto f(d, d')$ are continuous.*

We omit the proofs of these results, rather standard (see Barendregt's handbook pp. 10-3).

Theorem 58. (Continuity of application) *Let $B^A = [A \to B]$ and A, B, C cpo's. Hence the map $App : B^A \times A \to C$ defined as $App(f, a) = f(a)$ is continuous.*

Proof. Let us consider the components $\lambda f.f(x)$ (namely the function $f \mapsto App(f, x)$) and $\lambda x.f(x)$ (namely the function $x \mapsto App(f, x)$). The latter is clearly continuous, as for the former, we argue as follows. Let $X \subseteq B^A$ be directed and let $h = \lambda f.f(x)$; hence:

$$h(\sqcup X) = \lambda f.f(x)(\sqcup X) = (\sqcup X)(x) =$$

$$= \sqcup\{f(x) | f \in X\} = \sqcup\{h(f) | f \in X\} = \sqcup h(X)$$

Hence *App* is continuous in both components and by the previous result, it is continuous itself. QED

Theorem 59. (Continuity of abstraction) *Let $f \in [B \times A \to C]$ and $\Lambda(f)(x) = \lambda y.f(x, y)$ (i.e. the map that send x in $y \mapsto f(x, y)$). It follows that $\Lambda(f)$ is continuous.*

Proof. Observe that $\Lambda(f) : A \to [B \to C]$. Let $X \subseteq A$ be directed; hence:

$$\Lambda(f)(\sqcup X) = \lambda y.f(\sqcup X, y) = \lambda y.\, \sqcup_{x \in X}\, f(x,y) =$$

$$= \sqcup_{x \in X}(\lambda y.f(x,y)) = \sqcup_{x \in X}\Lambda(f)(X)$$

QED

Theorem 60. *Also the mapping $f \mapsto \Lambda(f)$ is continuous.*

Proof. Let $H = \lambda f.\Lambda(f)$. Then for $X \subseteq [B \times A \to C]$ directed. Hence:

$$H(\sqcup X) = \lambda x \lambda y.(\sqcup X)(x,y) = \lambda x \lambda y.(\sqcup_{f \in X} f(x,y)) =$$

$$= \sqcup_{f \in X} \lambda x \lambda y.f(x,y) = \sqcup H(X)$$

QED

So we have that if A, B are cpo's, also $A \times B$ and B^A are cpo's. The singleton cpo $\langle \{a\}, \{\langle a, a\rangle\}\rangle$ works as terminal object. Moreover *App* is continuous and for every continuous $f : A \times B \to C$, there is a unique continuous $\Lambda(f) : A \to [B \to C]$ such that the diagram of CCC. commutes. Hence we have the followiung result:

Corollary 13. CPO *is cartesian closed.*

We now consider reflexive cpo's A, namely equipped with continuous maps:

$$A \underset{G}{\overset{F}{\rightleftarrows}} A^A$$

Such that $F \circ G = Id_{A^A}$, where $A^A = [A \to A]$. We say that $[A \to A]$ is a *retract* of A.

We define an interpretation of terms $[\![t]\!]^\sigma$ on the basis of an evaluation $\sigma : Var \to A$ of variables:

1. $[\![x]\!]^\sigma = \sigma(x)$
2. $[\![ts]\!]^\sigma = F([\![t]\!]^\sigma)([\![s]\!]^\sigma)$
3. $[\![\lambda x.t]\!]^\sigma = G(d \mapsto [\![t]\!]^{\sigma[d/x]})$

This definition is correct, only if we prove the following:

Lemma 24. *The function $\lambda d.[\![t]\!]^{\sigma[d/x]}$, namely $d \mapsto [\![t]\!]^{\sigma[d/x]}$ is continuous.*

Proof. Induction on t. The only interesting case is when $t = \lambda y.s$. By (IH) we have that $h(d,c) = [\![s]\!]^{\sigma[d/x, c/y]}$ is continuous separately in d and c; hence by previous a result also $h(d,c)$ is continuous. Hence:

$$[\![\lambda y.s]\!]^{\sigma[d/x]} = G(c \mapsto h(d,c)) = G(\Lambda(h)(d))$$

namely the composition of continuous maps, itself continuous. QED

Also by induction on the complexity of terms it is provable the following substitution result:

$$[\![t[s/x]]\!]^\sigma = [\![t]\!]^{\sigma[[\![s]\!]^\sigma/x]}$$

From which follows the main result:

Theorem 61. *If $t \Rightarrow s$, then $[\![t]\!]^\sigma = [\![s]\!]^\sigma$.*

Proof. . We see just the case of contraction:

$$[\![(\lambda x.t)s]\!]^\sigma = F([\![(\lambda x.t)]\!]^\sigma)([\![s]\!]^\sigma) =$$

$$= F(G(d \mapsto ([\![t]\!]^{\sigma[d/x]}))([\![s]\!]^\sigma) = (d \mapsto ([\![t]\!]^{\sigma[d/x]})([\![s]\!]^\sigma) =$$

$$= [\![t]\!]^{\sigma[[\![s]\!]^\sigma/x]} = [\![t[s/x]]\!]^\sigma$$

QED

The extensionality axiom $\lambda x.tx$, where $x \notin FVar(t)$, holds only if $G \circ F = Id_A$ also holds, namely if we have an isomorphism $A \cong [A \to A]$. In 1969 Dana Scott introduced the model D_∞ satisying this isomorphism. In the 70s Scott and Plotkin introduced the easier, but non extensional *graph model* $\langle \mathcal{P}(\omega), \subseteq \rangle$ of subsets of natural numbers ordered by set-theoretical inclusion. Both models can be constructed in the category **CPO**. Here we illustrate some property of the graph-model. Actually, this is a cpo, with $\cup$ as supremum. Being a complete lattice, supremum and infimum exist for any set. We start showing some basic properties.

Theorem 62. *A function $f : \mathcal{P}(\omega) \to \mathcal{P}(\omega)$ is continuous iff for every $A \subseteq \omega$:*

$$f(A) = \bigcup_{D \subseteq_{finite} A} f(D)$$

Proof. $\Rightarrow$ by monotonicy if $D \subseteq A$, then $f(D) \subseteq f(A)$, hence $f(A) \supseteq \bigcup_{D \subseteq_{finite} A} f(D)$. For the other way round, consider $A_n = A \cap \{0, ..., n\}$, so that $A = \bigcup_n A_n$. Hence by continuity:

$$f(A) = f(\bigcup_n A_n) = \bigcup_n f(A_n)$$

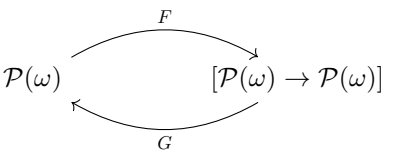
Since the A_n 's are finite, it follows $f(A) \subseteq \bigcup_{D \subseteq_{finite} A} f(D)$

$\Leftarrow$ Suppose that $f(A) = \bigcup_{D \subseteq_{finite} A} f(D)$. Note that if $A \subseteq B$ and $D \subseteq_{finite} A$, then $D \subseteq_{finite} B$, from which $f(A) \subseteq f(B)$. Take a directed set of sets A_0, A_1, A_2.... Hence:

$$f(\bigcup_n A_n) = \bigcup \{f(D) | D \subseteq_{finite} \bigcup_n A_n\} =$$

$$= \bigcup_n \bigcup \{f(D) | D \subseteq_{finite} A_n\} = \bigcup_n f(A_n)$$

Retraction pair. Now we define the retraction pair F, G as follows:



Let $D_x = \{x_0, ..., x_n\}$, where $x = 2^{x_0} + ... + 2^{x_n}$ and $x_0 < ... < x_n$. Let us define:

1. $F(S)(X) = \{m | \exists n(D_m \subseteq X \wedge \langle n, m \rangle \in S)\}$
2. $G(f) = \{\langle n, x \rangle | x \in f(D_n)\}$

Lemma 25. $F \circ G = Id_{[\mathcal{P}(\omega)\to\mathcal{P}(\omega)]}$
Let $f \in [\mathcal{P}(\omega) \to \mathcal{P}(\omega)]$. Hence:

$$F(G(f))(X) = \{x|\exists n(\langle x,n\rangle \in G(f) \wedge D_n \subseteq X)\} =$$

$$= \{x|\exists n(x \in f(D_n) \wedge D_n \subseteq X)\} =$$

$$= \bigcup\{f(D_n)|D_n \subseteq X\} = f(X)$$

Proof. By the previous result. QED

Now we show that F and G are continuous.

Theorem 63. *The maps F and G, defined as above, are both continuous.*

Proof. 1. Let $X \subseteq [\mathcal{P}(\omega) \to \mathcal{P}(\omega)]$. Then:

$$G(\bigsqcup X) = \{\langle n,m\rangle|m \in (\bigsqcup X)(D_n)\} = (by\ definition\ of\ sup.)$$

$$= \{\langle n,m\rangle|m \in \bigcup_{f\in X} f(D_n)\} =$$

$$= \bigcup_{f\in X}\{\langle n,m\rangle|m \in f(D_n)\} =$$

$$= \bigcup_{f\in X} G(f) = \bigsqcup G(X)$$

2. Let $X \subseteq \mathcal{P}(\omega)$. Then:

$$F(Z)(\bigcup_i X_i) = \{a|\exists n(\langle n,a\rangle \in Z \wedge D_n \subseteq \bigcup_i X_i)\} =$$

$$= \{a|\exists_i \exists n(\langle n,a\rangle \in Z \wedge D_n \subseteq X_i)\} =$$

$$= \bigcup_i \{a|\exists n(\langle n,a\rangle \in Z \wedge D_n \subseteq X_i)\} =$$

$$= \bigcup_i F(Z)(X_i)$$

QED

Corollary 14. *$\langle \mathcal{P}(\omega), \subseteq\rangle$ with the retraction pair F, G and the interpretation $[\![x]\!]$ is a model of the type free lambda calculus.*

Note the disparity between the countable amount of terms of the language, and the uncountable amount of object of the model. An interesting development is to restrict the model to the set $\mathcal{E}$ of recursively enumerable sets.

Definition 29. *A continuous $f : \mathcal{P}(\omega) \to \mathcal{P}(\omega)$ is effective iff $G(f)$ is recursively enumerable.*

Theorem 64. *Continuous effective f's, maps $\mathcal{E}$ in $\mathcal{E}$.*

Proof. $x \in f(A)$ iff $\exists n(\langle x,n\rangle \in G(f) \wedge D_n \subseteq A)$ (where D_n is the already mentioned coding of finite sets). Note that if $G(f)$ and A are r.e. the whole formula, being Σ_1 is r.e. Continuous effective f as above are called *enumeration operators.* QED

Theorem 65. *For all interpretations of variables in $\mathcal{E}$, $[\![t]\!]$ is an r.e. set, for all t.*

We remark that the graph model is not extensional (i.e. does not satisfies the $\eta - rule$). Scott introduced another *extensional* model in 1969. It was the first non-trivial model of the untyped Lambda-calculus. We conclude just sketching the basic steps of the construction of Scott's model named D_∞. The equation $D = D \longrightarrow D$ is solved not as an equality, but as an isomorphism. Let $D_0, D_1, D_2...$ be a countable sequence of cpo's with Scott's topology, defined as $D_0 = D$ and $D_{n+1} = [D_n \longrightarrow D_n]$ and let $j_n : D_{n+1} \longrightarrow D_n$ be continuous maps. Each D_n is a finer approximation to a limit than each of the preceding and each D_n will be embedded in D_{n+1} preserving its structure. The elements of this limit set are infinite sequences of functions. The application of an element $b = \langle b_0, b_1, b_2...\rangle$ to an element $a = \langle a_0, a_1, a_2...\rangle$ is defined as:

$$b \cdot a = \langle b_1(a_0), b_2(a_1), b_3(a_2)...\rangle$$

so that self-application is made possible $a \cdot a = \langle a_1(a_0), a_2(a_1), a_3(a_2)...\rangle$. The sequence $\langle D_n, j_n\rangle$ is called *projective* (or *inverse*) *system* and its limit is a cpo called *inverse limit*, also denoted D_∞, where:

$$D_\infty = \{x \in (\bigcup_n D_n)^{\mathbb{N}} | \forall n(x(n) \in D_n) \wedge j_n(x(n+1)) = x(n)\}$$

is a set of infinite sequences of functions (we can write x_n in place od $x(n)$), orderd pointwise: $x \leq y$ if an only if $\forall n(x(n) \leq y(n))$. For directed $X \subseteq D_\infty$ the supremum is defined as $\sqcup X = \lambda n. \sqcup \{x(n) | x \in X\}$. Scott characterises this embedding by means of pairs of functions (i, j) where i is injective and j is surjective.

Definition 30. *A pair of mappings (i, j) is called* projection *(or* embedding-projection*, or* retraction pair*) of D' on D if:*

1. *$i : D \longrightarrow D'$ and $j : D' \longrightarrow D$ are both continuous.*
2. *$j \circ i = Id_D$ and $i \circ j \leq Id_{D'}$.*

The function j is properly called projection *and the function i is called* embedding.

The function i (embedding) is injective, while the function j (projection) is suriective and these functions determine uniquely each other. Embedding-projection pairs are weakening of the concept of isomorphism: point 2. means that going in the direction $D \xrightarrow{i} D' \xrightarrow{j} D$ all information is completely preserved and we are back exactly where we started, while, on the contrary, going in the direction $D' \xrightarrow{j} D \xrightarrow{i} D'$ we may lose some information, i.e. $i(j(x)) \leq x$. Let us start by considering the projection (i_0, j_0) where $i_0 : D_0 \longrightarrow D_1$ and $j_0 : D_1 \longrightarrow D_0$ are defined as $i_0(x)(y) = x$ and $j(f) = f(\bot)$. Now, suppose a projection (i_n, j_n) has been defined, where $i_n : D_n \longrightarrow D_{n+1}$ and $j_n : D_{n+1} \longrightarrow D_n$. Hence it is provable that the following pair (i_{n+1}, j_{n+1}) is a projection too, where:

1. $i_{n+1}(f) = i_n \circ f \circ j_n$
2. $j_{n+1}(g) = j_n \circ g \circ i_n$

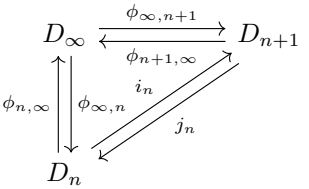
for $f \in D_{n+1}$ and $g \in D_{n+2}$. Our D_∞ is the inverse limit of the sequence $\langle D_m, j_n\rangle_{n\in\mathbb{N}}$, where $\bot = \langle \bot_0, \bot_1, \bot_2....\rangle$, where $\bot_n$ is the least member of D_n and $\sqcup X = \langle \sqcup X_0, \sqcup X_1, \sqcup X_2...\rangle$ with directed $X \subseteq D_\infty$ and $X_n = \{x(n) | x \in X\}$. In particular the projection $(\phi_{n,\infty}, \phi_{\infty,n})$ which embeds D_n into D_∞ is considered, where $\phi_{n,\infty} = \langle \phi_{n,k}(x)\rangle_{k\in\mathbb{N}}$ and $\phi_{\infty,n}(x) = x_n$ and:

$$\phi_{n,m} = \begin{cases} j_m \circ \phi_{n,m+1} & \text{if } n > m \\ i_{m-1} \circ \phi_{n,m-a} & \text{if } n < m \\ Id & \text{otherwise} \end{cases} \tag{4}$$

In other words, if $n < m$, then $\phi_{n,m} = i_{m-1} \circ i_{m-2} \circ i_{m-3} \circ ... \circ i_{n+1} \circ i_n$ and if on the contrary $n > m$, then $\phi_{n,m} = j_m \circ j_{m+1} \circ j_{m+2} \circ ... \circ j_{n-2} \circ j_{n-1}$. Hence $\phi_{n,\infty}$ has the form $y_0, y_1, y_2....$ where:

$$
\begin{aligned}
&y_0 = j_0(j_1(j_2(...(y_{n-1}(x)...)\\
&y_1 = j_1(j_2(...(y_{n-1}(x)...)\\
&\vdots\\
&y_{n-k} = j_{n-k}(...(y_{n-1}(x)...)\\
&\vdots\\
&y_{n-1} = j_{n-1}(x)\\
&y_n = x\\
&y_{n+1} = i_n(x)\\
&y_{n+2} = i_{n+1}(i_n(x))\\
&\vdots\\
&y_{n+k} = i_{n+k-1}(...(i_n(x)...)\\
&\vdots
\end{aligned}
$$

For instance, $\phi_{2,\infty}(x)$ has the form $\langle j_0(j_1(x)), j_1(x), x, i_2(x), i_3(x), i_4(x)...\rangle$, where $j_0(j_1(x)), j_1(x)$ are approximations of x and $i_2(x), i_3(x), i_4(x)...$ are copies of it. Since the range of $\phi_{n,\infty}$ is an isomorphic copy of D_n in D_∞ we can think of $\phi_{n,\infty}(x)$ just as x. Under this convention, since it is not difficult to prove that $\phi_{n,\infty}(x_n) \leq \phi_{n+1,\infty}(x_{n+1})$ and $x = \sqcup_n \phi_{n,\infty}(x(n))$, then we can look at the sequence $x_0, x_1, x_2....$ as a series of successive approximations to x and (up to isomorphisms) $D_0 \subseteq D_1 \subseteq D_2 \subseteq ... \subseteq D_\infty$. Notice that $\phi_{\infty,n+1} = i_n \circ \phi_{\infty,n}$ and $\phi_{n+1,\infty} = \phi_{\infty,n} \circ j_n$, where $\phi_{n,\infty} : D_n \longrightarrow D_\infty$ and $\phi_{\infty,n} : D_\infty \longrightarrow D_n$. Hence if we consider the category of cpo's with projective (or retraction) pairs as morphisms, we see that the following diagram commutes:



It follows that D_∞ is an upper bound for the sequence $D_0, D_1, D_2...$ Without investigating in depth the interesting structure of D_∞, let us emphasise a few facts. The first fact is that if $x \in D_n$, then $i_n(x) \in D_{n+1}$, hence this second element has in D_∞ the form:

$$\langle \phi_{n+1,0}(x), \phi_{n+1,1}(x), ..., \phi_{n+1,n+1}(x), ...\rangle$$

However $\phi_{n+1,m} = \phi_{n,m} \circ j_n$, so the above sequence actually is:

$$\langle \phi_{n,0}(x), \phi_{n,1}(x), ..., x, i_n(x), ..., \phi_{n,m}(x), ...\rangle$$

which is just x. It follow that if $x \in D_n$, then $i_n(x) = x$. With similar arguments we also obtain that if $x \in D_{n+1}$, then $j_n(x) \leq x$. The second fact to which we would like to draw attention is that if $n \leq k$, $x \in D_{n+1}$ and $y \in D_k$, then $x(y_n) = x_{k+1}(y)$. Actually this is immediate for $k = n + 1$, since $i_{n+1}(x) = i_n \circ x \circ j_n$ and therefore $i_{n+1}(x)(y) = i_n(x(j_n(y)) = i_n(x(y_n)) = x(y_n)$ by definition of projection and of the structure of D_∞. Analogously if $n \leq k$, $x \in D_{k+1}$ and $y \in D_n$, then $x_{n+1}(y) = (x(y_k))_n$. We will use these remarks in what follows. The next step consist in defining the application as follows, for $x, y \in D_\infty$:

$$x \cdot y = \sqcup_n \phi_{n,\infty}(x_{n+1}(y_n))$$

and $F(x)(y) = x \cdot y$ and its inverse G, where $G(f) = \sqcup_n G_n(f)$, where $G_n(f)(y) = (f(y))_n$. Application is continuous in both variables, i.e. $x \cdot \sqcup X = \sqcup\{x \cdot y | y \in X\}$ and $\sqcup X \cdot y = \sqcup\{x \cdot y | x \in X\}$.

Theorem 66. *The maps F and G are inverse isomorphisms.*

Proof. In these arguments we use, without explicitly mentioning it, the continuity of the functions involved:

1. $G \circ F = Id$. This follows from the fact that if $y \in D_n$, then $(x \circ y)_n = x_{n+1}(y)$ and this is a consequence of the second of the previous remarks:

$$(x \circ y)_n = (\sqcup_k (x_{k+1}(y)_k)_n = \sqcup_k (x_{k+1}(y)_k)_n = \sqcup_{k \geq n} (x_{k+1}(y)_k)_n = \sqcup_{k \geq n} x_{n+1}(y) = x_{n+1}(y)$$

   Hence $G(F(x)) = \sqcup_n x_{n+1} = x$.
2. $F \circ G = Id$. First observe that, thanks to the above remarks:

$$(G(f)) \cdot x = \sqcup_i (G(f))_i(x_i) = \sqcup_{i \geq n} (G(f))((x_i)_n) = (G(f))(x_n)$$

   Hence by definition $G(f) \cdot x = \sqcup_n (f(x_n))_n$. Still by using continuity of f we obtain:

$$G(f) \cdot x = \sqcup_n (f(x_n))_n = \sqcup_{p \geq n} (f(x_n))_p = f(x)$$

QED

Hence we have obtained a "solution" to the equation $D = [D \longrightarrow D]$, as desired. The functions F and G are used to interpret terms, as in the cases we have already analysed. Moreover, this model is extensional, in the sense that if $a \cdot c = b \cdot c$, for all c, then $a = b$. We show that $a_{n+1} = b_{n+1}$, that is, for all $x \in D_n$, $a_{n+1}(x) = b_{n+1}$: take $c = \phi_{n,\infty}(x)$, so that $\phi_{r,n}(x) = c_n$; but we have seen that $(a \cdot c)_n = a_{n+1}(x)$ and $(b \cdot c)_n = b_{n+1}(x)$ and therefore if $a \cdot c = b \cdot c$ we conclude that $a_{n+1}(x) = b_{n+1}(x)$, for all $x \in D_n$. In particular $a_0 = j_0(a_1) = j_0(b_1) = b_0$ (see Amadio and Curien (1996) 67-76, Hindley and Seldin (2008) 256-270 and Barendregt (1984) 477-486 for a more detailed argument).

3.6. Feasibility in lambda calculus: guide for further study

So far we have talked mainly of definability in untyped lambda calculus of computable functions, but we have established little with regard to lower complexity and to the main complexity classes, only that the functions that can be defined in the simply typed lambda calculus are the extended polynomials. What about the polynomial time computable functions? In line with the purpose of these lecture notes we indicate now some research developments in this area more specifically oriented in the direction of *computational complexity*. This research has developed mainly in two directions. The first one comes from the area of (second order) *Linear Logic*: *Light Linear Logic*, *Light Affine Logic*, *Soft Linear Logic* etc. (see among others Girard (1998), Asperti (1998), Baillot and Mogbil (2004), Lafont (2004), Asperti and Roversi (2002)) propose different characterisations of functions computable in polynomial time. Such functions are representable by proofs within the proofs-as-programs correspondence and strategies are found which normalize them in polynomial time. These logics are *intrinsically polytime* in that any proof expressed in the so-called *proof nets*, or term of a lambda-calculus, can be converted into a normal (cut-free) one in polynomial time in the dimension of some parameter relative to the proof or term. Gentzen's sequent calculus contains, in addition to the logical rules, the so-called structural rules, in particular weakening and contraction (see sections 7.1 and 7.2 for an introduction to this calculus, of cut-elimination and the role of structural rules). Both, although perfectly plausible from a classical point of view, clash with the principles of *Linear Logic*, which is a resource-sensitive logic, where two resources do not count as one. Furthermore, the unrestricted use of the contraction rule actually causes an exponential explosion of complexity of the process of computation, i.e. of cut-elimination. In *Linear Logic*, contracting or weakening a formula is not permitted unless it is authorised by means of two operators ! (of course) and ? (why not). The !-modality corresponds to contraction and weakening on the left hand side, and the ?-modality corresponds to contraction and weakening on the right hand side. However, the cut-elimination complexity is still superexponential, but this

idea of controlling contraction through modal operators can be further developed, resulting in the *Light Linear Logic* that, unlike another variant, called *Bounded Linear Logic*, does not use explicit parameters such as $! \leq n$, which makes the system not purely logical. Rather, a new modality § ("neutral") is in this case added which refers to the unary cases of ! and ?. The associated calculus no longer allows certain laws of *Linear Logic* to be derived, such as $!\alpha \Longrightarrow \alpha$ ("dereliction ") or $!\alpha \Longrightarrow !!\alpha$ ("digging"), which are, however, compensated for by other laws as a weak version of the former $!\alpha \Longrightarrow \S\alpha$ or $\S\alpha \Longrightarrow ?\alpha$. This ends up being sufficient to prevent the uncontrolled explosion of computational complexity. Every polynomial time function is representable by an proof of this calculus, and every proof is normalizable in polynomial time.

Light Affine Logic is a variant of the *Light Linear Logic* that was introduced in Asperti (1998) by admitting the full weakening rule. This move does not undermine progress in complexity, but greatly simplifies the calculus. In Baillot (2004) it is shown that *Intuitionistic Light Affine Logic* provides a typing for lambda-calculus which guarantees that a well-typed program is executable in polynomial time. Echoing an observation by Girard that normalisation in polynomial time is largely independent of types, Terui (2007) introduces instead an *untyped light affine lambda calculus*, which embodies the essentials of this logic, with the remarkable strong normalization property that each term is normalizable in polynomial time regardless the strategy applied. Logic is then re-introduced as a Curry-style type assignment system for this calculus.

We will not deal with this here because of the large number of prerequisites that we would have to set in advance to address this issue in detail. Rather, we conclude by broadly stating what are the insights behind the second direction of the research, that proposes restrictions such that the functions definable thus obtained are those of PTIME and it is more accessible with the tools we have introduced so far: this is for instance the approach that originated in Bellantoni and Cook (1992). In Leivant (1990) and Leivant and Marion (1993) the underling idea is that data are used computationally at different levels of abstraction and this is called *data ramification*, or *data tiering*. The proposal consists of an *implicit* characterisation of complexity classes, i.e. by language restrictions rather than by explicit resource bounds. The first paper proposes a particular formalism in which the arguments of a function on binary strings are divided into two blocks separated by a semi-colon $f(x_0, ..., x_n; y_0, ..., y_k)$. Those occurring to the left of the semi-colon are called *normal*, and variables to the right are called *safe*. Performing a recursion, the previous step value must be placed in *safe* position:

1. $f(0, x; y) = g(x; y)$
2. $f(z * i, x; y) = h_i(z, x; y, f(z, x; y))$ $(i = 0, 1)$

The composition scheme ensures that the recursive value will stay in a safe position and will not be copied into a normal position:

$$f(x; y) = h(g_0(x;), ..., g_m(x;); h_0(x; y), ..., h_k(x; y))$$

This is an example of use of tiering techniques. The polynomial time functions will be exactly those functions which have all normal inputs, i.e. no safe inputs. As for the lambda calculus, the second paper introduce a notion of "tiering" in the framework of a typed lambda-calculi with a base type o, by extending the set of terms of the typed calculus we have considered in the strong normalization theorem with new operators $\mathsf{0}$, $\mathsf{1}$ and predecessor $\mathsf{p}(cx) = x$ (for $c = 0, 1$), all of type $\mathsf{o} \to \mathsf{o}$ and discriminator function d of type $\mathsf{o} \times \mathsf{o} \times \mathsf{o} \times \mathsf{o} \to \mathsf{o}$ and an operator ϵ (the empty string) of type o, and show that the polynomial-time functions are precisely the functions representable with *abstract* input (Church-like numerals) and *concrete* output (represented by binary strings). To explain this idea, let us consider the set of binary strings $\{\mathsf{0}, \mathsf{1}\}^*$ (the free algebra generated by $\mathsf{0}, \mathsf{1}$ and ϵ). Observe that a word e.g. $w = (\mathsf{0}(\mathsf{1}(\mathsf{1}\epsilon)))$ can be also represented abstractly *à la Church* as $w^* = \lambda x_0 \lambda x_1 \lambda x_\epsilon . x_0(x_1(x_1 x_\epsilon)))$. The reader can check that this is a term of type $\omega[\tau] = (\tau \to \tau) \to (\tau \to \tau) \to \tau \to \tau$.

When w^* has type $\omega[\mathsf{o}^q]$, where $\mathsf{o}^q = \overbrace{\mathsf{o} \times ... \times \mathsf{o}}^{q-times}$, we will write $w^{*,q}$. A function $f : \{0, 1\}^k \to \{0, 1\}^r$ is *explicitly* defined in this calculus by a term E of type $\mathsf{o}^k \to \mathsf{o}^r$, if $\mathsf{E}w_0...w_{k-1} =$

$f(w_0...w_{r-1})$. It is instead *2-tiers* definable by a term E of type $\omega[\mathsf{o}_0^q] \to ... \to \omega[\mathsf{o}_{k-1}^q] \to \mathsf{o}$, for some $q_0, ..., q_{k-1}$, if $\mathsf{E}w_0^{*,q_0}...w_{k-1}^{*,q_{k-1}} = f(w_0...w_{r-1})$. Well, it is provable that a function over $\{\mathsf{0},\mathsf{1}\}^*$ is polynomial time computable if and only if it is *2-tier* definable in this calculus.

Part II The incompleteness theorems

4. First and second Gödel's theorems and related results

4.1. Definability and representability

We discussed the definability of the sets, so let's begin this section by talking in particular about the definability *of functions*, to come next to their *representability*, as in Gödel's original approach, when he proved that every computable function and computable set is representable in Robinson's Arithmetic Q. Roughly speaking, the notion of definability is semantic, given in terms of the notion of *truth*, while representability is a corresponding tighter syntactic notion. We will later see the close relationship between these two concepts. To our pourpose, it will be necessary to investigate the Σ_1 definability in the standard model. Saying that $\phi(x_0, ..., x_k)$ is Σ_1-definable e.g. in the standard model we mean therefore to say that *its graph*:

$$G_\phi = \{\langle x_0, ..., x_k, y\rangle | \phi(x_0, ..., x_k) \simeq y\}$$

is Σ_1-definable in it. We will see later the relationship with the notion of representability.

Theorem 67. *Consider the language of* PA. *The following holds:*

1. *If a set X is Δ_0-definable in the standard model, then X is primitive recursive.*
2. *X is Σ_1-definable in the standard model, iff X is computably enumerable*
3. *ϕ is Σ_1-definable in the standard model, iff ϕ is partial recursive.*

Proof. The proof is articulated in several passeges. As for 1., note that all terms of the language based on $+, \times, S, 0, \leq$ denote primitive recursive functions. In addition the bounded quantifiers do not get out of this level. Regarding point 2., we have already seen at p. 43. Finally 3. $\Rightarrow$, if $G_\phi = \{\langle x, z\rangle | \phi(x) \simeq z\}$ is Σ_1, then is r.e and therefore is the domain of a partial recursive function $\phi_e(x, z)$. Hence $\langle x, z\rangle \in G_\phi$ iff $\exists y T(e, x, z, y)$. Let therefore:

$$\phi(x) \simeq (\mu u.u = \langle z, y\rangle \wedge T(e, x, (u)_0, (u)_1))_1$$

The crucial direction is $\Leftarrow$: all partial recursive functions are Σ_1-definable in the language of PA. We associate to any function $\phi(x)$, a Σ_1-formula $\psi_\phi(x, y)$ defining its graph G_ϕ, that is to say:

$$\psi_\phi(x, y) \text{ is true iff } \langle x, y\rangle \in G_\phi \text{ iff } \phi(x) \simeq y$$

1. *Initial functions*:

$Function$	$Definition$
$Z(x) \simeq 0$	$(x = x \wedge y = 0)$
$S(x) \simeq x + 1$	$(y = S(x))$
$U_i^m(x_0, ..., x_n) = x_i$	$\bigwedge_{0 \leq j \leq n} x_j = x_j \wedge x_i = y$

(1)

Duccio Pianigiani, University of Siena, Italy, duccio.pianigiani@unisi.it, 0000-0001-9441-7226
Referee List (DOI 10.36253/fup_referee_list)
FUP Best Practice in Scholarly Publishing (DOI 10.36253/fup_best_practice)

Book References DOI 10.36253/979-12-215-0778-2.references

2. *Composition*: Suppose ψ_g, $\psi_{h_0}, ..., \psi_{h_k}$ let Σ_1-definitions of the functions $g, h_0, ..., h_k$ and let $\phi(x) \simeq g(h_0(x), ..., h_k(x))$.
Let therefore $\psi_\phi(x, z)$ the Σ_1-formula:

$$\exists y_0 ... \exists y_k(\psi_{h_0}(x, y_0) \wedge ... \wedge \psi_{h_k}(x, y_k) \wedge \psi_g(y_0, ..., y_k, z))$$

3. *Minimization*: let $\phi(x) \simeq \mu z.g(x, z) \simeq 0$ where $\forall v < z(g(x, v) \downarrow \wedge g(x, v) \neq 0)$ and let ψ_g a Σ_1-definition of g; let then $\psi_\phi(x, z)$ the following Σ_1-formula:

$$\psi_g(x, z, \overline{0}) \wedge \forall v < z \exists w(\psi_g(x, v, w) \wedge \neg(w = \overline{0}))$$

4. *Primitive recursion*: let ϕ defined as follows:

   (a) $\phi(x, 0) \simeq h(x)$

   (b) $\phi(x, y + 1) \simeq g(x, y, \phi(x, y))$

   Intuitive idea. Gödel's idea was to formalize the steps of computation of $\phi(x, y) \simeq z$ as a sequence $\langle s_0, ..., s_y \rangle$, where:

   (a) $s_0 = h(x)$

   (b) $s_{i+1} = g(x, i, s_i)$

   (c) $s_y = z$

   To this goal, we need to develop functions that handle sequences. When we had primitive recursion, we could define things like the $n -$ th prime and code finite sequences by means of primes and factorization. But here we do not have primitive recursion: in fact we want to show that we can do primitive recursion. Hence Gödel used the machinery based on the so-called *Chinese remainder theorem*:

   > "Given $m_0, ..., m_k$ pairwise coprime (that is, each pair of them has no common divisors > 1), if $s_0 < m_0, ..., s_k < m_k$, then there exist a unique B such that $B < \Pi_{i \leq k} m_i$ and $Rem(B, m_i) = s_i$ (i.e. s_i is the remainder of the division of B by m_i)".

   But how to obtain a succession of $y + 1$ numbers pairwise coprime? Given $s_0, ..., s_y$ let us define $\nu = max\{s_0, ..., s_y, y + 1\}$ and $A = \nu!$. Hence the numbers $1 + A, 1 + 2A, ..., 1 + (y + 1)A$ are pairwise coprime (and clearly, for all $i \leq y$, $s_i < 1 + (i + 1)A$). To check it, suppose that it is not true: let then p prime that divides both $1 + rA$ and $1 + r'A$; but then it will divide also their difference $(r - r')A$. Also applies in general that if $p|ab$ (with p prime), then either $p|a$, or $p|b$. Thus in our case, either $p|(r - r')$, or $p|A$. Since A is a multiple of $r - r'$ (considering that $r, r' \leq y + 1$) we have that $(r - r')|A$. Ultimately we have these alternatives:

   (a) either $p|(r - r')$, and then $p|A$,

   (b) or $p|A$.

   In both cases $p|A$. Hence $Rem(1+rA, p) = 1$ (but this is contradictory, because, by assumption $p|1 + rA$).
   Equivalent formulation of the intuitive idea. We can thus formulate the intuitive idea by saying, equivalently, that there are A, B such that:

   (a) $Rem(B, A + 1) = h(x)$

   (b) $Rem(B, A(y + 1) + 1) = z$

   (c) $\forall i < y(Rem(B, A(i + 2) + 1) = g(x, i, Rem(B, A(i + 1) + 1))$

Let us denote with (*) these conditions. Suppose these conditions (*) hold. Actually, defining $s_i = Rem(B, A(i+1)+1)$, the sequence $s_0, ..., s_y$ satisfies Gödel's intuitive idea. On the other hand, if $s_0, ..., s_y$ is a sequence that satisfies the conditions expressed by the intuitive idea and A is defined as above, then for the Chinese remainder theorem there is a number B such that each s_i can be expressed as $s_i = Rem(B, A(i+1)+1)$ in such a way that the above conditions (*) are met.
Formalization of the intuitive idea. At this point we define a formula that we denote $\beta(B, A, i, v)$, which expresses in the formal language of arithmetic the relation $(Rem(B, A \cdot (i+1)+1) = v)$, by means of a Δ_0-formula:

$$v < \overline{A} \cdot (i + \overline{1}) + \overline{1} \wedge \exists w \leq \overline{B}(\overline{B} = w \cdot (\overline{A} \cdot (i + \overline{\overline{1}}) + \overline{1}) + v)$$

Let us finally consider these three formulas:

(a) $\exists w(\beta(\overline{B}, \overline{A}, \overline{0}, w) \wedge \psi_h(x, w))$

(b) $\beta(\overline{B}, \overline{A}, u, z)$

(c) $\forall i < y \exists v \exists u(\beta(\overline{B}, \overline{A}, i, v) \wedge \beta(\overline{B}, \overline{A}, i+1, u) \wedge \psi_g(x, i, v, u))$

Note that the conjunction of these three formulas is Σ_1.
If now we call for brevity $\Theta_a(x, \overline{B}, \overline{A}), \Theta_b(\overline{B}, \overline{A}, y, z), \Theta_c(x, \overline{B}, \overline{A})$ respectively, these three formulas, finally get the desired Σ_1 formula $\psi_\phi(x, y, z)$:

$$\exists A \exists B(\Theta_a(x, B, A) \wedge \Theta_b(B, A, y, z) \wedge \Theta_c(x, B, A))$$

Note that we did not use the power-of-primes coding of sequences. The particular coding of sequences that we have shown here (the $\beta-$function) dates back from the same Gödel.

QED

As anticipated, Gödel used the notion of *representability*, restricted to recursive primitive functions (the proof of representability is almost the same), rather than definability in the model (see Odifreddi (1989-1999) 39-44).

Definition 31. *Let* T *be a theory that extends the predicate logic with identity and that contains terms* $\overline{n}$ *for all natural numbers* n*. Then, if* f *is a function, we say (for all sequence of natural numbers* $n_0, ..., n_m$*) that:*

1. f *is* weakly representable *in* T *(or* numerable*), if for some formula* ϕ *of the language of the theory* T*, we have that* $f(n_0, ..., n_m) = r$ *iff* $\mathsf{T} \vdash \phi(\overline{n_0}, ..., \overline{n_m}, \overline{r})$.
2. f *is* representable *in* T *(or* binumerable*), if for some formula* ϕ *of the language of* T*, we have that:*
   (a) $f(n_0, ..., n_m) = r$ *implies* $\mathsf{T} \vdash \phi(\overline{n_0}, ..., \overline{n_m}, \overline{r})$ *and*
   (b) $f(n_0, ..., n_m) \neq r$ *implies* $\mathsf{T} \vdash \neg\phi(\overline{n_0}, ..., \overline{n_m}, \overline{r})$.
3. f *is* strongly representable *in* T*, if moreover:*

$$\mathsf{T} \vdash \forall y \forall z(\phi(n_0, ..., n_m, y) \wedge \phi(n_0, ..., n_m, z) \rightarrow y = z)$$

The condition 3. joint to the 2.(a) is equivalent to:

$$\mathsf{T} \vdash \forall y(\phi(\overline{n_0}, ..., \overline{n_m}, y) \leftrightarrow y = \overline{f(n_0, ..., n_m)})$$

These definitions are extended to relations.

Definition 32. *Let* R *be a relation and* T *as in the previous definition:*

1. *R is* weakly representable *in* T *(or* numerable*), if for some formula ϕ of the language of* T*, we have that $R(n_0, ..., n_m)$ is true iff* $\mathsf{T} \vdash \phi(\overline{n_0}, ..., \overline{n_m})$.
2. *R is* representable *in* T *(or* binumerable*), if for some formula ϕ of the language of* T*, we have that if $R(n_0, ..., n_m)$ is true, then* $\mathsf{T} \vdash \phi(\overline{n_0}, ..., \overline{n_m})$ *and if $R(n_0, ..., n_m)$ is false, then* $\mathsf{T} \vdash \neg\phi(\overline{n_0}, ..., \overline{n_m})$

Let us now look at some relationships between the concepts we have introduced (see Hájek and Pudlák (1993), pp. 155-57).

Theorem 68. *The notions of representability and definability are related as follows*

1. *If R is defined in $\mathbb{N}$ by a $\phi \in \Sigma_1$, then ϕ* numerates *R in* Q.
2. *If R is Δ_1 defined in $\mathbb{N}$, then there is a $\theta \in \Sigma_1^0$ that* binumerates *R in* Q.

Proof. 1. It follows immediately from Σ_1-soundness and Σ_1-completeness of Q. Regarding 2. if $R(x_0, ..., x_n)$ is Δ_1-definable, then exists a Σ_1-formula $\exists y\phi(x_0, ..., x_n, y)$ and a Π_1-formula $\forall y\psi(x_0, ..., x_n, y)$ that define it, with $\phi, \psi \in \Delta_0$. Let now $\theta(x_0, ..., x_n)$ the formula:

$$\exists y(\phi(x_0, ..., x_n, y) \wedge \forall z \leq y\psi(x_0, ..., x_n, z))$$

Observe that:

1. if $\langle k_0, ..., k_n\rangle \in R$, then there exists an m such that $\mathbb{N} \models \phi(\overline{k_0}, ..., \overline{k_n}, \overline{m})$ and $\mathbb{N} \models \forall y\psi(\overline{k_0}, ..., \overline{k_n}, y)$. In particular, for all $s \leq m$, we have that $\mathbb{N} \models \psi(\overline{k_0}, ..., \overline{k_n}, s)$.
   It follows that Q proves $\phi(\overline{k_0}, ..., \overline{k_n}, \overline{m}) \wedge \forall z \leq \overline{m}\psi(\overline{k_0}, ..., \overline{k_n}, z)$, from which immediately $\theta(\overline{k_0}, ..., \overline{k_n})$.
2. $\langle k_0, ..., k_n\rangle \notin R$, then $\mathbb{N} \models \neg\forall y\psi(\overline{k_0}, ..., \overline{k_n}, y)$ e $\mathbb{N} \models \neg\exists y\phi(\overline{k_0}, ..., \overline{k_n}, y)$. Hence there exist an m, such that the formula:

   $$\neg\psi(\overline{k_0}, ..., \overline{k_n}, \overline{m}) \wedge \forall y \leq \overline{m}\neg\phi(\overline{k_0}, ..., \overline{k_n}, y)$$

   is true and also provable in Q.
   Reason therefore inside the theory Q and consider that if the formula $\theta(\overline{k_0}, ..., \overline{k_n})$ were true, then for some u would be true also $\phi(\overline{k_0}, ..., \overline{k_n}, u) \wedge \forall y \leq u\psi(\overline{k_0}, ..., \overline{k_n}, y)$.
   We have therefore the following alternatives:

   (a) if $u \leq \overline{m}$, then $\neg\phi(\overline{k_0}, ..., \overline{k_n}, u)$, from our assumptions, however, and for (2) we will also $\phi(\overline{k_0}, ..., \overline{k_n}, u)$ (contradiction).

   (b) $u > \overline{m}$, then $\psi(\overline{k_0}, ..., \overline{k_n}, \overline{m})$, for (2), but at the same time $\neg\psi(\overline{k_0}, ..., \overline{k_n}, \overline{m})$, by assumptions.

   Contradiction. Hence $\mathsf{Q} \vdash \neg\theta(\overline{k_0}, ..., \overline{k_n})$.

It follows that $\theta(x_0, ..., x_n)$ is the desired binumeration. QED

Corollary 15. *All recursive relations (being Δ_1-definable) are binumerable in* Q*, and all computably enumerable relations (being Σ_1-definable) are numerable in* Q.

To summarise, the following equivalences hold:

1. As for the relations, R is r.e. iff R is Σ_1-definable iff R is weakly representable (numerable). Moreover R is recursive iff R is Δ_1-definable iff R is representable (binumerable).
2. As for the functions, ϕ is a *partial* recursive function iff the graph of ϕ is Σ_1-definable iff ϕ is numerable in Q. Moreover f is a total recursive function iff the graph of f is Δ_1-definable iff f is binumerable in Q.

Axiomatizability. In the current literature, sometimes it is only required that the set of axioms or of the theorems is recursively enumerable; sometimes instead, more strictly, it is assumed that the axiomatic theories are axiomatizable in primitive recursive way. That these alternatives are ultimately equivalent, can be proved thanks to variants of a method known as "Craig's Trick". Define two theories to be deductively equivalent if they prove the same theorems. In other words, they are two different axiomatizations of a deductively closed set of formulas.

Theorem 69. *Suppose the theory* T *is computably enumerable, i.e., there is a computable function that lists the set* Thm_{T} *of its theorems; then the following are equivalent:*

1. T *is axiomatizable in primitively recursive way.*
2. Thm_{T} *is recursively enumerable.*

Proof. We proof 2. $\Rightarrow$ 1. Remember first that if a set is computably enumerable, then is the codomain of a *primitive recursive* function (see the remark on p.103). So let f primitive recursive such that $Thm_{\mathsf{T}} = f[\mathbb{N}]$. Now we apply the so-called *Craig's Trick* and replace each $f(n) = \ulcorner \alpha_n \urcorner$ with $\overbrace{\ulcorner \alpha_n \wedge ... \wedge \alpha_n \urcorner}^{n-times}$. Observe that the theory T^* axiomatized by the set so obtained is deductively equivalent to T. Moreover, being enumerable in increasing order, arguing as on p.43 it is recursive. Actually, by a similar argument, we can show that it is *primitive* recursive, since we ca check whether $\ulcorner \alpha \urcorner \in \mathsf{T}^*$ in a primitive recursive way as follows: count the number m of conjuncions in α, the compute $f(0), f(1), f(2), ..., f(m+1)$. Hence $\ulcorner \alpha \urcorner \in \mathsf{T}^*$ if it is a conjunction having code $f(i)$, for some $i \leq m+1$. 1. $\Rightarrow$ 2. we already know that Thm_{T} is creative. More constructively we will see that in the relation:

$$\exists d[d \text{ is a proof of } \phi \text{ from } \Gamma]$$

the part in brackets is computable, when Γ is computable. Note that we have defined Thm_{T} by means of a Σ_1 formula. QED

4.2. Arithmetization of metamathematics

We need to encode metamathematical concepts: certain assertions about formulas will be converted into assertions about natural numbers and thus expressed in the formal language, and therefore we will express facts *about formulas* by expressing facts *about numbers*. *Arithmetizing* a certain discrete domain means encode objects and notions of that domain using natural numbers; in particular, symbols, formulas (thought of as chains of symbols) and statements about formulas of a formal language can be converted to natural numbers and statements about the natural numbers using a numerical coding.

Suppose we have assigned a number to each logical constant, e.g.:

$\forall$	$\exists$	$\wedge$	$\vee$	$\neg$	$\rightarrow$	$=$	$($	$)$
$\langle 0,0\rangle$	$\langle 0,1\rangle$	$\langle 0,2\rangle$	$\langle 0,3\rangle$	$\langle 0,4\rangle$	$\langle 0,5\rangle$	$\langle 0,6\rangle$	$\langle 0,7\rangle$	$\langle 0,8\rangle$

1. variables x_i are coded as pairs $\langle 1, i\rangle$ (Possibly i can be written in unary or binary notation).
2. a constant c_i will be coded by $\langle 2, i\rangle$;
3. an n-ary function symbol f_j will be coded as $\langle 3, j, n\rangle$;
4. an n-ary relational symbol P_j will be coded by $\langle 4, j, n\rangle$

If $s_0, ..., s_n$ is a sequence of symbols, and c_i is the code of the symbol s_i, its Gödel number is the coded sequence:

$$\langle c_0, ..., c_n\rangle = p_0^{n+1} \cdot p_1^{c_0} \cdot ... \cdot p_n^{c_n}$$

1. a generic composite term $f_j(t_0, ..., t_m)$ will be coded as a sequence:

$$\ulcorner f_j(t_0, ..., t_m)\urcorner = \langle \ulcorner f_j \urcorner, \ulcorner (\urcorner, \ulcorner t_0 \urcorner, ..., \ulcorner t_m \urcorner, \ulcorner) \urcorner \rangle$$

2. A generic atomic formula $P_j(t_0, ..., t_m)$ will be coded as a sequence:

$$\ulcorner P_j(t_0, ..., t_m)\urcorner = \langle \ulcorner P_j \urcorner, \ulcorner(\urcorner, \ulcorner t_0 \urcorner, ..., \ulcorner t_m \urcorner, \ulcorner)\urcorner \rangle$$

The non atomic formulas will be finally codified in this way:

(a) $\ulcorner(\neg\phi)\urcorner = \langle \ulcorner(\urcorner, \ulcorner\neg\urcorner, \ulcorner\phi\urcorner, \ulcorner)\urcorner \rangle$

(b) $\ulcorner(\alpha \vee \beta)\urcorner = \langle \ulcorner(\urcorner, \ulcorner\alpha\urcorner, \ulcorner\vee\urcorner, \ulcorner\beta\urcorner, \ulcorner)\urcorner \rangle$

(c) $\ulcorner(\alpha \wedge \beta)\urcorner = \langle \ulcorner(\urcorner, \ulcorner\alpha\urcorner, \ulcorner\wedge\urcorner, \ulcorner\beta\urcorner\ulcorner)\urcorner \rangle$

(d) $\ulcorner(\alpha \leftrightarrow \beta)\urcorner = \langle \ulcorner(\urcorner, \ulcorner\alpha\urcorner, \ulcorner\leftrightarrow\urcorner, \ulcorner\beta\urcorner, \ulcorner)\urcorner \rangle$

(e) $\ulcorner(\alpha \rightarrow \beta)\urcorner = \langle \ulcorner(\urcorner, \ulcorner\alpha\urcorner, \ulcorner\rightarrow\urcorner, \ulcorner\beta\urcorner, \ulcorner)\urcorner \rangle$

(f) $\ulcorner(\forall x_j \beta)\urcorner = \langle \ulcorner(\urcorner, \ulcorner\forall\urcorner, \langle 0, j\rangle, \ulcorner\beta\urcorner.\ulcorner)\urcorner \rangle$

(g) $\ulcorner(\exists x_j \beta)\urcorner = \langle \ulcorner(\urcorner, \ulcorner\exists\urcorner, \langle 0, j\rangle, \ulcorner\beta\urcorner, \ulcorner)\urcorner \rangle$

The idea goes back to Leibniz. There are many sophisticated ways to accomplish this, which in some cases take into account the complexity, but the essential point is that it is always possible to pass, purely mechanically, from an expression to its code number, and from a number to the corresponding expression. In turn, these coded expressions can be translated into the language of the same theory, if is sufficiently expressive: this is therefore a method which enables the theory *to speak indirectly of itself*, stratagem that allows us to formalize an argument based on *self-reference*.

The relations "x is a variable", "x is a constant", "x is an $n-$ary function symbol", "x is an $n-$ary predicative symbol", are clearly primitive recursive. For example "x is a variable" is defined as $Var(x) = \exists y < x(x = \langle 1, y\rangle)$ (analogously for the other concepts).

Lemma 26. *The relations $Term(x)$ and $Form(x)$, respectively "x is a variable", "x is a constant", "x is a term" and "x is a formula", are primitive recursive.*

Proof. To show that the predicate $Term(x)=$ "x is a term" is primitive recursive, we build a 'constructional history' for a term, or a term-sequence. Let $Termseq(n)$ encodes a term sequence, i.e. a sequence of expressions $t_0, t_1, ..., t_n$ such that each expression t_i in the sequence either is a constant; or is a variable; or else is built by using an $m - place$ function symbol from m terms occurring prior to place i:

$$Termseq(n) = Seq(n) \wedge \forall k \leq lh(n)[\Psi_0 \vee \Psi_1 \vee \Psi_2 \vee \Psi_3]$$

where:

1. $\Psi_0 = (n)_k = \ulcorner\overline{0}\urcorner$
2. $\Psi_1 = \exists j < k((n)_k = \langle \ulcorner S\urcorner, (n)_j\rangle)$
3. $\Psi_2 = \exists j < k \exists i < k((n)_k = \langle \ulcorner +\urcorner, (n)_i, (n)_j\rangle$
4. $\Psi_3 = \exists j < k \exists i < k((n)_k = \langle \ulcorner \cdot\urcorner, (n)_i, (n)_j\rangle$

Hence $Term(x)$ can be defined as $\exists y(Termseq(y) \wedge ((y)_{lh(y)} = x)$. But we need to find a bound to the existential quantifier. Notice that the last element x of the sequence is the bigger and this sequence y has the form $p_0^n \cdot p_1^{x_1} \cdot ... \cdot p_n^{x_n}$ where $x_n = x$. Hence a bound could be $y < (p_x!)^x$.

Now $Form(z)$ can be defined repeating the same strategy we used in defining Term where $Formseq(x)$ is a sequence such that for all $i \leq lh(x)$, either:

1. $(x)_i = \langle \ulcorner =\urcorner, y, z\rangle \wedge Term(y) \wedge Term(z)$, or
2. $(x)_i = \langle \ulcorner \wedge\urcorner, y, z\rangle \wedge Form(y) \wedge Form(z)$, or
3. (analogous conditions by replacing the conjunction with the other connectives), or

4. $(x)_i = \langle \ulcorner \forall \urcorner, y, z \rangle \wedge Var(y) \wedge Form(z)$, or
5. (analogous condition for the existential quantifier)".

where $y, z < x$, and the last element is z.

QED

Along the same lines, it is shown that the following are primitive recursive :

1. $ClTerm(x)$, "x is a closed term".
2. $FTVar(x, y)$, "y is a term that contains free occurrences of the variable with Godel number x".
3. $FFVar(x, y)$, "x is a free variable in the formula y".
4. $FreeF(x, y, z)$, "x is a term free for variable y in the formula z".
5. $Sent(x)$," x is the Gödel number of a sentence is primitive recursive."

If we consider axiomatic systems, the following are primitive recursive (for the sake of readability we will abbreviate $\langle \ulcorner(\urcorner, x, \ulcorner \rightarrow \urcorner, y, \ulcorner)\urcorner \rangle$ with $x \dot{\rightarrow} y$):

1. $Equax(x)$, "x encoding an axiom of equality".
2. $PropAx(x)$, "x is a propositional axiom".
3. $QAx(x)$, "x is a quantifiers axiom". Recall that the quantifier axioms are $\phi(t) \rightarrow \exists x \phi$ and $\forall x \phi \rightarrow \phi(t)$ (t free for x in ϕ). To say this we must say that there are $y, z, w \leq x$ such that the conjunction of the following formulas holds:

   (a) $Var(y) \wedge Form(z) \wedge Term(w) \wedge FreeF(w, y, z)$

   (b)
   $$(x = \langle \ulcorner(\urcorner, \ulcorner \forall \urcorner, y, z, \ulcorner)\urcorner \rangle \dot{\rightarrow} Subst(z, y, w)) \vee$$
   $$\vee (x = Subst(z, y, w) \dot{\rightarrow} \langle \ulcorner(\urcorner, \ulcorner \exists \urcorner, y, z, \ulcorner)\urcorner \rangle)$$

4. $Logax(x)$ iff "x code a logical axiom" i.e. the disjunction of $PropAx(x)$ and $QAx(x)$.
5. $MP(x, y, z)$ iff "y is obtained from z and x by Modus Ponens" as $z = x \dot{\rightarrow} y$.
6. Recall the rule of existential generalization: "from $\phi \rightarrow \psi$ conclude $\exists u \phi \rightarrow \psi$, assuming u not free in ψ". Now "x is obtained from y by existential generalization" can be expressed in primitive recursive way:

   $$\begin{aligned} &\exists v < x \exists w < x \exists z < x (Var(v) \wedge Form(w) \wedge Form(z) \wedge \\ &\wedge \neg FFvar(v, z) \wedge y = w \dot{\rightarrow} z \wedge x = \langle \ulcorner(\urcorner, \ulcorner \exists \urcorner, v, w, \ulcorner)\urcorner \rangle \dot{\rightarrow} z) \end{aligned} \tag{2}$$

   We could do the same for the universal generalization: "from $\phi \rightarrow \psi$ conclude $\phi \rightarrow \forall u \psi$, where u not free in ϕ", and so clearly we could then express $Gen(y, x)$ as " x is obtained from y by generalization (existential or universal)".

In the above, we also used this primitive recursive *substitution function*:

$$Subst(\ulcorner \phi \urcorner, i, \ulcorner t \urcorner) = \ulcorner \phi[t/x_i] \urcorner$$

whose effect is to take (Gödel number) of a formula $\phi(x_i)$ and replace in it the term t, in place of all occurrences of the variable x_i. This can be obtained by primitive recursion on the course of values:

1. Terms:

   (a)
   $$Subst(\ulcorner x_j \urcorner, i, y) = \begin{cases} y & \text{if } i = j \\ \ulcorner x_i \urcorner & \text{if } i \neq j \end{cases}$$

   (b) $Subst(\ulcorner f_i(t_0, ..., t_n) \urcorner, j, y) = \langle \ulcorner f_i \urcorner, Subst(\ulcorner t_0 \urcorner, i, y), ..., Subst(\ulcorner t_n \urcorner, i, y) \rangle$

2. Formulas:

   (a) $Subst(\ulcorner R_i(t_0, ..., t_n) \urcorner, j, y) = \langle \ulcorner R_i \urcorner, Subst(\ulcorner t_0 \urcorner, i, y), ..., Subst(\ulcorner t_n \urcorner, i, y) \rangle$

   (b) $Subst(\ulcorner \neg \phi \urcorner, j, y) = \langle \ulcorner \neg \urcorner, Subst(\ulcorner \phi \urcorner, j, y) \rangle$

   (c) $Subst(\ulcorner \phi \wedge \psi \urcorner, j, y) = \langle \ulcorner \wedge \urcorner, Subst(\ulcorner \phi \urcorner, j, y), Subst(\ulcorner \psi \urcorner, j, y) \rangle$ (analogously for $\vee, \rightarrow$)

   (d)
   $$Subst(\ulcorner \forall x_i \phi \urcorner, j, y) = \begin{cases} \langle \ulcorner \forall \urcorner, \ulcorner x_i \urcorner, Subst(\ulcorner \phi \urcorner, j, y) \rangle & \text{if } i \neq j \\ \ulcorner \forall x_i \phi \urcorner & \text{if } i = j \end{cases}$$

   (analogously for $\exists$)

   (e) $Subst(x, j, y) = 0$ in all other cases.

The provability predicate

For our purposes we need in particular a primitive recursive relation $\mathsf{Prf}_\mathsf{T}(x, y)$ whose intende meaning is "x code a correct proof of y in the theory T". Let's say that a theory is *presented in primitive recursive way* if the predicate Ax_T that defines the proper axioms of the theory T is in turn primitive recursive; in familiar theories of arithmetic that we have previously here quoted, the Ax_T it is. The complexity of the provability predicate $\mathsf{Prf}_\mathsf{T}(x, y)$ depends essentially on that of the formula Ax_T. To fix the ideas, let $\mathsf{T} = \mathsf{PA}$, then $Ax_\mathsf{PA}(y)$, a standard definition of proper axioms has the form $a_0 \vee ... \vee a_7 \vee Ind(y)$, where $a_0, ..., a_7$ are the codes of the eight axioms of *Peano Arithmetic* different from the induction, and $Ind(y)$ is true iff y codes an instance of the induction principle. To formalize the predicate $Ind(y)$ the idea is to say: "there are $x, z < y$ such that $Form(x)$ and $\exists i < z(z = \langle 1, i \rangle)$ and if $Subst(x, i, \ulcorner \overline{0} \urcorner)$ and for all v, if $Subst(x, i, \ulcorner \overline{v} \urcorner)$, then $Subst(x, i, \ulcorner \overline{S(v)} \urcorner)$, then for all v, $Subst(x, i, \ulcorner \overline{v} \urcorner)$". The provability predicate $\mathsf{Prf}_\mathsf{T}(x, y)$, "$x$ codes a proof in T of y", is now defined as:

$$\begin{gathered} Seq(x) \wedge ((x)_{ln(x)} = y) \wedge \\ \wedge \forall i \leq lh(x)(Logax((x)_i) \vee Equax((x)_i) \vee Ax_\mathsf{T}((x)_i) \vee \\ \vee \exists h, j < i MP((x)_h, (x)_j, (x)_i) \vee \exists j < i Gen((x)_j, (x)_i)) \end{gathered} \tag{3}$$

where, according to our coding of sequences, $ln(x)$ denotes the length of the sequence x and is defined as $(x)_0$. In what follows, we will denote by Prf_T the formula in the language of Peano Arithmetic that represents the primitive recursive predicate of provability Prf_T.

Remark 4. *We emphasise once again that we have defined the predicate of provability for Hilbert-style axiomatic systems, where proofs are sequences of formulas, solely for the sake of simplicity of exposition. A similar definition is possible for Gentzen-style systems, where derivations are labeled trees, although it is more laborious (see for instance Van Dalen (2013) 248-51 for Natural Deduction, or Buss (1986) 130-34 and Girard (1987) for Sequent Calculus).*

> we can readily see that the proof just given is constructive; that is...proved in an intuitionistically unobjectionable manner... (K. Gödel, 1931)

That the 1931 proof of the first theorem is *syntactic* and *constructive*. This means that it doesn't appeal to truth and that we concretely *give examples* of sentences that are independent (neither provable nor refutable). We start with the central result from which we derive the first Gödel theorem.

Theorem 70. (Fixed point theorem) *Let* T *a consistent extension of* Q *in the same language; then for all formulas* $\psi(x)$*, exists another formula* ϕ*, such that:*

$$\mathsf{T} \vdash \phi \leftrightarrow \psi(\overline{\ulcorner \phi \urcorner})$$

We say that ϕ *costitutes a* fixed point *of* ψ*.*

Proof. Let $Num(x)$ the (easily proved promitive recursive) function whose effect is to take a number n and return Gödel's number of the numeral of that number $\ulcorner \overline{n} \urcorner$ and let therefore $Sub(\ulcorner \phi(x_i) \urcorner, n) = Subst(\ulcorner \phi(x_i) \urcorner, i, Num(n)) = \ulcorner \phi(\overline{n}) \urcorner)$. First we remind the reader that according to what we said around definability and representability of functions if f is recursive we have:

$$\mathsf{Q} \vdash \alpha_f(\overline{m}, y) \leftrightarrow (y = \overline{f(m)})$$

In particular, if $f(x) = Sub(x, x)$, then exists a formula $\alpha(x, y)$ such that $\mathsf{T} \vdash \alpha(\overline{n}, v) \leftrightarrow (v = \overline{Sub(n, n)})$. Let therefore $\chi(x)$ the formula $\exists v(\alpha(x, v) \wedge \psi(v))$; then suppose that $\ulcorner \chi(x) \urcorner = m$ and lastly let ϕ the formula $\chi(\overline{m})$, i.e.the formula χ "applied" to itself.

We see that the following holds in T:

$$\begin{aligned} \phi &\longleftrightarrow \chi(\overline{m}) \\ &\longleftrightarrow \exists v(\alpha(\overline{m}, v) \wedge \psi(v)) \\ &\longleftrightarrow \exists v(v = \overline{Sub(\ulcorner \chi(x) \urcorner, \ulcorner \chi(x) \urcorner)} \wedge \psi(v)) \\ &\longleftrightarrow \exists v(v = \overline{\ulcorner \phi \urcorner} \wedge \psi(v)) \\ &\longleftrightarrow \psi(\overline{\ulcorner \phi \urcorner}) \end{aligned} \tag{4}$$

QED

A first remarkable application of this result, is the Tarski theorem

Theorem 71. (Tarski's undefinability theorem 1933) *A sufficiently strong consistent theory can not express its truth.*

Proof. To make the proof more accessible, we give a semantic version and we show that there is no definition of $Th(\mathbb{N})$ in the standard model. So we consider a semantic version of the fixed point theorem, for which we shall have $\mathbb{N} \models \phi$ iff $\mathbb{N} \models \psi(\overline{\ulcorner \phi \urcorner})$. Suppose by contradiction that $Th(\mathbb{N})$ is definable, i.e. there exist a "definition of truth", namely a a formula $True(x)$ such that $\mathbb{N} \models \phi$ iff $\ulcorner \phi \urcorner \in Th(\mathbb{N})$ iff $\mathbb{N} \models True(\overline{\ulcorner \phi \urcorner})$. Take a fixed point of $\neg True(x)$, let for instance τ this fixed point; then we have: $\mathbb{N} \models \neg True(\overline{\ulcorner \tau \urcorner})$ iff $\mathbb{N} \models \tau$ iff $\mathbb{N} \models True(\overline{\ulcorner \tau \urcorner})$, a contradiction. QED

A language that contains its own truth predicate and names for all its sentences, is called by Tarski *semantically closed* (e.g. the natural languages). The notion of truth cannot be defined in these languages. However many scientific languages are not semantically closed and can be placed in a hierarchy, where the truth for an obiect language can be define in a higher meta-language. In particular there is a *second order* truth-definition for first order sentences over $\mathbb{N}$. That is, $Th(\mathbb{N})$ is an analytical set. As remarked in the introductory chapter, there are also *partial* truth-predicates for all fixed level of arithmetical hierarchy and this can be proved already in the subteory of PA with induction restricted to Σ_1 formulas.

The first incompleteness theorem receives today several presentations which nevertheless reflect the same general structure (see Hájek and Pudlák (1993), ch.III):

1. It is required that the formal theory of arithmetic T is an extension of Q, which thus proves the *Fixed point theorem*, and also that the set of its theorems is recursively enumerable, and therefore have a Σ_1-definition $P(x)$.

2. Then we take a fixed point ν of the formula $\neg P(x)$, i.e. $\nu \longleftrightarrow \neg P(\overline{\ulcorner \nu \urcorner})$ where the fixed point ν is just Gödel's sentence, that will be proved undecidable.

3. Hence we prove that:

   (a) If T is consistent, then $T \nvdash \nu$

   (b) Under certain conditions (e.g. 1-consistency) $T \nvdash \neg\nu$.

1. In fact, Gödel did not consider a generic enumeration of theorems., but one based on a particular formula, namely the Σ_1 formula that enumerates the theorems of T based on the formula that represents the provability predicate, i.e. take $P(x) = \exists y \mathsf{Prf}_{\mathsf{T}}(y, x)$. Thus the theory T shows the existence of a fixed point ν for its *negation*:

$$\nu \longleftrightarrow \neg\exists y \mathsf{Prf}_{\mathsf{T}}(y, \overline{\ulcorner \nu \urcorner})$$

The formula $\exists y \mathsf{Prf}_{\mathsf{T}}(x, y)$ is usually abbreviated as $\mathsf{Pr}_{\mathsf{T}}(\overline{\ulcorner \nu \urcorner})$.

Note that ν says of itselfs of not being provable: hence, if it is unprovable it will be true. Note also the resemblance to the paradox of the liar and so with the Tarski theorem of undefinability of truth, where in place of provability (definable) appears the truth (undefinable). But what are these " additional conditions "? Alternatively, one of the following is usually placed:

1. $\omega-$consistency. A theory is ω-consistent, iff it is not the case that $\mathsf{T} \vdash \exists x \neg\phi(x)$ and $\mathsf{T} \vdash \phi(\overline{0}), \mathsf{T} \vdash \phi(\overline{1}), \mathsf{T} \vdash \phi(\overline{2}), ...etc.$ for all natural numbers. This is the original approach of K. Gödel's paper of 1931.

2. Σ_1-*soundness* (or 1-consistency). Kreisel observed that the assumption of ω-consistency was unnecessarily strong and can be replaced by 1-consistency, namely ω-consistency restricted to Σ_1 sentences. This property is equivalent to Σ_1-soundness, namely the property according to which if a Σ_1 sentence is provable, then it is true. Note that this property implies consistency. It is not difficult to check that, if we denote with $Th_{\Pi_1}(\mathbb{N})$ the set of true Π_1 sentences, then we have that T is Σ_1-sound, iff $T + Th_{\Pi_1}(\mathbb{N})$ is consistent.

Let's see some relations among these concepts. If a theory is $\omega-$ consistent, then is Σ_1-sound: indeed suppose that $\mathbb{N} \not\models \exists x\theta$, where $\exists x\theta \in \Sigma_1$; hence $\mathbb{N} \models \neg\exists x\theta$ and therefore $\mathbb{N} \models \neg\theta(\overline{n})$, for all numbers n. But $\theta \in \Delta_0$ and therefore we will have also $\mathsf{T} \vdash \neg\theta(\overline{n})$, for all numbers n and by $\omega-$ consistency $T \nvdash \exists x\theta$. If a theory T is ω-consistent, then is also consistent: note indeed that $\mathsf{T} \vdash x = x$ and therefore $\mathsf{T} \vdash \overline{n} = \overline{n}$ for all n; hence $T \nvdash \exists x(x \neq x)$ by ω-consistency, and therefore T is consistent (i.e. there is at least an unprovable sentence). The opposite direction does not hold and therefore ω-consistency is a notion strictly stronger than consistency. Let us consider for example $\mathsf{T} = \mathsf{PA} + \mathsf{Pr}_{\mathsf{PA}}(\overline{\ulcorner \overline{1} = \overline{0} \urcorner})$. Since $\mathsf{PA} \nvdash \neg\mathsf{Pr}_{\mathsf{PA}}(\overline{\ulcorner \overline{1} = \overline{0} \urcorner})$, we have that T is consistent. Moreover for all n, the relation $\neg\mathsf{Prf}_{\mathsf{PA}}(n, \ulcorner \overline{1} = \overline{0} \urcorner)$ is true. Hence, by binumerablility, $\mathsf{PA} \vdash \neg\mathsf{Prf}_{\mathsf{PA}}(\overline{n}, \overline{\ulcorner \overline{1} = \overline{0} \urcorner})$, from which $\mathsf{T} \vdash \neg\mathsf{Prf}_{\mathsf{PA}}(\overline{n}, \overline{\ulcorner \overline{1} = \overline{0} \urcorner})$ and at the same time $\mathsf{T} \vdash \exists y \mathsf{Prf}_{\mathsf{PA}}(y, \overline{\ulcorner \overline{1} = \overline{0} \urcorner})$.

Gödel's first incompleteness theorem. Let's see how we can get the first theorem, in a version closer to the original due to Gödel, through the concepts of ω-consistency and binumerability, and the provability predicate. Let ν a fixed point of $\neg\mathsf{Pr}_{\mathsf{T}}$, where T is an extension of Q. That is, the following is provable it:

$$\nu \leftrightarrow \neg\mathsf{Pr}_{\mathsf{T}}(\overline{\ulcorner \nu \urcorner})$$

We show that:

(a) under the hypothesis of *consistency*, ν is not provable in T, e

(b) under the hypothesis $\omega - consistency$, is not provable neither in $\neg\nu$.

Proof. Let's look at the two cases:

(a) Let us suppose that $\mathsf{T} \vdash \nu$; then a number n codes a proof of ν in T. But the primitive recursive relation "n codes a proof of ν in T", as we have seen, is binumerable in T, i.e. we have $\mathsf{T} \vdash \mathsf{Prf}_{\mathsf{T}}(\overline{n}, \overline{\ulcorner \nu \urcorner})$. Hence $\mathsf{T} \vdash \exists x \mathsf{Prf}_{\mathsf{T}}(x, \overline{\ulcorner \nu \urcorner})$, namely $\mathsf{T} \vdash \neg\nu$, from the definition of ν, against consistency.

(b) Suppose, therefore, to have demonstrated (a) in the way that we said; if T is ω-consistent, then is also consistent and therefore $\mathsf{T} \nvdash \nu$; as a consequence, for no natural number n we will have that it codes a proof of ν in T, namely, the primitive recursive relation "n codes a proof of ν" does not hold for any natural number; it follows that $\mathsf{T} \vdash \neg\mathsf{Prf}_{\mathsf{T}}(\overline{n}, \overline{\ulcorner \nu \urcorner})$, for all number n. Lastly, from ω-consistency it follows $\mathsf{T} \nvdash \exists y \mathsf{Prf}_{\mathsf{T}}(y, \overline{\ulcorner \nu \urcorner})$ and therefore $T \nvdash \neg\nu$.

QED

Rosser's version of the first incompleteness theorem. The assumption of $\omega-$consistency may seem too restrictive, in view of generalizations of the method to generic consistent theories that extend Q: we mentioned an example of a theory that is consistent, though not $\omega-$consistent. Similarly if we take the indipendent sentence ν of Gödel's theorem, being of complexity Π_1, its negation will be Σ_1; then $\mathsf{Q} + \neg\nu$ will be consistent (since ν is not derivabile from Q), but is not Σ_1-sound: indeed clearly $\mathsf{Q} + \neg\nu \vdash \neg\nu$, but $\neg\nu$ is *false* (being Gödel sentence ν true). A result shown by J.B. Rosser in 1936, allows us to use only the hypothesis of simple *consistency*, for both (3.a) that for (3.b). It makes use of the technique of the so-called *witness comparison* to produce a provability predicate rather particular. Note that the following proof uses only elementary means already available in Robinson's Arithmetic.

Theorem 72. (Rosser 1936) *Let* $\mathsf{T} \supseteq \mathsf{Q}$ *computably enumerable consistent; then exists un sentence* β*, such that (a)* $\mathsf{T} \nvdash \beta$ *e (b)* $\mathsf{T} \nvdash \neg\beta$.

Proof. Check this for Q. Let $\mathsf{Pr}^{\mathsf{R}}_{\mathsf{Q}}(\overline{\ulcorner \alpha \urcorner})$ be the following sentence:

$$\exists x(\mathsf{Prf}_{\mathsf{Q}}(x, \overline{\ulcorner \alpha \urcorner}) \wedge \forall y < x \neg \mathsf{Prf}_{\mathsf{Q}}(y, \overline{\ulcorner \neg\alpha \urcorner}))$$

and consider the fixed point $\beta \leftrightarrow \neg\mathsf{Pr}^{\mathsf{R}}_{\mathsf{Q}}(\overline{\ulcorner \beta \urcorner})$.

(a) Suppose that $\mathsf{Q} \vdash \beta$ and let a a number that witness this, coding a proof of β; from binumerability we have $\mathsf{Q} \vdash \mathsf{Prf}_{\mathsf{Q}}(\overline{a}, \overline{\ulcorner \beta \urcorner})$ and $\mathsf{Q} \vdash \neg\mathsf{Prf}_{\mathsf{Q}}(\overline{n}, \overline{\ulcorner \neg\beta \urcorner})$ for all n. Using the principle, provable in Q, according to which $\forall x \leq \overline{a}(\bigvee_{n \leq a} x = \overline{n})$ we obtain:

$$\exists x(\mathsf{Prf}_{\mathsf{Q}}(x, \overline{\ulcorner \beta \urcorner}) \wedge \forall y < x \neg \mathsf{Prf}_{\mathsf{Q}}(y, \overline{\ulcorner \neg\beta \urcorner}))$$

namely: $\mathsf{Q} \vdash \neg\beta$, against consistency.

Indeed, thanks to the identity axioms:

$$\neg\mathsf{Prf}_{\mathsf{Q}}(\overline{n}, \overline{\ulcorner \neg\beta \urcorner})) \rightarrow (x = \overline{n} \rightarrow \neg\mathsf{Prf}_{\mathsf{Q}}(x, \overline{\ulcorner \neg\beta \urcorner}))$$

and since we have $\neg\mathsf{Prf}_{\mathsf{Q}}(\overline{n}, \overline{\ulcorner \neg\beta \urcorner}))$ for all n, by "modus ponens" we get, for all n, $x = \overline{n} \rightarrow \neg\mathsf{Prf}_{\mathsf{Q}}(x, \overline{\ulcorner \neg\beta \urcorner})$.

Now we use the tautology

$$((A \rightarrow C) \wedge (B \rightarrow C) \rightarrow ((A \vee B) \rightarrow C))$$

From the conjunction $\bigwedge_{n \leq a}(x = \overline{n} \rightarrow \neg\mathsf{Prf}_{\mathsf{Q}}(x, \overline{\ulcorner \neg\beta \urcorner})$ we get therefore the formula $((\bigvee_{n \leq a} x = \overline{n}) \rightarrow \neg\mathsf{Prf}_{\mathsf{Q}}(x, \overline{\ulcorner \neg\beta \urcorner})$ and from the Q theorem $\forall x(x \leq \overline{a} \rightarrow \bigvee_{n \leq a} x = \overline{n})$ by modus ponens from these two formulas we obtain $x \leq \overline{a} \rightarrow \neg\mathsf{Prf}_{\mathsf{Q}}(x, \overline{\ulcorner \neg\beta \urcorner})$, namely $\forall x \leq \overline{a}(\neg\mathsf{Prf}_{\mathsf{Q}}(x, \overline{\ulcorner \neg\beta \urcorner})$, from which, having already $\mathsf{Prf}_{\mathsf{Q}}(\overline{a}, \overline{\ulcorner \beta \urcorner})$, it follows:

$$\exists x(\mathsf{Prf}_{\mathsf{Q}}(x, \overline{\ulcorner \beta \urcorner}) \wedge \forall y < x \neg \mathsf{Prf}_{\mathsf{Q}}(y, \overline{\ulcorner \neg\beta \urcorner}))$$

(b) Suppose that $\mathsf{Q} \vdash \neg\beta$ and let a a number that witness this, coding a proof of $\neg\beta$; it follows that no b and in particular no $b \leq a$ codes a proof of β (consistency). This relation is binumerable, hence we have:

(i) $\mathsf{Q} \vdash \mathsf{Prf}_{\mathsf{Q}}(\overline{a}, \overline{\ulcorner \neg\beta \urcorner})$

(ii) $\mathsf{Q} \vdash \neg\mathsf{Prf}_{\mathsf{Q}}(\overline{b}, \overline{\ulcorner \beta \urcorner})$, for all number $b \leq a$.

However, by using, as in the previous case, the identity axioms and the principle:

$$\forall x(x \leq \overline{a} \rightarrow \bigvee_{n \leq a} x = \overline{n})$$

from (ii) we get $\forall y(y \leq \overline{a} \rightarrow \neg\mathsf{Prf}_{\mathsf{Q}}(y, \overline{\ulcorner \beta \urcorner})$ that (again by axioms of Q) is equivalent to $\forall y(\mathsf{Prf}_{\mathsf{Q}}(y, \overline{\ulcorner \beta \urcorner}) \rightarrow \overline{a} < y)$, while from (i) follows $\forall y(\overline{a} < y \rightarrow \exists z < y \mathsf{Prf}_{\mathsf{Q}}(z, \overline{\ulcorner \neg\beta \urcorner})$. From these last two formulas obtained thereby follows:

$$\forall x(\mathsf{Prf}_{\mathsf{Q}}(x, \overline{\ulcorner \beta \urcorner}) \rightarrow \exists y < x \neg\mathsf{Prf}_{\mathsf{Q}}(y, \overline{\ulcorner \neg\beta \urcorner}))$$

namely $\mathsf{Q} \vdash \beta$ (against consistency).

QED

4.4. The limit of incompleteness

But how much mathematical information is actually *needed* for obtaining the first incompleteness theorem? Is Robinson's Q the best framework for explaining incompleteness and undecidability? Since we want to consider theories in different languages, we need a method to compare them and for this reason we introduce the notion (due to Tarski) of *relative interpretability* between theories, as a measure of their strength, so that Gödel's theorem, instead of speaking of consistent extensions of Q, can be rephrased as follows:

Each consistent axiomatizable theory S that *interprets* Q is incomplete.

Definition 33. *An interpretation of a theory* S *in a theory consists of a pair* $\langle \delta(x), \tau \rangle$*:*

1. *The formula* $\delta(x)$ *of the language of* T *is called domain of the intepretation (the objects of* S *from the point of view of* T*) and* τ *is a computable function from the language of* S *to the language of* T*.*
2. *There are definitions in* T *of all symbols of* S*. The translation* τ *sends each* $n-ary$ *relational symbol* R *in a formula* ψ_R *with the same ariety; maps each* $n-ary$ *functional symbol* f *in a formula* ψ_f *of ariety* $n-1$ *and each constats (i.e.* $0-ary$ *function) in a formula* $\psi_c(y)$*.*
3. *The translation extends to terms and atomic formulas. Let us denote* $(t^{\tau,w}) = w$ *is the value of* t *according to* τ*:*
   (a) *variables:* $(x^{\tau,w}) = (w = x)$
   (b) *Constants:* $(c^{\tau,w}) = \psi_c(w)$
   (c) *Functions.* $(f(t_0, ..., t_n)^{\tau,w}) = \exists w_0, ..., \exists w_n(\bigwedge_i \delta(w_i) \wedge \bigwedge_i (t_i^{\tau,w_i}) \wedge \psi_f(w_0, ..., w_n, w)$
   (d) *Atomic formulas.* $R(t_0, ..., t_n)^{\tau} = \exists v_0, ..., \exists v_n(\bigwedge_i \delta(v_i) \wedge \bigwedge_i (t_i^{\tau,v_i}) \wedge \psi_R(v_0, ..., v_n))$
4. *This translation* τ *commutes with connectives and relativises to* $\delta(x)$ *the quantifiers:*
   (a) $(\forall x\theta)^{\tau} = \forall x(\delta(x) \rightarrow \theta^{\tau})$
   (b) $(\exists x\theta)^{\tau} = \exists x(\delta(x) \wedge \theta^{\tau})$
5. S *is relatively intepretable in* T *if there exists such a pair* $\langle \delta(x), \tau \rangle$ *and:*

(a) $\mathsf{T} \vdash \exists x \delta(x)$

(b) If f is a symbol of s, then T proves:

$$\bigwedge_i \delta(x_i) \to \exists! y(\delta(y) \wedge \psi_f(x_0, ..., x_n, y))$$

(c) In particular, for constants c, T proves $\exists! y(\delta(y) \wedge \psi_c(y))$

(d) It is asked that the translations of all axioms of S are theorems of T.

(e) The interpretation is transitive and reflexive.

Theorem 73. *The following hold:*

1. *If the theory S is intepretable in a theory T, then the consistency of T implies the consistency of S.*
2. *Moreover, if S is essentially undecidable, then T is essentially undecidable too.*

Proof. (See e.g. Murawski (1999), pp. 250-260) QED

The notion of interpretability can therefore be used as a means to measure strength of axiomatic theories: if T and S are mutually interpretable, then we can think that they represent the same expressive and deductive strength. A theory has a minimal degree of interpretation if no theory is strictly intepretable in it. It is an open question whether a theory S exists that is *minimal* in this sense, i.e. no theory is stricty interpretable in it, and Gödel result holds. We know that the other Robinson's theory, named R, is essentially incomplete and essentially undecidable too: can we find a theory S strictly intepretable in it and such that these results hold? Actually there are several of these theories: Jerabék, for example, has found an essentially undecidable and essentially incomplete theory that is unable to interpret R.

Considering theories formulated in different languages, for instance, the following set theory with only these two axioms:

1. $\exists x \forall y \neg(y \in x)$
2. $\forall x \forall y \exists z \forall v(v \in z \leftrightarrow (v = x \vee v = y))$

denoted AST (Adjunctive Set Theory), has the same strength as Q, since the two theories are mutually interpretable (see Montagna and Mancini (1994) for an in-depth study of extremely weak set theories and Švejdar (2007) for the mutual interpretability between Q and the theory of string concatenation). About the somewhat neglected "little sister"R we can actually say more:

Theorem 74. (Visser 2009) T *is locally finitely satisfable (i.e. any finite subtheory has a finite model) iff it is intepretable in* R.

Proof. (see Visser (2009) and here on p.193.) QED

Intepretability logic was deeply investigated in a multi-modal logics framework, with an additional binary modality $\phi \rhd_{\mathsf{T}} \psi$, whose intended meaning is that there is a relative interpretation of $\mathsf{T} + \psi$ in $\mathsf{T} + \phi$. It was the subject of intense studies, started partly by the Dutch school De Jongh and Veltman (1990), Visser (1990), partly by the Italian school Montagna (1987), Berarducci (1990), partly by the Czech school Švejdar (1983) and Russian school Shavrukov (1988).

4.5. A complete and decidable theory

We have already mentioned in the introduction some historical examples of decidable theories, such as the theory of real ordered fields. The example we are considering here instead concerns a theory closer to those analyzed in this chapter, namely *the first-order theory of the addition* of

the natural numbers. This result was proven independently by M. Presburger in 1929 and Thoralf Skolem in 1930 by the method of quantifier-elimination, first introduced in 1919 by Skolem in proving the completeness and decidability of the fIrst-order theory of a special class of boolean algebras. The full details were first published by Paul Bernays in *Grundlagen der Mathematik I* in 1934. Since the non-negative integers are definable in $\langle \mathbb{Z}, +, -, <, 0, 1\rangle$, these decision procedure covered both structures, $\langle \mathbb{Z}, +, -, <, 0, 1\rangle$ and $\langle \mathbb{N}, +, <, 0, 1\rangle$. Skolem and Presburger gave a quantifier elimination algorithm for these theories, but we show here the more efficient algorithm due to Cooper (1970). We actually show the decidability of $Th(\langle \mathbb{Z}, +, -, <, 0, 1\rangle)$, sometimes identified with Presburger's arithmetic (indeed, it has the same expressiveness, since formulas of these two theories are translable each other). Decidability follows from the fact that, given a formula, we can find an equivalent formula quantifier free, and the truth or falsity of such a formula is only a matter of computation. Actually the above theory *does not* satisfy the quantifier elimination, since for instance, the formula $\exists x(y = x + x)$ ("y is even") has no quantifier free equivalent. Indeed, it is provable that for all quantifier free formula $\phi(x)$ in the language of this structure, there is a number n such that for all $m > n$, either $\phi(m)$ or $\neg\phi(m)$ is true in this structure. However, decidability is proved by producing a quantifier elimination algorithm for an *extension* of this theory, obtained by adding a divisibility predicate $k|x$ defined as $\forall x(k|x \leftrightarrow \exists y(x = ky))$. Hence we actually are interested in the structure:

$$\langle \mathbb{Z}, +, -, <, 0, 1, \mathbb{Z}, 2\mathbb{Z}, 3\mathbb{Z}...\rangle$$

The theory of this structure *does* admits the quantifiers elimination and the basic formulas of this language are decidable. However, the theory of the original structure is a subset of the theory of this structure, hence in turn it is itselfs decidable.

We now show the Cooper's method for eliminating quantifiers from this theory, but before we recall this result.

Lemma 27. *Suppose that for every formula of the language of a theory* T *of the form:*

$$\exists x(\alpha_0 \wedge ... \wedge \alpha_n)$$

where each α_i *is atomic or negated atomic, there is a quantifier-free formula provably in* T *equivalent to it: then* T *admits elimination of quantifiers.*

Proof. See Enderton (2001), ch.3.2 and Harrison (2009) 328-49. Thanks to this result we will be concerned only about existential formulas. QED

The algorithm consists of the following steps:

1. Reprocessing $\exists x\phi(x)$, where $\phi(x)$ is quantifier-free:

   (a) by using boolean equivalents and definitions, we reduce the formula to the language $\neg, \wedge, \vee, <, =$.

   (b) replace $\neg(s < t)$ with $t < s + 1$.

   (c) put 0 to the left, so $s = t$ becomes $0 = t - s$ and $s < t$ becomes $0 < t - s$.

   (d) terms are written in canonical form $c_0x_0 + ... + c_nx_n + k$, with the c_i and k integers.

   Hence literals are of the forms

   $$0 = t, \neg(0 = t), 0 < t, k|t, \neg(k|t)$$

   and t is canonical. Atoms containing x are of the form $cx + s$,

2. Compute the least common multiple δ of all coefficients of x in $\phi(x)$ and then replace terms as in the following table, where x has always $\pm\delta$ as coefficients (where $hh' = \delta$):

literal	*is replaced by*
$0 < t - hx$	$0 < h't - \delta x$
$0 < hx - t$	$0 < \delta x - h't$
$k \mid hx + t$	$h'k \mid \delta x + h't$
$\neg(k \mid hx + t)$	$\neg(h'x \mid \delta x + h't)$

(5)

3. Obtain an equivalent formula of the form $\exists x(\psi(x) \wedge \delta|x)$ where $\psi(x)$ is $\phi[x/\delta x]$, i.e. we have replaced δx with x. Let us call $\phi^*(x)$ the formula $\psi(x) \wedge \delta|x$.

We have now the following alternatives:

1. either $\forall y \exists x < y \phi^*(x)$,
2. or $\exists x \phi^*(x) \wedge \forall y < x \neg\phi^*(x)$

Hence replace $0 = t$ and $0 < t$ with $\bot$, if x occours in t, and replace $0 < t$ with $\top$, if $-x$ occours in t (do nothing on other atoms). Call $\phi^*_\infty(x)$ the result of this replacement.

Lemma 28. *For sufficiently small x the formulas $\phi^*(x)$ and $\phi^*_\infty(x)$ are equivalent. Formally:*

$$\exists y \forall x < y(\phi^*(x) \leftrightarrow \phi^*_\infty(x))$$

Proof. Let us consider the atomic cases where $\phi^*(x)$ is $0 = x + a$ $0 < x + a$. In these cases $\phi^*_\infty(x)$ is $\bot$ and it is true that:

$$\forall x < -a(\phi^*(x) \leftrightarrow \bot)$$

because $x \geq -a$. Hence $y = -a$. In the case $0 < -x + a$, we have that $\phi^*_\infty(x)$ is $\top$ and clearly we have:

$$\forall x < a(\phi^*(x) \leftrightarrow \top)$$

so that $y = a$. As for non atomic cases, let us consider for example the conjunction: if $\phi^*(x)$ is $\beta(x) \wedge \gamma(x)$ and by (IH) we have $\forall x < a(\beta(x) \leftrightarrow \beta_\infty(x))$ and $\forall x < b(\gamma(x) \leftrightarrow \gamma_\infty(x))$. The result follows taking $y = min\{a, b\}$. Other case for exercise. QED

Lemma 29. *$\forall y \exists x < y \phi^*(x)$ is equivalent to $\exists x \phi^*_\infty(x)$.*

Proof. $\Rightarrow$ if $\forall y \exists x < y \phi^*(x)$ holds, then $\phi^*(x)$ holds for arbitrary small values of x, sufficent to make $\phi^*(x)$ and $\phi^*_\infty(x)$ equivalent, according to the previous lemma.

$\Leftarrow$ If $\exists x \phi^*_\infty(x)$ holds, this means that $\phi^*_\infty(n)$ holds for some n. We now show that indeed it holds for infinitely many $x < n$. Note that in a true statement $k| \pm x + t$, the formula remains true if x is altered by a multiple of k. Take the LCM δ of all k occurring in formulas of the kind $k|s$ in $\phi^*_\infty(x)$, observing that x occurs only in these formulas. Hence subtract to x a multiple fo δ. The truth value of $\phi^*_\infty(x)$ does not change and therefore $\phi^*_\infty(n) \leftrightarrow \phi^*_\infty(n - z\delta)$ for all $z \geq 0$. From this follows $\forall y \exists x < y \phi^*(x)$, by applying the previous lemma. QED

Corollary 16. *$\forall y \exists x < y \phi^*(x)$ is equivalent to $\bigvee_{i=1}^{\delta} \phi^*_\infty(i)$.*

Proof. Since $\forall y \exists x < y \phi^*(x)$ is equivalent to $\exists x \phi^*_\infty(x)$ and $\phi^*_\infty(x)$ is invariant modulo δ (i.e. changing x to $x \pm h\delta$), then $\phi^*_\infty(n)$ holds for some n iff it holds for at least one $i \in [1, \delta]$ (or any other set of δ consecutive integers). This depends on the fact that any n is congruent with some $i \in [1, \delta]$ modulo δ, and therefore $n = k\delta \pm i$.

This as regards the alternative 1. Now let us consider the alternative 2. Let us consider again δ as above (the LCM of all k in formulas $k|t$ of $\phi^*_\infty(x)$). Let i a number such that $\phi^*(i)$ holds, but $\phi^*(i - \delta)$ does not hold. By the invariance of divisibility predicate under δ the change of values of ϕ^* is due to other literals containing x, becoming false when x decreases. QED

We come to the following definition:

Definition 34. *Let $L(x)$ be a literal of ϕ^* containing x , but different from a divisibility predicate. A boundary point for it is given by the following table:*

literal	*boundary point*
$0 = x + t$	$-(t + 1)$
$\neg(0 = x + t)$	$-t$
$0 < x + t$	$-t$
$0 < -x + t$	*none*

(6)

If b is a boundary point for a literal, this literal is false for $x = b$, but true for $x = b + 1$.

Remarks. If there is a minimum i s.t. $\phi^*(i)$, then $\phi^*(i - \delta)$ does not hold.

Theorem 75. *Let δ the LCM as above and let B the boundary set. For all integers i, if $\phi^*(i)$ holds, but $\phi^*(i - \delta)$ does not hold, then there is $b \in B$ such that $1 \leq j \leq \delta$ and $i = b + j$.*

Proof. We check only the base cases, leaving the induction relative to the complex formulas as exercise.

1. If ϕ^* is $(0 = x + t)$, then $i = (-t)$. The boundary point is $-(t + 1)$. Hence there exists $b \in B$ such that $i = b + 1$.
2. ϕ^* is $\neg(0 = x + t)$, the boundary point is $-t$; note that $\phi^*(i - \delta)$ is false only if $i = \delta + b = \delta - t$.
3. If ϕ^* is $0 < x + t$ a boundary point is again $-t$; by assumption $-t + 1 \leq i$ and $i \leq -t + \delta$. Ergo $i = b + j = -t + j$ for some $1 \leq j \leq \delta$.
4. For other kind of literals, it is not possible that $\phi^*(i)$ holds, but $\phi^*(i - \delta)$ does not hold.

Now we compose the alternatives 1. and 2. QED

Corollary 17. *Let $\phi^*(x)$ as above, where all coefficients of x are ± 1. Let δ be the LCM of all k such that $k|t$, for some t containing x occurs in the formula. Then $\exists x \phi^*(x)$ is equivalent to:*

$$\bigvee_{i=1}^{\delta} (\phi^*_\infty(i) \vee \bigvee_{b \in B} \phi^*_\infty(b + i))$$

Proof. By collecting the above results concerning alternatives 1. and 2. $\Rightarrow$ Suppose $\exists \phi^*(x)$ holds. Hence either:

1. $\forall y \exists x < y \phi^*(x)$ or
2. $\exists x \phi^*(x) \wedge \forall y < x \neg \phi^*(y)$

In case (1) we obtain $\bigvee_{i=1}^{\delta} \phi_\infty(i)$ by the previous result. In case (2) there is $i = b + j$ such that $\phi^*(i)$ and $\neg\phi^*(i-\delta)$ and therefore $\bigvee_{j=1}^{\delta} \bigvee_{b\in B} \phi^*(b+j)$.

$\Leftarrow$ If the disjunction holds, then, if $\bigvee_{i=1}^{\delta} \phi_\infty(i)$ then by previous results we have an arbitrarily large negative x with $\phi^*(x)$ and therefore $\exists x\phi^*(x)$. If on the other hand $\bigvee_{j=1}^{\delta} \bigvee_{b\in B} \phi^*(b+j)$ holds, so does $\exists x\phi^*(x)$. QED

Corollary 18. *$Th(\langle \mathbb{Z}, +, -, <, 0, 1, \mathbb{Z}, 2\mathbb{Z}, 3\mathbb{Z}...\rangle)$ is decidable, i.e., there is an effective procedure which, given any sentence of the language of this theory, will decide the truth or falsity of it in $\langle \mathbb{Z}, +, -, <, 0, 1, \mathbb{Z}, 2\mathbb{Z}, 3\mathbb{Z}...\rangle$.*

4.6. Another look at incompleteness: Tennenbaum's theorem

Another kind of "undecidability" theorem is due to Tennenbaum (1959): although the first-order theories of arithmetic are not categorical (see on p.15), it is impossible to build non-standard models of Peano Arithmetic in which the addition and multiplication operations, as well as the order relation are effectively computable. The phenomenon highlighted by Tennenbaum is actually very pervasive and applies to most subtheories of PA. Already Skolem (1955) built nonstandard models of arithmetic by an ultrapower-like construction and had raised the question of recursiveness of nonstandard models. Tennenbaum's theorem improved the result in Mostowski (1957), that no nonstandard model of primitive recursive arithmetic with predicates for all primitive recursive functions can be recursive. On the contrary Shepherdson (1964), using algebraic methods (a model of open induction is an integer part of a real closed field), produced a recursive nonstandard models of arithmetic with axiom schema of induction for quantifier-free formulas (see Kaye (2011) for a more detailed historical background and Berarducci and Otero (1996) and D'Aquino (1997) for further developments).

Definition 35. *Let $\mathcal{U}$ and $\mathcal{B}$ be first-order structures of the same language. A function $f : \mathcal{U} \longrightarrow \mathcal{B}$ is an* homomorphism *iff:*

1. *For all $a_0, ..., a_n \in A$, and functional symbol F, if $F^{\mathcal{U}} = h$ and $F^{\mathcal{B}} = g$, si ha:*
$$f(h(a_0, ..., a_n)) = g(f(a_0), ..., f(a_n))$$
2. *For all $a_0, ..., a_n \in A$, and relational symbol P, if $P^{\mathcal{U}} = R$ and $P^{\mathcal{B}} = \widetilde{R}$, we have:*
$$\langle a_0, ..., a_n\rangle \in R \Rightarrow \langle f(a_0), ..., f(a_n)\rangle \in \widetilde{R}$$
3. *For all constant a_i, if $a_i^{\mathcal{U}} = c_i$ and $a_i^{\mathcal{B}} = b_i$, then $f(c_i) = d_i$.*
If also the inverse of 2. holds, this is a strong *homomorphism. Also we say that it is an* embedding*, iff*
4. *f is injective: $x \neq y \Rightarrow f(x) \neq f(y)$,*
5. *f is a strong homomorphism.*
6. *It is an elementary embedding iff for all formulas ϕ and all $a_0, ..., a_m \in \mathcal{U}$,*
$$\mathcal{U} \models \phi(a_0, ..., a_m) \text{ if and only if } \mathcal{B} \models \phi(j(a_0), ..., j(a_m))$$

We say that f is an isomorphism *(notation $\mathcal{U} \cong_f \mathcal{B}$), iff it is an embedding and it is bijective.*

Related to the algebraic notion of isomorphism, we have the logical notion of *elementary equivalence*: the structures $\mathcal{U}, \mathcal{B}$ are elementarily equivalent (notation $\mathcal{U} \equiv \mathcal{B}$) iff for all *sentence* ϕ, $\mathcal{U} \models \phi \Leftrightarrow \mathcal{B} \models \phi$. Isomorphic structures are also elementarily equivalent, but the reverse implication is not true and we have that:

$$\mathcal{U} \equiv \mathcal{B} \not\Rightarrow \mathcal{U} \cong \mathcal{B}$$

Indeed it is provable that $\langle \mathbb{Q}, <\rangle \equiv \langle \mathbb{R}, <\rangle$, however (Cantor!) $\langle \mathbb{Q}, <\rangle \ncong \langle \mathbb{R}, <\rangle$. A structure which is elementarily equivalent, but not isomorphic, to the standard model is called a *nonstandard model* of arithmetic.

Definition 36. *A model $\mathcal{M}$ for the language of Peano Arithmetic is standard, if it isomorphic to the intended model $\langle \mathbb{N}, +, \cdot, S, 0\rangle$. Otherwise is called non-standard. Let $t^{\mathcal{M}}$ be the element of the model that interprets a term t. The elements x of $\mathcal{M}$ such that $x = \overline{n}^{\mathcal{M}}$ for some number n (i.e. the elements that interpret a natural number) are called standard.*

Theorem 76. *If $\mathcal{M}$ contains a nonstandard element, then it is nonstandard (i.e. non isomorphic to the standard model).*

Proof. Suppose by contradiction that $\mathcal{M}$ is standard but contains a nonstandard element x. Let $g : \mathbb{N} \to \mathcal{M}$ be the alleged isomorphism. Hence for all natural numbers n, $g(\overline{n}^{\mathbb{N}}) = \overline{n}^{\mathcal{M}}$, namely standard elements are mapped in standard elements. But $\mathcal{M}$ contains a non standard element x, and therefore g cannot be suriective, against the hypothesis that it is an isomorphism. QED

We can derive the existence of non-standard models from the compactness. For example, let $\mathcal{L}_{\mathsf{PA}}$ be the language of Peano arithmetic and let us consider the extended language $\mathcal{L}_{\mathsf{PA}}{}^* = \mathcal{L}_{\mathsf{PA}} \cup \{c\}$ and the theory:

$$T^* = \mathsf{PA} \cup \{c > \overline{n} | n \in \mathbb{N}\}$$

where:

$$\overline{n} = \overbrace{S(S(...S(\overline{0})...))}^{n-times}$$

Note that each finite subset of this theory is contained in some:

$$T_k := \mathsf{PA} \cup \{c > \overline{n} | n < k\}$$

But such a T_k is true in every structure $\mathcal{N}_k = \langle \mathbb{N}, k\rangle$, where $c^{\mathcal{N}} = k$. Hence every finite subset of T^* has a model and therefore also T^* has a model $\mathcal{M}$.

However in this model, we have $c^{\mathcal{M}} > k$ for all natural k; it will contain thus at least one non-standard element. In particular this means that the reduct of this model to our original language of PA is not isomorphic to the standard model. Applying Löwenheim-Skolem, we can take the model in question, at most countable. Rather, $\mathbb{N}$ constitutes an initial segment of each non-standard model; non-standard elements come "after", since both the standard model and non-standard, fulfils the following principle, for all $k \in \mathbb{N}$:

$$\forall x(x < \overline{k} \to (x = \overline{0} \vee x = \overline{1} \vee ... \vee x = \overline{k-1}))$$

There are exactly $2^{\aleph_0}$ non isomorphic countable models of arithmetics.

Theorem 77. *There are exactly $2^{\aleph_0}$ non isomorphic countable models of Peano arithmetic* PA.

Proof. Clearly there are *at most* (since there are continuum-many interpretations of $+$, continuum-many interpretations of $\times$ etc.). We can verify that there are *at least*, considering that if ν is Gödel's undecidable sentence relative to PA, then $S_0 = PA + \nu$ and $S_1 = PA + \neg\nu$ they are both consistent and they will therefore have a model; moreover, they will also have respectively a undecidable Gödel's sentence ν_0 and ν_1. Hence we can consider $S_{00} = S_0 + \nu_0$, $S_{01} = S_0 + \neg\nu_0$, $S_{10} = S_1 + \nu_1$, $S_{11} = S_1 + \neg\nu_1$ etc. in general,for all finite binary string τ, $S_{\tau 0} = S_\tau + \nu_\tau$ and $S_{\tau 1} = S_\tau + \neg\nu_\tau$. Recall that $|\{0,1\}|^{\mathbb{N}} = 2^{\aleph_0}$. Each branch of the binary tree corresponds to a completion of PA. Being all these theories consistent and inequivalent, each one has a model, which is not isomorphic to any model of the others. In addition these will be all models of PA. QED

In spite of the apparent chaotic nature of the non-standard part of nonstandard models, it actually has a very precise order type.

Definition 37. *If $\langle P, <_P\rangle$ and $\langle Q, <_Q\rangle$ are linear orders, let:*

1. *$P + Q = P \times \{0\} \cup Q \times \{1\}$ the sum of them, where $\langle a, i\rangle <_{P+Q} \langle b, j\rangle$ if and only if $j = i = 0$ and $a <_P b$, or $j = i = 1$ and $a <_Q b$, or $i = 0$ e $j = 1$ (i.e. the elements of P and* Q *maintain their original order, but the elements in P precede those in* Q*).*

2. *let* $P \times Q$ *be the product of them, and order it lexicographically* $\langle a, b\rangle <_{P\times Q} \langle c, d\rangle$ *iff either* $a <_P c$ *or* $a = c$ *and* $b <_Q d$.

Theorem 78. (Henkin 1950) *Each non-standard countable model* $\mathcal{U}$ *of Peano arithmetic has order type* $\mathbb{N} + \mathbb{Q} \times \mathbb{Z}$.

Proof. Now we observe that $\mathcal{U}$ has $\mathbb{N}$ as initial segment and then its order will be of the form $\mathbb{N} + A$, for some linear order A.

Let $a, b \in \mathcal{U}\backslash\mathbb{N}$; we define the following equivalence $a \sim b \Leftrightarrow |a - b| \in \mathbb{N}$, namely, a, b are in this relation, iff the absolute value of the difference is standard. Moreover, let $a \prec b$ iff $a < b$ and $|a - b| \notin \mathbb{N}$ ("b is much bigger than a"). Lastly:

$$[a] \prec^* [b] \Leftrightarrow a \prec b$$

Observe that a is non-standard, then $[a]_\sim = \{a - n, a + n | n \in \mathbb{N}\}$, namely, the equivalence class of a has the form of $...a - 2, a - 1, a, a + 1, a + 2...$ Note that this order is isomorphic to that of integers $\mathbb{Z}$. Also notes that the following holds:

1. $\prec$ is transitive and antireflexive.
2. $a \prec b, c \sim a, d \sim b \Rightarrow c \prec d$
3. $a \prec b \Rightarrow \neg(b \prec a)$

Lastly, note that $\prec^*$ lacks a maximum, lacks a minimum, and is also a dense order, since the following hold:

1. $[a] \prec^* [a + a]$ (no last element)
2. $[a/2] \prec^* [a]$ *or* $[a + 1/2] \prec^* [a]$ (no first element)
3. $[a] \prec^* [b] \Rightarrow [a] \prec^* [a + b/2] \prec^* [b]$ (density)

But this order is isomorphic to that of rational $\mathbb{Q}$ (remember that for a result of Cantor, *all dense countable linear orders without endpoints are isomorphic*); hence, consider that $...\xi_{q_0}, \xi_{q_1}, \xi_{q_2}...$ are the equivalence classes with respect to $\sim$, ordered by $\prec^*$, that is isomorphic to the order of $\mathbb{Q}$, while inside them, they are ordered as $\mathbb{Z}$. Hence the equivalence classes $\xi_{q_0} \prec^* \xi_{q_1} \prec^* \xi_{q_2}...$ are ordered like $\mathbb{Q}$ and each one has the order of $\mathbb{Z}$:

$$\xi_q = ...a_{q_{z-2}} < a_{q_{z-1}} < a_{q_z} < a_{q_{z+1}} < a_{q_{z+2}}...$$

then the elements a_{q_z} of the non-standard part of $\mathcal{U}$ can be ordered in this way:

$$a_{q_z} <_U a_{q'_{z'}} \Leftrightarrow q <_{\mathbb{Q}} q' \vee (q =_{\mathbb{Q}} q' \wedge z <_{\mathbb{Z}} z')$$

This, according to the above definition of product between orders, is the order type of $\mathbb{Q} \times \mathbb{Z}$. So the whole model has the order type $\mathbb{N} + \mathbb{Q} \times \mathbb{Z}$. QED

For a remark of Potthoff (1969), in the previous theorem, we will see now, we cannot replace $\mathbb{Q}$ with $\mathbb{R}$, for any non-standard model. To see this, a key result of this field is needed. First observe that the standard part of a non standard model M, is not definable in M. Suppose by contradiction that it is, i.e. that for some $\sigma(x)$, $\mathcal{M} \models \sigma[a]$ (i.e. a satisfies σ in the model $\mathcal{M}$) iff $a \in \mathbb{N}$, and therefore $\neg\sigma(x)$ will define the non standard part $\mathcal{M} \setminus \mathbb{N}$: but then there will be a minimum (the principle of the minimum number is equivalent to induction) non-standard, that is contradiction, as we saw about the structure of non-standard models.

Hence the following obtain.

Theorem 79. (Overspill) *If* $\phi[n]$ *holds for all* $n \in \mathbb{N}$, *then exists* $a \in M \setminus \mathbb{N}$, *such that* $\phi[b]$ *holds for all* $b \leq a$.

Proof. Indeed, if no a non standard satisfied $\phi[a]$, then the formula $\sigma(x)$ defined as $\exists y(x < y \wedge \phi(y))$ would define the standard part, against what we said. In other words, if $\mathcal{M} \models \sigma[a] \Leftrightarrow a \in \mathbb{N}$, then in particular, since $0 \in \mathbb{N}$ and if $n \in \mathbb{N}$, also $n+1 \in \mathbb{N}$, we would have $\mathcal{M} \models \sigma(\overline{0}) \wedge \forall x(\sigma(x) \rightarrow \sigma(x+1))$. However, being $\mathcal{M}$ a model of PA it must satisfy the induction principle, from which follows $\mathcal{M} \models \sigma(\overline{a})$, for all a, included non standard elements. QED

An interesting application (see Bovykin and Kaye (2002)) of the *overspill* theorem is the proof of the above-mentioned fact that in the order type $\mathbb{N} + \mathbb{Q} \times \mathbb{Z}$, is not possible to replace $\mathbb{Q}$ with $\mathbb{R}$; in other words we cannot consider the equivalence classes $[a] = \{...a-2, a-1, a, a+1, a+2...\}$ as "reals". Recall that the real numbers fulfill the so-called *Dedekind completeness property*, that can be also expressed in this form:

> Let X be a set of real numbers; an upper bound for X is a real r such that $r \geq a$, for all $a \in X$. The least upper bound property says that any non-empty set of reals having an upper bound, has a least upper bound.

Now we just apply this principle.

Let us take a non standard model $\mathcal{M}$, an element a nonstandard of the model and a sequence $\{ia\}_{i \in \mathcal{M}}$. Note that $i < j$, where i, j are standard, implies

$$[ia] \prec^* [ja]$$

Suppose there is a least upper bound $[b] = sup\{[na] | n \in \mathbb{N}\}$. Then, for all $n \in \mathbb{N}$, $na \prec b$ and therefore by overspill, there will be an element c non standard such that $ca \prec b$. Hence, for all $n \in \mathbb{N}$, $na \prec ca \prec b$. But all reals $\{[sa]\}_{\mathbb{N} < s < c}$ are between $\{[na] | n \in \mathbb{N}\}$ and $[b]$ (contradiction, against the hypothesis that $[b] = sup\{[na] | n \in \mathbb{N}\}$).

We formulate this result with respect to PA, but, as said, it appropriately extends to all its subtheories, until minimum extensions of the *open induction* theory IOpen, for which, as proved for the first time by Sheperdson, there exist recursive models. It states that in these theories, the standard model is the only one in which the operations of addition and multiplication are recursive. This can be seen as a kind of "incompleteness result", which states that we can not build a non standard model of Peano arithmetic.

Definition 38. *The "standard system" $SSy(\mathcal{M})$ of a non-standard model $\mathcal{M}$ is the set of subsets $A \subseteq \mathbb{N}$ such that $n \in A$ iff $\mathcal{M} \models \phi(a, n)$, for some $\phi(x, y)$ and some $a \in \mathcal{M}$.*

Remember that in PA is Σ_1^0-definable the primitive recursive predicate "p is the x^{th}-prime". Note that $\mathbb{N}$ and $\mathcal{M}$ agree on the standard prime numbers, in the sense that p_n, for $n \in \mathbb{N}$ is the n^{th}-prime in both models. It is not hard to show that the "standard system" can actually be equivalently defined also as the set of subsets $S \subseteq \mathbb{N}$ such that: $S = \{n \in \mathbb{N} | \mathcal{M} \models p_n | a\}$ for some a of the model.

Theorem 80. (Tennenbaum 1959) *Let $\mathcal{M} = \langle M, \oplus, \otimes, S, 0\rangle$ a non standard countable model of* PA *non isomorphic to the standard model. Then $\mathcal{M}$ is not recursive.*

Proof. We show first that $SSy(\mathcal{M})$ contains a non recursive set. Let indeed $A, B \subseteq \mathbb{N}$ disjoint recursively enumerable and recursively inseparable. Being computably enumerable they will be definable in $\mathbb{N}$ respectively by Σ_1^0 formulas $\exists y \alpha(x, y)$ and $\exists y \beta(x, y)$; but these are preserved in the extensions of $\mathbb{N}$. In particular, since A and B have empty intersection, we will have for each k:

$$\mathbb{N} \models \forall x < \overline{k} \forall y < \overline{k} \forall z < \overline{k} \neg(\alpha(x, y) \wedge \beta(x, z))$$

But by "overspill" this holds also in M with some $a \in M \setminus \mathbb{N}$.

If now we define $C = \{n \in \mathbb{N} | \mathcal{M} \models \exists y < a \alpha(n, y)\}$, then we note that $A \subseteq C$ and $C \cap B = \emptyset$. Hence, being $A, B \subseteq \mathbb{N}$ recursively inseparable, $C \in SSy(\mathcal{M})$ but it cannot be recursive. QED

Given $C \in SSy(\mathcal{M})$, that we assume not recursive, it can be coded in $\mathcal{M}$ in this way: $n \in C$ if and only if $\mathcal{M} \models \exists y(c = y \otimes p_n)$, for some $c \in M$, where p_n is the n-th prime number. We show that $\oplus$ is not recursive. It is crucial that, being b and p_n standard, we can express $(y \otimes p_n) \oplus b$ as $y \oplus y \oplus ... \oplus y \oplus 1 \oplus ... \oplus 1$. Let us consider therefore the disjunction of the following formulas:

$$(c = \overbrace{y \oplus ... \oplus y}^{p_n - times}), (c = \overbrace{y \oplus ... \oplus y}^{p_n - times} \oplus 1), ..., (c = \overbrace{y \oplus ... \oplus y}^{p_n - times} \oplus \overbrace{1 \oplus ... \oplus 1}^{p_n - 1 - times})$$

Since it is a model of PA, the non standard model $\mathcal{M}$ will verify the "Euclidean division":

$$\forall x \forall z(z \neq 0 \rightarrow \exists y \exists b(x = yz + b \wedge 0 \leq b < y))$$

Fix $z = p_n$ and $x = c$. Hence in the model there are (unique) y and $0 \leq b < p_n$ such that $c = (y \otimes p_n) \oplus b$.

Since p_n is standard, the theory proves:

$$\forall x(x < \overline{p_n} \leftrightarrow \bigvee_{k < p_n} x = \overline{k})$$

so that we have these alternatives:

1. If $b = 0$, then $c = (a \otimes p_n^M)$, hence the disjunction is true, since the first disjunct is true, and therefore $\mathcal{M} \models \exists y(c = y \otimes p_n)$, from which $n \in C$.

2. If $b = 1 \oplus 1 \oplus ... \oplus 1$, for some number of summands less than p_n. Then $n \notin C$, because is true one of the other disjucts and therefore $\mathcal{M} \not\models \exists y(c = y \otimes p_n)$.

But then, if $\oplus$ is recursive, in $\mathcal{M}$ is decidable if either $n \in C$ or $n \notin C$. Given n, compute p_n and search for the unique $y \in \mathcal{M}$ such that:

$$(\overbrace{y \oplus ... \oplus y}^{p_n - times}) \oplus \overbrace{1 \oplus 1 \oplus ... \oplus 1}^{b - times} = c$$

Indeed, our search is guaranteed to terminate, being b and y uniquely determined by n and c . Hence, at the end, if $b = 0$, then we know that $n \in C$, otherwise, we know that $n \notin C$. Contradiction, because C was not recursive. The case of $\otimes$ is treated in a similar way.

A more refined version of Tennenbaum's theorem can be expressed in this form (see Smorynski (1984)).

Theorem 81. *In a nonstandard model $\mathcal{M} = \langle \mathbb{N}, \otimes, \oplus, S, 0 \rangle$, every set X in the standard system is recursive in each $\otimes$ and $\oplus$.*

Proof. Recall that $n \in X$ iff $\mathcal{M} \models \phi(n, b)$, for some $b \in \mathcal{M}$ and formula ϕ, iff for some $b \in \mathcal{M}$, $\mathcal{M} \models p_n | b$, iff $\exists c(p_n^{\mathcal{M}} \otimes c = b)$ iff $\exists c(c \oplus ... \oplus c = b)$ (p_n-times). If we call the (by hypothesis, recursive) relation expressed in parentheses "A", then this means that X is Σ_1^A definable, in the relativized arithmetical hierarchy, and therefore X is r.e. in A (by the relativised version of a well-known result). Since the complement of X correspond to $\neg\phi$ the same argument leads to the conclusion that also the *complement* of X is r.e. in A and therefore X is recursive in A. Analogously, for multiplication, if c, b are as above, take $d = 2^c$ and $e = 2^b$; hence $d \otimes ... \otimes d = e$, because $2^c \otimes ... \otimes 2^c = 2^{c \oplus ... \oplus c}$ and the same argument apply, so that X is also recursive in $\otimes$. Tennenbaum's theorem follows: if these operations $\otimes$ and $\oplus$ were computable, also all elements of $SS(\mathcal{M})$ would be computable, because if A is computable and $X \leq_T A$, then X is computable too; but we have shown that this is false, because there exist non computable sets in the standard system. QED

4.7. Tennenbaum's and Gödel's theorems

We would now investigate the relationship between Tennenbaum's and Gödel's theorems, showing how to use Tennenbaum's result to obtain the incompleteness theorem. A proof of the completeness theorem for countable vocabularies can be based on König's lemma ("every finitely branching infinite tree has an infinite branch"). Refinements of König's lemma are possible, e.g. the so-called "Kreisel basis theorem", according to which every infinite computable binary tree has an infinite path *recursive in* $\emptyset'$, i.e.. computable from the halting problem, or, equivalently, by Post's theorem, a Δ_2 path.

These refinements yield refinements to the first-order Completeness Theorem, leading to the conclusion that that there exists a model $\mathcal{M}$ of PA where $\otimes, \oplus$ are Δ_2-definable. More precisely, let S_M be the set of pairs $\langle \ulcorner \phi \urcorner, n \rangle \in \mathbb{N} \times \mathbb{N}$ such that $\mathcal{M} \models \phi(\overline{n})$. Say that $\mathcal{M}$ *is a* Δ_2 *model* if S_M is Δ_2 definable in the standard model. In other words, a Δ_2 model, is a model $\mathcal{M}$ with domain $\mathbb{N}$ such that the set of formulas (with parameters) which are true in the model is Δ_2-definable. Then the *Arithmetized Completeness Theorem* says that each consistent recursively axiomatizable theory has a Δ_2 model $\mathcal{M}$ (see Berarducci and Mamino (2023), Kaye (1991), Kennedy (2022) pp. 153-188, Kossak and Schmerl (2006), p. 163 for further details on all the points mentioned).

Hence, by the generalized Tennenbaum theorem above and Post theorem, every set in $SS(\mathcal{M})$ is Δ_2-definable too: indeed, by Post's theorem, $A \in \Delta_2$ iff $A \leq_T \emptyset'$ and therefore, by transitivity of the reduction, if $B \leq_T A$, then also $B \in \Delta_2$. Now suppose by contradiction that the model of the arithmetized completeness theorem $\mathcal{M}$ is an *elementary extension* of the standard model (in symbols $\mathbb{N} \preceq \mathcal{M}$): this means that for every formula $\phi(x)$ and every element $b \in \mathbb{N}$, this b satisfies the formula in $\mathbb{N}$ iff it satisfies the formula in $\mathcal{M}$. Then *all* arithmetically definable sets, i.e. all Σ_n sets for every n are in $SS(\mathcal{M})$ and at the same time are computable from the $\otimes, \oplus$ of this model. Therefore these operations cannot be Δ_2-definable or arithmetical at all (contradiction). Hence $\mathbb{N} \not\preceq \mathcal{M}$ and therefore there is some sentence ψ true in $\mathbb{N}$ but false in $\mathcal{M}$ from which follows $\mathsf{PA} \nvdash \psi$.

5. Second incompleteness theorem: research developments and consequences

5.1. Intensionality of the consistency statements

In Gödel's 1931 article there is only the sketch of what is been called "second theorem", with the comment that it is a formalization of the first theorem, accompanied by the announcement of the imminent publication of a second part of the article, which is never happened. The first real proof of the " second theorem " appeared in Hilbert and Bernays (1934), vol. II, and it is in this context that emerges the decisive importance of *the way* in which the sentence that affirms the consistency of a certain theory is formalized. Hilbert and Bernays pointed out some conditions that the predicate numbering theorems must satisfy, called "derivability conditions", that were further developed by Löb. These conditions are sufficient for obtaining the second incompleteness theorem.

Gödel's second theorem, states that given a consistent extension T of Robinson's Q that meet certain conditions, there a un sentence (that we denote $Con(T)$) who claim to represent in a natural way the consistency of the theory, and which behaves as the undecidable sentence of the first theorem; i.e. we can produce, under the same conditions, a sentence undecidable which is the arithmetization of the metatheoretical sentence expressing in the language of the theory, the consistency of the theory itself:

1. If T is consistent, then $\mathsf{T} \nvdash Con(\mathsf{T})$
2. If T is $\omega-$consistent, then $\mathsf{T} \nvdash \neg Con(\mathsf{T})$

Actually, the second theorem implies that no theory for the formal arithmetic that incorporates the " finitary mathematics" is able to demonstrate the consistency of transfinite mathematics with the only means of finitary mathematics and this, as Von Neumann first observed, undermines Hilbert's programme. However, in view of the second theorem, the meaning of the provability predicate deserves a deep reflection. As we have seen, Rosser used a particular predicate. The statement of consistency $Con^R(\mathsf{T})$ that results does not fulfil the second Gödel's theorem. If we mistakenly thought we could use $\mathsf{Pr}^R(x)$, instead of the standard proof predicate, we would be faced with the following result, from which "Rosser consistency" $\neg\mathsf{Pr}^R(\overline{\ulcorner \overline{1} = \overline{0} \urcorner})$ follows by considering that $\mathsf{Q} \vdash \neg(\overline{1} = \overline{0})$.

Lemma 30. *If* $\mathsf{Q} \vdash \neg\phi$*, then* $\mathsf{Q} \vdash \neg\mathsf{Pr}^R(\overline{\ulcorner \phi \urcorner})$.

Proof. Suppose that $\neg\phi$ is provable; we claim that $\neg\mathsf{Pr}^R_{\mathsf{T}}(\overline{\ulcorner \phi \urcorner})$, namely:

$$\forall x(\neg\mathsf{Prf}_{\mathsf{Q}}(x, \overline{\ulcorner \phi \urcorner}) \vee \exists y < x \mathsf{Prf}_{\mathsf{Q}}(y, \overline{\ulcorner \neg\phi \urcorner}))$$

is provable too.

Hence work in Q and recall that the theory proves $x \leq \overline{n} \vee x > \overline{n}$. Now, if $\neg\phi$ is provable, then a number n exists that codes a proof of it and therefore by binumerability $\vdash \mathsf{Prf}_{\mathsf{Q}}(\overline{n}, \overline{\ulcorner \neg\phi \urcorner})$. This n cannot be a code of a proof of ϕ too. For any x we have two possibilities:

Duccio Pianigiani, University of Siena, Italy, duccio.pianigiani@unisi.it, 0000-0001-9441-7226

Referee List (DOI 10.36253/fup_referee_list)
FUP Best Practice in Scholarly Publishing (DOI 10.36253/fup_best_practice)

Book References DOI 10.36253/979-12-215-0778-2.references

1. If $\overline{n} < x$, then the second disjunct of the above disjunction is true.
2. $\overline{n} \geq x$ then, since $x = \overline{0} \vee ... \vee x = \overline{n}$, if any of these numbers codes a proof of ϕ we have a contradiction, again consistency. Hence $x \leq \overline{n} \rightarrow \neg \mathsf{Prf}_{\mathsf{Q}}(x, \overline{\ulcorner \phi \urcorner})$, i.e. the first disjunct holds.

We conclude that the above disjunction holds for all x. QED

On the other hand, if we maintain the usual proof-predicate, we must pay attention on the definition of proper axioms. The American logician Solomon Feferman was highly insistent on this point:

> At a given theory T we associate a class of formulas $\tau(x)$ that number the set of proper axioms of T[...] Then we can associate with each of these formulas, in a uniform way, a formula $\mathsf{Prf}_{\tau}(x, y)$ and therefore the sentence Con_{τ}. When the formula $\tau(x)$ is recognized to be a correct expression of the fact that x codes an axiom of T, then the sentence associated with it Con_{τ} will be recognized as correct expression of the proposition that T is consistent (Feferman (1960), p. 38).

The problem of how arithmetize the predicate "x is an axiom of T" does not have a single answer: different formulas defining the proper axioms correspond to different notions of provability. Some of these actually result in provability predicates of Prf_{τ} for which it is provable that $\neg \exists y \mathsf{Prf}_{\tau}(y, \overline{\ulcorner \overline{1} = \overline{0} \urcorner})$, against Gödel's result. According to Feferman, although extensionally such τ correspond to the set of axioms, actually they do not express properly belonging to them. In order to be not only correct extensionally, but also *intensionally* (from the Latin word *intentio* , which descends from medieval logic), i.e. conceptually appropriate, it is necessary that the correspondig provability predicates meet also other conditions, typically, those due to Löb. Following Feferman, many logicians actually favour an intensional approach to arithmetisation in general, which is not satisfied with the mere representability of functions and relations, but demands that the theory be able to prove general properties of them. On a similar approach is based, for example, Buss (1986), where an intensional arithmetisation of metamathematics is developed, already in the weak theory S^1_2.

Theorem 82. (Feferman (1962)) *There is a binumeration τ of the PA's axioms such that $\mathsf{PA} \vdash Con_{\tau}$.*

Proof. Since PA is recursively axiomatizable, it will have a Σ_1 binumeration σ and a Π_1 binumeration π. Let now $\tau(x)$ the conjunction of the following:

1. $\sigma(x) \wedge \forall y \leq x(\sigma(y) \leftrightarrow \pi(y))$
2. $\neg \exists z \mathsf{Prf}_{\pi \restriction x}(z, \overline{\ulcorner \overline{1} = \overline{0} \urcorner})$

where $\pi \restriction x(y) \leftrightarrow \pi(y) \wedge y \leq x$. Recalling now that PA is *reflexive*, i.e. demonstrate the consistency of all its finite subtheories, we establish the equivalence between $\tau(\overline{k})$, $\sigma(\overline{k})$ and $\pi(\overline{k})$, for all k. (*Exercise*).

Reasoning inside PA, it can be proved that:

1. if Con_{π} holds, since $\tau(x) \rightarrow \pi(x)$, we also have Con_{τ}.
2. if $\neg Con_{\pi}$ holds, take the minimum z such that $\exists y Prf_{\pi \restriction_{z+1}}(y, \overline{\ulcorner \overline{1} = \overline{0} \urcorner})$. But then $Con_{\pi \restriction z}$ and $\tau(x) \rightarrow \pi \restriction z(x)$: suppose $\neg \pi \restriction z(x)$. This means that $\pi(x) \rightarrow (x > z)$ and by the definition of z, $\exists y \mathsf{Prf}_{\pi \restriction x}(y, \overline{\ulcorner \overline{1} = \overline{0} \urcorner})$, namely $\neg \tau(x)$. But from $\tau(x) \rightarrow \pi \restriction z(x)$ follows that $Con_{\pi \restriction z} \rightarrow Con_{\tau}$.

In both cases Con_{τ}. Notice that, unlike the standard case, the formula Con_{τ} is provably Π_2 in PA, the formula τ is provably Δ_2 in PA (see Hájek and Pudlák (1993), ch.III for further details). QED

A deep investigation of Feferman's interesting, though strange proof predicate and its relations with Gödel's standard one was made in Montagna (1978), Montagna (1987) and Visser (1989). But why should we favor a formalization of the intuitive notion of consistency to another? Thinking in more general terms, in the first theorem we required only a certain formula $P(x)$ that numerate the theorems of a $\omega-$consistent extension of T of Robinson arithmetic: to it, we asked just of being *extensionally correct*, i.e. it was required the following:

$$\mathsf{T} \vdash P(\overline{n}) \text{ iff } n \text{ codes a theorem of } \mathsf{T}.$$

This is not sufficient for deriving the second theorem: there are numerations extensionally correct, but that give rise to predicates of consistency derivable in T. In *second* Gödel's theorem it is required something more: if $P(x)$ is an enumeration of T's theorems, then sufficient condition to the sentence Con_P and its negation are both unprovable, is that are fulfilled certain conditions of provability, introduced by Hilbert and Bernays in 1939 later simplified in this propositional form by Löb in 1955 and Jeroslow (1973). Such conditions are indeed so natural that one may wonder if it really would be a provability predicate, that predicate that does not satisfy it. However, there is no general consensus on this fact.

Definition 39. *Let $P \in \Sigma_1$ a definition of theorems of T. Löb's conditions are the following:*

1. *If $\mathsf{T} \vdash \phi$, then $\mathsf{T} \vdash P(\overline{\ulcorner \phi \urcorner})$.*
2. $\mathsf{T} \vdash P(\overline{\ulcorner \phi \urcorner}) \wedge P(\overline{\ulcorner \phi \to \psi \urcorner}) \to P(\overline{\ulcorner \psi \urcorner})$.
3. $\mathsf{T} \vdash P(\overline{\ulcorner \phi \urcorner}) \to P(\overline{\ulcorner P(\overline{\ulcorner \phi \urcorner}) \urcorner})$.

The standard proof predicate $\mathsf{Pr}_{\mathsf{T}}(x) = \exists y \mathsf{Prf}_{\mathsf{T}}(y, x)$ satisfies these conditions in sufficently strong theories. But Robinson's arithmetic is not strong enough for this purpose. It proves just the first condition. Concerning the third condition, this is an application of the so-called Σ_1 formalized completeness. This means that if T is a sufficiently strong theory, it proves $\phi \to \mathsf{Pr}_{\mathsf{T}}(\ulcorner \phi \urcorner)$, for $\phi \in \Sigma_1$, noting that $\mathsf{Pr}_T(y) \in \Sigma_1$. Once more, Robinson's arithmetic is not strong enough to prove it. The third condition is actually the most problematic and difficult to justify (see Detlefsen (2001) for an in-depth discussion around this condition). In fact the proof can be performed in the fragment of Peano arithmetic denoted $I\Delta_0 + exp$, i.e. with induction restricted to Δ_0-formulas, together with an axiom that states the totality of the function 2^x. Falling below these levels, for example in the important theory of *Bounded Arithmetic* named $\mathsf{I}\Delta_0 + \Omega_1$ (see 3) some problems may arise. In Berarducci and Verbrugge (1991) it is shown that if this theory proves the Σ_1-completeness, then it would also be valid the equation $\mathsf{NP} = \mathsf{co} - \mathsf{NP}$. More exactly, if $\mathsf{NP} \neq \mathsf{co} - \mathsf{NP}$, then there exist ϕ and ψ such that the Σ_1-completeness of Σ_1-formula:

$$\exists x(Prf_{I\Delta_0+\Omega_1}(x, \overline{\ulcorner \phi \urcorner}) \wedge \forall z \leq x \neg Prf_{I\Delta_0+\Omega_1}(z, \overline{\ulcorner \psi \urcorner}))$$

is not provable in $\mathsf{I}\Delta_0 + \Omega_1$. However, we can carefully obtain versions of the Σ_1-completeness restricted to specific subclasses, but sufficient for deducing the derivability conditions already in Buss's S^1_2. As regards the second condition, it is in practice to demonstrate that, for all a, b, if a codes a proof of ϕ, if b codes a proof of $\phi \to \psi$, then $a * b * \ulcorner \psi \urcorner$ codes a proof of ψ, and therefore it is needed the amount of induction required (induction on Σ_1-formulas is sufficient). Rosser's predicate does not meet at least one of them and Feferman's predicate of theorem on p.122 does not satisfy 3. because it is not Σ_1.

The idea behind the derivability conditions was that all formulas correctly expressing provability would have to satisfy them, although, this generally accepted idea has also found proud opponents (see primarily the works Detlefsen (1986) and Detlefsen (1979)). Regarding the second incompleteness theorem for Q, we wonder which relation exists, in Q, between the arithmetic sentence $Con(\mathsf{Q})$ and the metamathematical statement "Q is consistent". Pivotal to this debate, albeit of controversial interpretation, was a version of the second incompleteness theorem for Q in Bezboruah and Shepherdson (1976) using different methods, where nevertheless the two logicians, amazingly, sharing Kreisel's point of view, do not attach importance to this result.

Indeed, according to Georg Kreisel, which had shown similar results well in advance, $Con(\mathsf{Q})$ should not be considered, in Robinson arithmetic, as expressing its formalized consistency, but a mere algebraic property, that only in stronger systems can be said to constitute a formalization of consistency of Q. Nowadays many people think that Q can actually prove the unprovability of it's own consistency, on the basis of an argument due to Pudlák (1996), which starts from the consideration that the weak theory $\mathsf{I}\Delta_0 + \Omega_1$ is interpretabile in Q (see on p.192). The argument is the following. A formula $J(x)$ is called a *cut* of a theory T, if this theory proves:

1. $J(0)$
2. $J(x) \to J(x+1)$
3. $J(x) \wedge y \leq x \to J(y)$

Let therefore $Con_T^J = \neg\exists x(J(x) \wedge Prf_T(x, \overline{\ulcorner \overline{1} = \overline{0} \urcorner})$. Pudlàk showed that for every consistent extension T of Q and every cut $J(x)$, $\mathsf{T} \nvdash Con_{\mathsf{T}}^J$. Actually it is possible to build cuts $J(x)$ with further properties, in particular, in such a way that satisfy the axioms of $I\Delta_0 + \Omega_1$ relativized to $J(x)$. Take $\mathsf{T} = \mathsf{Q}$. Hence it is consistent to assume that a proof of a contradiction from Q is encoded in a model of $I\Delta_0 + \Omega_1$, a theory in which the meaning of Con_{Q} is not ambiguous.

But let us now see the usual argument based on the derivability conditions.

Lemma 31. *Let T be a consistent extension of Robinson arithmetic and let Con_P the formula $\neg P(\overline{\ulcorner \overline{0} = \overline{1} \urcorner})$, where P satisfies Löb's conditions; then $\mathsf{T} \vdash Con_P \leftrightarrow \neg P(\overline{\ulcorner \phi \urcorner}) \vee \neg P(\overline{\ulcorner \neg\phi \urcorner})$, for any ϕ.*

Proof. $\Leftarrow$ Recall that $\neg(\overline{0} = \overline{1})$ is an axiom; from the *first derivability condition* in T we get $P(\overline{\ulcorner \neg(\overline{0} = \overline{1}) \urcorner})$, from which, by propositional tautology

$$A \to (B \to (A \wedge B)$$

we obtain $\neg Con_P \to (P(\overline{\ulcorner \neg(\overline{0} = \overline{1}) \urcorner}) \wedge P(\overline{\ulcorner \overline{0} = \overline{1} \urcorner})$. From the first condition and tautology $A \to (\neg A \to B)$ we get $P(\overline{\ulcorner (\overline{0} = \overline{1}) \to (\neg(\overline{0} = \overline{1}) \to \phi) \urcorner})$ and by the second derivability condition and tautology $(A \to (B \to C)) \leftrightarrow ((A \wedge B) \to C)$ lastly $P(\overline{\ulcorner (\overline{0} = \overline{1}) \urcorner}) \wedge P(\overline{\ulcorner \neg(\overline{0} = \overline{1}) \urcorner}) \to P(\overline{\ulcorner \phi \urcorner})))$. By modus ponens we obtain $\neg Con_P \to P(\overline{\ulcorner \phi \urcorner})$; we can repeat the argument and in specular way to obtain $\neg Con_P \to P(\overline{\ulcorner \neg\phi \urcorner})$. Now $\Rightarrow$. In T, from a propositional tautology and the first condition we get $P(\overline{\ulcorner \phi \to (\neg\phi \to (\overline{0} = \overline{1})) \urcorner})$ and still for simple propositional transformations, using the second condition, we finally obtain $P(\overline{\ulcorner \neg\phi \urcorner}) \wedge P(\overline{\ulcorner \phi \urcorner}) \to \neg Con_P$. Take the contronominal. QED

Theorem 83. (The second incompleteness theorem) *Let T a consistent extension of Robinson arithmetic and $P(x)$ a Σ_1-definition of its theorems, satisfying the derivability conditions. Hence $T \nvdash Con_P$; if moreover T is also $\omega-$consistent (or $\Sigma_1 -$ sound), then $T \nvdash \neg Con_P$.*

Proof. Let ν be a fixed point of $\neg P(x)$ in T, namely $\mathsf{T} \vdash \nu \leftrightarrow \neg P(\overline{\ulcorner \nu \urcorner})$. By the previous lemma $\neg P(\overline{\ulcorner \neg\nu \urcorner}) \vee \neg P(\overline{\ulcorner \nu \urcorner}) \to Con_P$ and therefore consider the meaning of ν, lastly $\nu \to Con_P$. On the other hand T we also have $\neg\nu \to P(\overline{\ulcorner \nu \urcorner})$. By the third condition $P(\overline{\ulcorner \nu \urcorner}) \to P(\overline{\ulcorner P(\overline{\ulcorner \nu \urcorner}) \urcorner})$ and by the meaning of ν, from this we get $P(\overline{\ulcorner \nu \urcorner}) \to P(\overline{\ulcorner \neg\nu \urcorner})$ (where we have replaced $\neg\nu$ in place of $P(\overline{\ulcorner \nu \urcorner})$ in the consequent); hence we have $\neg\nu \to P(\overline{\ulcorner \neg\nu \urcorner}) \wedge P(\overline{\ulcorner \nu \urcorner})$ and in a few logical steps, using the previous lemma, finally, we get also $\neg\nu \to \neg Con_P$. QED

What about sentences asserting their own *provability*? This queston is known as the *Henkin's question.*

Theorem 84. (Löb 1955) *Let ϕ be a sentence and $\mathsf{Pr}_{\mathsf{T}}(x)$ the standard proof predicate for T . Then $\mathsf{T} \vdash \mathsf{Pr}_{\mathsf{T}}(\overline{\ulcorner \phi \urcorner}) \to \phi$ iff $\mathsf{T} \vdash \phi$.*

Proof. $\Leftarrow$ clear. $\Rightarrow$ Let us suppose that $\mathsf{T} \nvdash \phi$. Hence $\mathsf{T} + \neg\phi$ is consistent and therefore $\mathsf{T} + \neg\phi \nvdash Con_{\mathsf{T}+\neg\phi}$, namely $\mathsf{T} \nvdash \neg\phi \rightarrow \neg\mathsf{Pr}_{\mathsf{T}}(\overline{\ulcorner \neg\phi \rightarrow (1 = 0) \urcorner})$ and therefore $\mathsf{T} \nvdash \neg\phi \rightarrow \neg\mathsf{Pr}_{\mathsf{T}}(\overline{\ulcorner \phi \urcorner})$. QED

We get a soluyion to Henkin's question as a corollary.

Corollary 19. $\mathsf{T} \vdash \mathsf{Pr}_{\mathsf{T}}(\overline{\ulcorner \phi \urcorner}) \leftrightarrow \phi$ *iff* $\mathsf{T} \vdash \phi$.

This result (sometimes seen as a generalization of Gödel's result, taking $1 = 0$ for ϕ) focuses our attention on so called *reflection shemas*.

Hilbert's program of elimination of abstract entities. Principles like the above "$\mathsf{Pr}_{\mathsf{T}}(\overline{\ulcorner \phi \urcorner}) \rightarrow \phi$", represent a formalization of *soundness* called *reflection principles*; the main schematic representations of them are the following:

1. (*Local* Reflection Rfn_{T}) $\mathsf{Pr}_{\mathsf{T}}(\overline{\ulcorner \phi \urcorner}) \rightarrow \phi$

2. (*Uniform* Reflection RFN_{T}) $\forall x(\mathsf{Pr}_{\mathsf{T}}(\overline{\ulcorner \phi(\dot{x}) \urcorner}) \rightarrow \phi(x))$

(where $\ulcorner \phi(\dot{x_0}, ..., \dot{x_n}) \urcorner$ is the function that associates to a given sequence of numbers $m_0, ..., m_n$, the number $\ulcorner \phi(\overline{m_0}, ..., \overline{m_n}) \urcorner$). The second principle is stronger than the first. Actually $\mathsf{T} + RFN_{\mathsf{T}}$ proves $Con(\mathsf{T} + Rfn_{\mathsf{T}})$.

If we consider *partial scheme* $RFN_{\mathsf{T},\Pi_1}, Rfn_{\mathsf{T},\Pi_1}$, i.e. restricted, in particular, to Π_1 -formulas, and if Con_{T} formalize consistency of T, then we have that the principles Con_{T}, RFN_{T,Π_1} and Rfn_{T,Π_1} are equivalent. This allows us to better clarify the link between the Hilbert program of consistency and that of Π_1-conservativity, i.e. the statement that is exposed by Hilbert in the 1926 essay on infinity, in these terms:

> the extension by the addition of ideal elements is legitimate, if no contradiction is determined in the old, restricted domain, namely if relations that result for old objects, when ideal objects are eliminated are valid in the old domain.

Since we tied the finitary mathematics to Skolem arithmetic, suppose here to fix ideas, that T is an extension of PRA. Hence consider the formula $\psi(x) \in \Pi_1$: since $\neg\psi(x)$ will be then Σ_1, if the theory T is sufficiently strong (as it is PRA) to show the Σ_1- formalized completeness, then it will show $\neg\psi(x) \rightarrow \mathsf{Pr}_{\mathsf{T}}(\overline{\ulcorner \neg\psi(\dot{x}) \urcorner})$, and therefore $\neg\mathsf{Pr}_{\mathsf{T}}(\overline{\ulcorner \neg\psi(\dot{x}) \urcorner}) \rightarrow \psi(x)$.

But $\mathsf{PRA} + Con_{\mathsf{T}}$ proves $\mathsf{Pr}_{\mathsf{T}}(\overline{\ulcorner \psi(\dot{x}) \urcorner}) \rightarrow \neg\mathsf{Pr}_{\mathsf{T}}(\overline{\ulcorner \neg\psi(\dot{x}) \urcorner})$ - see the preparatory lemma to the proof of the second Gödel 's theorem - and therefore, lastly, $\mathsf{Pr}_{\mathsf{T}}(\overline{\ulcorner \psi(\dot{x}) \urcorner}) \rightarrow \psi(x)$, as claimed. On the other hand, if we use the principle of reflection for Π_1-formulas, this will apply in particular to the formula $\overline{1} = \overline{0}$; however it is an axiom that $\neg(\overline{1} = \overline{0})$, from which $\neg\mathsf{Pr}_{\mathsf{T}}(\overline{\ulcorner (\overline{1} = \overline{0}) \urcorner})$, namely Con_{T}.

Thus we have shown in Skolem's theory PRA the equivalence between RFN_{Π_1} and Con_{T}. But from this follows that if a formula $\psi(x)$ of the language of Skolem arithmetic, of complexity Π_1 is provable in T, then it is also in $\mathsf{PRA} + Con_{\mathsf{T}}$: indeed, for a theory strong enough as Skolem arithmetic, the first provability condition can be better refined in this sense: "if $\mathsf{T} \vdash \psi(x)$, then $\mathsf{PRA} \vdash \mathsf{Pr}_{\mathsf{T}}(\overline{\ulcorner \psi(\dot{x}) \urcorner})$", from which $\mathsf{PRA} + RFN_{\Pi_1} \vdash \psi(x)$ and for the equivalences have just demonstrated, $\mathsf{PRA} + Con_{\mathsf{T}} \vdash \psi(x)$.

Suppose, then, that T embodies the infinitary mathematics in sense of Hilbert, and PRA, as in Tait thesis, incorporates finitary mathematics. Remember that, according to Hilbert, the "concrete" sentences have the form $\forall x(f(x) = 0)$, with $f(x)$ primitive recursive, and are therefore Π_1- sentences of language PRA. If as expected by Hilbert (and refuted by the second theorem of incompleteness) the consistency of T was demonstrated in PRA, then, as now established, we would have that if $\psi(x)$ be provable in T, the would already in PRA: in the proof of "concrete sentences", the "ideal entities" (to quote Hilbert) could be eliminated.

This allows us to make another point. For a time it is approached primarily to *second* Gödel theorem as a refutation of Hilbert's program. Subsequently, between the '70s and late' 80s, by logicians like Kreisel, Prawitz, Simpson, or Smorynski,the focus is shifted more on the first theorem.

For example Smorynski drew from the connection that we have now illustrated between the program of *consistency* and the *conservativity* program, the conclusion that already from the first theorem we can derive a refutation of Hilbert program on consistency, in this sense: the sentence which gives us the first incompleteness theorem, has precisely the form of a Π_1- sentence " concrete " $\forall x R(x)$ which is undecidable (and notice that its particular instances $R(\overline{0})$, $R(\overline{1})$, $R(\overline{2})$... *are* provable).

If T is the transfinite system containing ideal elements, it shows all true sentences of finitary mathematics; but the program of conservativity precisely requires all Π_1- sentences "concrete" provable in T are then demonstrable already in PRA; but the first theorem of incompleteness, rather, shows a true Π_1- sentence of language PRA (and therefore provable in T), but unprovable in PRA itself. The program of conservativity is therefore impossible: hence, for what we said above, it is also that of consistency.

5.2. Beating incompleteness: Turing's progressions

The second Gödel's theorem gives a way to extend a theory to a stronger theory and allows to see the phenomenon of inexaustibility of mathematics from a particular point o view (see Franzen (2003)): the consistency statement $Con(\mathsf{T})$ is indipendent from the consistent theory T, although true; hence, if T is a sound theory (i.e. does noot prove false things), also the theory T' obtained from T adding the sentence $Con(\mathsf{T})$ as a new axiom will be sound; moreover it will be stronger than T (for instance it will show that $Con(\mathsf{T})$). But also T' will be incomplete and therefore $Con(\mathsf{T}')$ will be indipendent from it, but we can define a stronger theory T'' that decides $Con(\mathsf{T}')$, by adding to T' the true sentence $Con(\mathsf{T}')$ and so on.

This leads to the idea of an effective association of formal systems S_α with ordinals α, but that can be done only for countable ordinals and to deal with limits in an effective way, it turns out that we must work not with ordinals per se, but with recursive ordinals, or *notations* for ordinals. In his PhD thesis. 1937 in Princeton, Alan Turing formalized this intuition, by introducing the notion of ordinal logic and a suggestive idea to "overcome incompleteness" iterating a transfinite number of times the operation of adding an undecidable sentence to a theory, of the kind of reflection statements or consistency statements, hoping to get to a certain point a complete theory, with respect to a significant class of sentences. This began a series of studies around transfinite recursive hierarchies axiomatic theories. As Turing (1939) says:

> The well-known theorem of Gödel (1931) shows that every system of logic is in a certain sense incomplete, but at the same time it indicates means whereby, from a system L of logic, a more complete system L' may be obtained. By repeating the process we get a sequence $L_1 = L', L_2 = L_1^{'}$... each theory more complete than the preceding... A logic L_ω may be constructed in which the provable theorems are the totality of theorems provables with the help of logics $L, L^{'}, L^{''}$....

Recall Feferman's approach to the proof-predicate which we discussed earlier. He consider primarily the arithmetization of axiom-hood: in the proof-predicate Prf_τ he clearly distinguishes a proof-predicate Prf for the first order logic, which is fixed, from the definition $\tau(x)$ of the specific mathematical axioms, which may change. A major case is when $\tau(x) \in \Sigma_1^0$, and *binumerates* the axioms, and therefore the whole expression $\exists y Prf_\tau(y, x)$ becomes in turn Σ_1^0. The derivability conditions are satisfied.

We now need to recall some set theory concepts. Recall that from a set theoretic point of view an ordinal is just a transitive set (i.e. an y such that if $x \in y$, then $x \subseteq y$) well (and therefore linearly) strictly ordered by the appartenence $\in$, i.e. every non-empty subset of it contains a $\in$-least element. This notion generalises the definition of natural numbers *à la* Von Neumann $0 = \emptyset$, $n + 1 = n \cup \{n\} = \{0, 1, 2, ..., n\}$ by admitting limit ordinals as for example $\omega = \{0, 1, 2, ...\}$, and so on. Let us recall some basic properties (see e.g. Jech (1978)):

1. Every member of an ordinal is an ordinal, however the class of all ordinals is not a set, but a proper class: since $\in$ well-orders this class, otherwise it would be in turn an ordinal.

2. For every ordinal α, also $\alpha \cup \{\alpha\}$ is an ordinal, and this is the successor of it, i.e., $\alpha < \alpha \cup \{\alpha\}$ and there is no β such that $\alpha < \beta < \alpha \cup \{\alpha\}$. An ordinal is a *successor*, if it has this form: otherwise, either is 0, or it is a *limit*.

3. If α is 0 or a limit ordinal then $sup\ \alpha = \alpha = \bigcup \alpha$.

4. If α is a successor ordinal then $sup\ \alpha$ is the predecessor of α.

5. If α is a limit ordinal and $\gamma < \alpha$, then there is an ordinal δ such that $\gamma < \delta < \alpha$, and in particular $\gamma < \gamma \cup \{\gamma\} < \alpha$. Say that α is a finite ordinal, or a natural number, if $\alpha = 0$, or α is a successor and every ordinal $\beta < \alpha$, in turn is 0 or a successor.

6. Every well ordered set is isomorphic to a unique ordinal and the isomorphism too is unique.

A *countable* ordinal is an ordinal whose set of predecessors can be matched up with natural numbers. For instance, all these are countable:

$$\left.\omega,\ \omega + n,\ \omega + \omega,\ \omega \times n,\ \omega \times \omega,\ \omega^{\omega} ...\ \epsilon_0 = \omega^{\omega^{\omega^{\omega^{\cdot^{\cdot^{\cdot}}}}}}\right\} \omega - times$$

These have all the same cardinality and are essentially different ordering of natural numbers. *Uncountable* ordinals exist. The first is denoted by ω_1. Recall that each well ordering is isomorphic to a unique ordinal. A *computable* ordinal is an ordinal isomorphic to a computable well ordering, i.e. if there is a computable relation on a subset of the integers that is well-ordered and isomorphic to it. The first *non-computable* ordinal is denoted by ω_1^{CK}. It holds that $\omega_1^{CK} < \omega_1$ There are uncountably many countable ordinals, so, if we want "give a name" to each one (i.e. code by numbers), we cannot invent disinct names. We shall see a method, due to Kleene, to assign names to so-called *constructive* ordinals. Spector has proved that the computable ordinals are just Kleene's constructive ordinals.

Given an effective description of formal systems $\mathsf{S}_0, \mathsf{S}_1, \mathsf{S}_2$.... in the same language, starting e.g. from $\mathsf{S}_0 = \mathsf{PA}$, we can form the union $\mathsf{S}_\omega = \bigcup_i \mathsf{S}_i$ i.e. the formal system made of the set the axioms of all S_i, but if each S_{i+1} results by the interaction of one and the same operation, for example $\mathsf{S}_{i+1} = \mathsf{S}_i + Con(\mathsf{S}_i)$, we can continue in the transfinite, defining $\mathsf{S}_{\omega+1} = \mathsf{S}_\omega + Con(\mathsf{S}_\omega)$ and so on. To the limit steps λ we take the union $S_\lambda = \bigcup_{\beta<\lambda} \mathsf{S}_\beta$. If S_i is sound (i.e. correct), then (being $Con(\mathsf{S}_i)$ a true sentence, although independente from S_i) also S_{i+1} will be sound. From here the idea to associate in an effective way to a countable ordinal α, a formal system S_α, in the framework of an inductive construction of the kind which we have referred.

There are several ways to make rigorous this topic: it is necessary first to give a Σ_1-formula defining arithmetically the axioms of the theories that constitute the sequence, so as to be able to express, for example, that a certain theory is consistent (however, in relation to what we said about consistency predicates, it should be emphasized the character in some respects intensional of this construction, which depends significantly on how the axioms of a theory S_i are defined) and the problematic case concerns the limit steps. For instance, if $\tau_0(x)$ is a definition of axioms of PA, then we can define $\mathsf{S}_0 = \mathsf{PA}$ and the axioms of S_{i+1} will be defined by the formula τ_{i+1} satisfying $\tau_{i+1}(x) \leftrightarrow \tau_i(x) \vee x = \overline{\ulcorner Con_{\tau_i} \urcorner}$.

The first problem now concerns the definition of the axioms of S_ω. For this purpose we can introduce a notation for countable ordinals, i.e. their coding by integers that allows us to describe them arithmetically. The idea is then to think a limit ordinal as a sequence of ordinals converging to it, enumerated from a function $\phi_e(x)$, so that for the definition of the axioms of a theory idexed with a limit ordinal, is intuitively fulfilled this condition "$\tau_{lim(e)}(y)$ iff there is an n such that $\tau_{\phi_e(n)}(y)$". With Ordinal Logic we mean a sequence of theories $\mathsf{S}_{\mathsf{a}_0}, \mathsf{S}_{\mathsf{a}_1}, \mathsf{S}_{\mathsf{a}_2}$.... where each a_i is the name of an ordinal, i.e. a number in Kleene's $\mathcal{O}$ we are going to define.

It should be noted, however, that the same limit ordinal can have different notations (think for example that ω is the limit of all strictly increasing computable infinite sequence of natural numbers) and is not sure that if two numbers a_j e a_i denote the same ordinal, then the theories

$\mathsf{S}_{\mathsf{a}_i}, \mathsf{S}_{\mathsf{a}_j}$ prove the same theorems: if this is the case, then the logic is *invariant*. Turing proved, however, that an ordinal logic can be invariant, or complete for Π_1 statements, but not both things together.

Definition 40. *An ordinal is said computable iff it is isomorphic to a recursive well order.*

It is well known that there is at least one *countable* ordinal, but *not computable*, and the least of these ordinals is denoted with ω_1^{CK}.

1. In 1936 Church and Kleene introduced a satisfactory characterization for this concept, through the notion of *constructive ordinal* and a system $\mathcal{O}$ of notations, i.e. of codes for countable ordinals.

2. the set $\mathcal{O}$ of constructive ordinals provably coincide with that of computable ordinals. If we write $\widetilde{a_i}$ for the ordinal denoted by a_i, then we can assert that ω_1^{CK} is the smallest ordinal not in the form $\widetilde{a}$, for some $a \in \mathcal{O}$.

A significant part of the theory of ordinal numbers can be formulated as a theory of *ordinal notations*.

Notation system. It is a function ν having domain a subset D of the naturals and codomain $X \subset On$, for which there exists a recursive partial function $k(x)$ such that:

$$k(x) = \begin{cases} 0 & \text{if } \nu(x) = 0 \\ 1 & \text{if } \nu(x) = \beta + 1 \\ 2 & \text{if } \nu(x) = \lambda \text{ limit} \end{cases}$$

1. There is a partial recursive function $p(x)$ such that $p(x)$ converges and $\nu(x) = \nu(p(x)) + 1$, if $\nu(x) = \beta + 1$.

2. there is a function $q(x)$ such that $\phi_{q(x)}$ is total and the sequence:

$$\nu(\phi_{q(x)}(0)) < \nu(\phi_{q(x)}(1)) < \nu(\phi_{q(x)}(2)) < ..$$

has limit $\nu(x)$, if $\nu(x) = \lambda$ limit.

In 1936, Church and Kleene had introduced a system of constructive ordinal notations, given by certain expressions in the lambda-calculus. A variant of this uses numerical codes, and associates with each number a countable ordinal.

Kleene's system $\mathcal{O}$. The intuitive idea is to code $0, 1, 2, 3...$ with the powers $1, 2, 2^2, 2^{2^2}...$ and if α is a limit ordinal, then its notations are all numbers $3 \cdot 5^e$ such that ϕ_e is a total function such that $\phi_e(n)$ is a number in $\mathcal{O}$ that denotes α_n and $\alpha_0, \alpha_1, \alpha_2...$ is an increasing sequence converging to α.

Kleene's system $\mathcal{O}$. For the sake of simplicity let us write $\widetilde{b}$ in place of $\nu(b)$.

1. 0 receives notation 1.

2. Suppose we have assigned a notation to all ordinals less than α and having defined $<_o$ on it:

   (a) if $\alpha = \beta + 1$ and $\widetilde{b} = \beta$, then $\widetilde{2^b} = \alpha$ and add the pairs $\langle z, 2^b \rangle$ to the relation $<_o$, for all $z \leq_o b$.

   (b) If α is a limit ordinal, it can be understood as a sequence of ordinals converging to it. Suppose that this sequence is enumerated by a total function ϕ_e with values in $\mathcal{O}$, such that for all n, $\phi_e(n) <_o \phi_e(n+1)$, where $\phi_e(n) = a_n$ and the increasing sequence $\widetilde{a_0}, \widetilde{a_1}, \widetilde{a_2}, ...$ has limit α; then $\widetilde{3 \cdot 5^e} = \alpha$; add moreover each pair $\langle z, 3 \cdot 5^e \rangle$ such that $z <_o \phi_e(n)$ for some n, to the relation $<_o$.

(c) $k(1) = 0$, $k(2^x) = 1$, $k(3 \cdot 5^y) = 2$, $p(2^x) = x$, $q(3 \cdot 5^y) = y$.

Note that the natural numbers receive a fixed notation, but not limit ordinals. If the sequence $\phi_e(0), \phi_e(1), \phi_e(2), ...$ denotes effective enumeration of ordinals converging to α, then there are infinite choices for e, both because each function has infinitely many index and because there are infinitely many increasing sequences of ordinals converging to α. In summary, the system begins with $1, 2, 2^2, 2^{2^2}, 2^{2^{2^2}}$...., as notations for $0, 1, 2, 3, ...$ To the limit ordinal α we assign a notation $3 \cdot 5^e$, where $\phi_e(x)$ is total recursive such that if $\phi_e(n) = b_n$, then $b_0 <_o b_1 <_o b_2 <_o ...$ and $\alpha = lim_n \alpha_m$, where b_n denotes α_n.

Hence, for instance, ω is denoted by $3 \cdot 5^e$, for infinite e. The order $<_o$ is therefore a tree, that in correspondence with a node labeled by a limit ordinal, branches in infinite branches.

$$1 <_o 2 <_o 2^2 <_o 2^{2^2} <_o ... \begin{cases} 3 \cdot 5^{e_0} <_o 2^{3 \cdot 5^{e_0}} <_o 2^{2^{3 \cdot 5^{e_0}}} <_o & ... \\ 3 \cdot 5^{e_1} <_o 2^{3 \cdot 5^{e_1}} <_o 2^{2^{3 \cdot 5^{e_1}}} <_o & ... \\ ... & ... \\ 3 \cdot 5^{e_n} <_o 2^{3 \cdot 5^{e_n}} <_o 2^{2^{3 \cdot 5^{e_n}}} <_o & ... \\ ... & ... \end{cases}$$

The following facts are well known (see Sacks (2017), ch.1):

1. There exist a recursive function $+_o$ such that for all $x, y \in \mathcal{O}$:
   (a) $x +_o y \in \mathcal{O}$
   (b) $\widetilde{x +_o y} = \widetilde{x} + \widetilde{y}$
   (c) $y \neq 1 \rightarrow x <_o x +_o y$
2. For every other notation system S exists $\phi : S \longrightarrow \mathcal{O}$ partial recursive, such that:
3. if $x \in S$, then $\nu_S(x) \leq \widetilde{\phi(x)}$.
4. $\omega_1^{CK} = \bigcup_{a \in \mathcal{O}} \widetilde{a}$=set of recursive ordinals.
5. (Kleene) $\mathcal{O}$ is Π_1^1-complete, i.e. is Π_1^1-definable and if $Y \in \Pi_1^1$, then $Y \leq_m \mathcal{O}$.

The version of the second incompleteness theorem presented by Feferman (1960) has this general form.

Theorem 85. *Let* T *be a consistent extension of* PA*, and let* $\tau(x)$ *be a* Σ_1*-formula which numerates the axioms of* T*, and* $Cons(\mathsf{T})$ *be a consistency statement constructed from* $\tau(x)$ *and* $Prf_\tau(x, y)$*. Then* $Cons(\mathsf{T})$ *is not provable in* T*.*

Hence we start with a Σ_1 numeration of the axioms of such a theory T. Given a numeration $\tau(x)$, we naturally construct the formula $Prf_\tau(y, x)$ that expresses the predicate is the Gödel number of the proof, in T, of the formula with the number x" and the formula of provability in T, $\exists y Prf_\tau(y, x)$. Following Beklemishev (1992) we can now make more precise what we have said in the introduction, through this definition. First of all we introduce a provability predicate with a parameter n for the "level" $Prf_\tau(n, y, x)$, which is the provability predicate for the theory axiomatized by $\tau(n, x)$ defined below (hence $Con_\tau(n)$ will be $\neg \exists y Prf_\tau(n, y, \ulcorner \overline{1} = \overline{0} \urcorner)$). Moreover, recall that "$\phi_e(x) \simeq y$" can be formalized by a Σ_1 formula $\sigma(e, x, y)$.

Definition 41. *A* Σ_1 *formula* $\tau(n, x)$ *is a* verifiable enumeration *for* $\tau(x)$ *if* PA *proves the following:*

1. $a = 1 \vee (\forall u(a \neq 2^u) \wedge \forall u(a \neq 3 \cdot 5^u)) \rightarrow (\tau(a,x) \leftrightarrow \tau(x))$
2. $a \geq 1 \rightarrow (\tau(2^a,x) \leftrightarrow \tau(a,x) \vee \overline{\ulcorner Con_\tau(\dot{a})\urcorner})$
3. $\tau(3 \cdot 5^a,x) \leftrightarrow \tau(x) \vee \exists u \exists w(\sigma(e,u,w) \wedge \tau(w,x))$

Theorem 86. *For all Σ_1 binumeration $\tau(x)$, there exists a Σ_1 formula $\tau(z,x)$ provably equivalent to the disjunction of the following:*

1. $(z = \overline{1} \vee \forall u \leq z(z \neq 2^u \wedge z \neq s \cdot 5^u)) \wedge \tau(x)$
2. $\exists u \leq z(u \neq \overline{0} \wedge z = 2^u \wedge (\tau(u,x) \vee x = \overline{\ulcorner Con_\tau(\dot{u}))\urcorner}$
3. $\exists u \leq z(z = 3 \cdot 5^u \wedge (\tau(x) \vee \exists v \exists w(\sigma(u,v,w) \wedge (w \neq z \wedge \tau(w,x)) \vee (w = z \wedge x = v)$

Proof. By using the fixed point theorem and partial truth predicates (recall that for all $e \in \mathcal{O}$ and n, $\phi_e(n) \neq 3 \cdot 5^e$). QED

We have the following:

1. If $a \in \mathcal{O}$, the formula $\tau(\overline{a},x)$ gives the axioms of the theory $\mathsf{S_a}$ in the progression.
2. $\mathsf{S_1} = \mathsf{PA}$.
3. $\mathsf{S_{2^a}} = \mathsf{S_a} \cup \{Con_\tau(\overline{a})\}$.
4. If λ denotes a limit ordinal, then $\mathsf{S}_\lambda = \bigcup_{d<_O\lambda} \mathsf{S_d}$,
5. Lastly, for all $a, b \in \mathcal{O}$, PA proves that if $a <_O b$, then $\tau(a,x) \rightarrow \tau(b,x)$.

The *Turing progression* with enumeration $\tau(e,x)$, is the sequence $\{S_e\}_{e\in\mathcal{O}}$ of theories enumerated by the formula $\tau(e,x)$.

Theorem 87. (Turing 1937) *For all progressions and all true Π_1^0-sentence $\forall x\psi(x)$, there exist a notaton $a \in \mathcal{O}$ such that $\widetilde{a} = \omega + 1$ and $\forall x\psi(x)$ is provable in $\mathsf{S_a}$.*

Proof. We give an informal sketch of the proof, following Feferman (2006). Turing proceeded as follows: we denote for simplicity $\mathsf{S(a)}$ the code of the successor 2^a and with $lim(a)$ the code of the limit $3 \cdot 5^a$. Let n_O the notation for the natural number n. We begin by defining (using the recursion theorem) function:

$$\phi_e(n) = \begin{cases} n_O & \text{if for all } k \leq n, \psi(\overline{k}) \text{ is true} \\ \mathsf{S(lim(e))} & \text{if there is } k \leq n \text{ such that } \psi(\overline{k}) \text{ is false} \end{cases}$$

where $\psi(x)$ is decidable. Note that if by hypothesis $\forall x\psi(x)$ is a true Π_1 sentence then for all n, we have $\phi_e(n) = n_O$ and therefore the sequence of values $\phi_e(0), \phi_e(1), \phi_e(2)$, is just the sequence $0_O, 1_O, 2_O, ...$ and $lim(e)$ is therefore an element of $\mathcal{O}$ that denotes ω. Reason inside $\mathsf{S_{S(lim(e))}}$, checking in this theory that if $\mathsf{S(lim(e))}$ is consistent, then the sentence $\forall x\psi(x)$ is true: indeed, suppose by contradiction that the sentence $\forall x\psi(x)$ is false; hence we have that for some n, the sentence $\psi(\overline{n})$ is false. But then the theory $\mathsf{S(lim(e))}$, i.e. the union of the sequence $\mathsf{S}_{\phi_e(0)}, \mathsf{S}_{\phi_e(1)}, S_{\phi_e(2)}...$ will be such that for some n and for all $k \geq n$ (i.e. from a certain point onwards) we will have $\phi_e(k) = \mathsf{S(lim(e))}$; hence from a certain point onwards $\mathsf{S_{S(lim(e))}}$ and $\mathsf{S(lim(e))}$ will coincide, and therefore $\mathsf{S(lim(e))}$ will prove its own consistency (being $Con(\mathsf{S(lim(e))})$ contained in $\mathsf{S_{S(lim(e))}}$);it follows from the second Gödel's result that $\mathsf{S(lim(e))}$ is inconsistent. But $\mathsf{S_{S(lim(e))}}$ actually proves the consistency of $\mathsf{S(lim(e))}$. Ergo $\mathsf{S_{S(lim(e))}}$ proves $\forall x\psi(x)$. Observe that $\mathsf{S(lim(e))}$ denotes $\omega + 1$. QED

Turing, however, was not satisfied with this result, which in his view shifted the problem to determine whether a Π_1^0-sentence is true, to the problem considerably more complex to determine whether a number belongs to $\mathcal{O}$:

> My completeness theorem [...] is completely useless for the purpose of producing proofs (Turing (1940)).

Turing also obtained a completeness result for Π_1 statements via the transfinite iteration of the local reflection principle, in place of the consistency statement. Further results were obtained in Feferman (1962) that strengthen Turing's theorems, where progressions obtained starting from $\mathsf{PA} = \mathsf{S}_0$ were studied, iterating the principle of *universal* reflection principle (see on p.125). It was proved, for instance, that for all true arithmetical sentences θ there exists an $a \in \mathcal{O}$, denoting an ordinal less or equal to $\omega^{\omega^{\omega+1}}$ such that $\mathsf{S}_a \vdash \theta$ (where on the contrary Turing's progression based on iteration of consistency statements is not complete for true Π_2-sentences). Feferman also proved that Turing's progression based on iteration of consistency statements is not complete for true Π_2 statements.

However, it was disappointing the fact that although two notations a and b denote the same ordinal, this does not imply that S_a proves the same sentence that S_b. We say that this ordinal logic is *not invariant* and Turing actually showed that an ordinal logic cannot be both complete for Π_1 sentences and invariant (see Feferman (2006)). In Franzen (2003) it is explained that the reason why a Π_1 sentence ψ can only be proved at stage $\omega + 1$ in a Turing consistency sequence is that at stage ω of the construction a non-standard definition of the axioms can be introduced in such a sequence, depending on the definition of ϕ_e. In Feferman's theorem a path (i.e. a subset $P \subseteq \mathcal{O}$ lineraly orderd by $<_{\mathcal{O}}$ and such that if $b \in P$ and $c <_{\mathcal{O}} b$, then $c \in P$) of length $\omega^{\omega^{\omega+1}}$ was generated, such that $\bigcup_{a \in P} S_a$ is complete for all arithmetical sentences, but that completeness result essentially depends on the choice of the path and the result does not hold for other paths. Hence the invariance property still fails. In other words, these result are sensitive to the choice of notation. As a consequence of this, the crucial problem then became to discover natural conditions to impose on the choice of ordinal notations used to index theories in a progression. In an attempt to resolve this issue, in Kreisel (1960) it was required that progressions should satisfy a kind of "autonomy" requirement and the purpose was to generate the hierarchy of theories via a kind of boot-strapping process, through so-called *autonomous progressions*, that are in some sense self-justifying, being characterised by the fact that we are allowed to advance to the step S_a, only if we have obtained a proof that $a \in \mathcal{O}$ in some previously accepted theory S_b where $b <_O a$.

This line of research has had a strong impact in Proof Theory. The notion of autonomy was used by Kreisel in his proposals to characterize constructive philosophies of mathematics as finitism and predicativity by suitable autonomous progressions of theories, say $\{F_a\}$ and $\{R_a\}$, respectively. While Kreisel (1960) established that a least upper bound for $\widetilde{a}$ appearing in the first sequence is ε_0, Feferman (1964) and Schütte (1964, 1965) considered the so-called *Veblen hierarchy* of functions ϕ_α:

1. $\phi_\alpha(\beta) = \omega^\beta$, if $\alpha = 0$,
2. ϕ_α enumerates the set $\{\xi | \phi_\gamma(\xi) = \xi, for\ all\ \gamma < \alpha\}$ if $\alpha > 0$

and came to the conclusion that an ordinal Γ_0 which is the least α such that $\phi_\alpha(0) = \alpha$ is the analogous limiting ordinal in the second progression. The received view, even if not agreed unanimously (see for instance Weaver (2005), that strongly disagree with this interpretation), is that this represents the smallest predicatively non-provable ordinal. See Beklemishev (1995), Franzen (2003), Franzen (2004) for a broader overview and Feferman's appendix to Takeuti (1987) to explore these developments further. On Feferman's predicativism see Crosilla (2017).

5.3. Propositional Provability Logic: classical vs. intuitionistic arithmetic

Provability Logic is a modal logic, in which the boxed formula $\Box\phi$ ("it is necessary that ϕ") is interpreted as "it is provable (in some formal theory) that ϕ". Directly related to the second incompleteness theorem, it was a vast research programme that involved many Italian logicians and characterised the Sienese school of logic in particular from the 1970s to the 1990s. The literature too is vast, but here, in keeping with the slant of the book, we would like to draw attention to a particular chapter: the relations with intuitionist constructive logic. Why consider

intuitionist theories? In fact, this interest dates back to the origins of studies in this field, and not only on the part of the Dutch school. Later on, we will explain in more detail that the first, partial examples of "arithmetical completeness" results come from this area. However, the generalisation of Solovay's seminal scientific achievement (see below), moving from classical mathematical theories to intuitionistic ones, although it has been a constant interest, especially on the part of the Dutch school (de Jongh, Visser, Iemhoff), has proved fraught with difficulties and has only made significant progress in very recent years. From the early years of this research programme, the issue also affected Italian school. What has been made clear is that the provability logic of HA is not a sublogic of that of PA. Let's start in this regard with a little history. At the three seminars of the Siena's annual meeting named ''Incontri di logica matematica" in 1982, many talks were focused on the topic of *constructive logic and mathematics*. At the first seminar Franco Montagna gave a talk on the subject: *Solovay's theorem and Heyting Arithmetic*. Solovay's theorem is the main result in the field of so-called *Provability Logic*, and essentially states that the modal logic GL captures everything Peano arithmetic PA (or other fragments of arithmetic considered) can truthfully say about their own provability predicate. Heyting arithmetic HA is the intuitionistic analogous of classical Peano arithmetic (only the logic is different). In the subsequent years, research on these two topics intertwined. Montagna concluded his talk by posing these problems:

1. Give an axiomatization of the provability logic of Heyting's intuitionistic arithmetic.

2. Determine whether this logic is decidable.

The whole problem remained unsolved for a long time (see Visser and Beklemishev (2006)) and important results have been achieved only recently. Research on provability logic was carried out mainly between the '70s and' 90s in various universities: Prague, Moscow, Amsterdam, Utrecht, Siena, Oxford, Manchester. In particular, in Siena, in the early 70s, R. Magari and other people proposed an algebraic approach to the formal provability which led to the concept of diagonalizable algebra. Sambin and Ursini studied an intuitionistic version. The research in this field continued after the 90s at other universities, and especially in the Netherlands addressed to the provability logic of *intuitionistic arithmetic*. Those who know the modal logic will recognize in the Löb conditions, discussed in the framework of the second incompleteness theorem, modal axioms and rules which are reflected in the propositional Provability Logic GL ("Gödel-Löb logic"). This is a (classical) modal logic in which the box operator $\Box$ is interpreted as the provability predicate Pr_{T}. GL can be given as an extension of the basic normal modal system K, where the distribution axiom $\Box(\alpha \to \beta) \to (\Box\alpha \to \Box\beta)$ interprets the second derivability condition on p.123, the Necessitation rule:

$$\frac{\vdash \alpha}{\vdash \Box\alpha}$$

interprets the first of these conditions and Löb's axiom $\Box(\Box\phi \to \phi) \to \Box\phi$ (which implies the third condition) internalizes Löb's theorem. This theory is incompatible with $\Box\phi \to \phi$ (note that $\Box(\Box\bot \to \bot)$ implies the provability of consistency $\Box\neg\Box\bot$). It is is (weakly) complete w.r.t. Kripke finite treelike models (hence is decidable), we are interested also in another kind of completeness: the arithmetical completeness, proved in Solovay (1976), of which Montagna (1979) gave a uniform version. Indeed, the most important results in this area are the so-called *fixed point theorem* (early results in the 70s as Bernardi (1975) or Sambin (1976)) and the fact that this logic was shown to be the logic of the formal proof predicate for many classical arithmetical theories: this is Solovay's arithmetical completeness theorem (1976). What is the problem with intuitionistic arithmetic? Recall that HA and PA are formulated in the same language, with the same mathematical axioms and share many properties: they have the same provably recursive functions and are proof-theoretically equivalent and are equi-consistent, i.e. if HA is consistent, then PA is also consistent. Moreover, each is capable of expressing its own proof predicate and by Gödel's results, if HA is consistent then neither HA nor PA can prove its own consistency (see e.g. Van Dalen (2001) and Avigad and Feferman (1998)). However, the provability logic of intuitionistic arithmetic goes beyond the intuitionistic version IGL of Löb's logic: Solovay's first

theorem does not hold in its immediate transposition, by replacing PA with HA and GL with IGL, this emerged clearly in the context of early algebraic researchs.

For reasons of space, we will not address here the thorny issue of Gentzen's proof systems for Provability Logic. In general, a Gentzen approach to modal systems is itself problematic, except for a few systems. Therefore, more complex formal systems, such as hypersequents, are often preferred. The cut elimination theorem for GL was first proved in Valentini (1983). There was a long controversy surrounding the validity of this theorem, where sequents were formalized as sets rather than multisets or sequences, however, improvements and corrections were then made in Sasaki (2001), in Goré and Ramanayake (2012), and, applying the hypersequent method, in Poggiolesi (2009). Many early investigations on provability logic, especially in Siena, where instead of algebraic character. We just quickly mention them, before reformulating these results in perhaps more understandable logical and proof-theoretical terms. Magari in the 70s introduced the equational class of *Diagonalizable algebras* DA, boolean algebras equipped with an additional unary operator $\tau(x)$ intended to "mimic" the provability predicate "x is provable", and fulfills conditions that reflect the Löb's conditions. Sambin and Ursini studied the *intuitionistic* version of these algebras, based on *Heyting algebras*, rather than boolean algebras. A primary example of diagonalizable algebra is the following. Let $\mathcal{M}_{\mathsf{PA}}$ be the *Lindenbaum algebra of Peano Arithmetic*, i.e. the well-known boolean algebra whose universe is the set of equivalence classes $[\phi]$ of sentences of this theory, modulo provable equivalence in PA:

$$\phi \sim_{\mathsf{PA}} \psi \text{ if and only if } \mathsf{PA} \vdash \phi \leftrightarrow \psi$$

equipped with a Boolean algebra structure, where the operations join $\sqcup$, meet $\sqcap$ and complementation $-$ are defined as usual by:

1. $[\phi] \sqcup [\psi] = [\phi \vee \psi]$
2. $[\phi] \sqcap [\psi] = [\phi \wedge \psi]$
3. $-[\phi] = [\neg\phi]$

where the square brackets denote the equivalence classes. This *is* a diagonalizable algebra, when enriched with an operator $\tau(x)$ defined as $\tau[\phi] = [\exists y \mathsf{Prf}_{\mathsf{PA}}(y, \ulcorner\phi\urcorner)]$. Solovay's celebrated theorem says that $\mathcal{M}_{\mathsf{PA}}$ is *functionally free* in the equational class DA of diagonalizable algebras. In other words, any identity true in $\mathcal{M}_{\mathsf{PA}}$ is true in *every* Diagonalizable algebras.

Since the beginning was raised the problem of extending this result to the intuitionistic case.

Problem. What, if in the above statement we replace PA with HA and diagonalizable algebras with *intuitionistic* diagonalizable algebra? Many problems arose immediately. Ursini in 1977 proved that $\mathcal{M}_{\mathsf{HA}}$ (the Lindenbaum algebra of HA) *is not functionally free* in the class of diagonalizable intuitionistic algebras (with Heyting's algebra in place of Boolean algebras): there are principles, as the law

$$\tau(x + y) \leq \tau(\tau(x) + \tau(y))$$

that holds in the diagonalizable Lindenbaum algebra $\mathcal{M}_{\mathsf{HA}}$, but not in all intuitionistic diagonalizable algebras. This is the problem on which Montagna's talk focused. The reasons why this happens were clearly highlighted by Montagna in the above mentioned conference:

> ... one of the reasons why this happens is that there are classically invalid rules that are valid in HA. By arithmetizing such rules, we obtain properties of the provability predicate for HA that are not deducible from the identities of intuitionist DA's.

We will better understand this problem after having reformulated Solovay's results in a general proof theoretical framework. From now on T will denote a "reasonable" (i.e. we assume that it is at least recursively enumerable and $\Sigma_1^0 - sound$) extension of $\mathsf{I}\Delta_0 + \mathsf{exp}$ (a formalization of the so-called *Elementary Arithmetic*, see at p. 192).

Definition 42. *An* arithmetical interpretation *in* T *is a function* $*$ *from the modal language to the arithmetical language that:*

1. *associates to each propositional variable p_i a formula p_i^* of the language of arithmetical theory* T.
2. *commutes with connectives,*
3. $\bot^* = (1 = 0)$.
4. $(\Box\psi)^* = \exists y \mathsf{Prf}_{\mathsf{T}}(y, \overline{\ulcorner \psi^* \urcorner})$.

A propositional formula ϕ is arithmetically valid in T *iff for all arithmetical interpretations in* T *as above, we have that ϕ^* is provable in* T.

The Provability Logic of T, that we denote PL_{T}, is the set of propositional modal formulas arithmetically valid in T. The most remarkable results in this area are the mentioned Solovay's arithmetical completeness theorems:

1. *Solovay's First theorem* (1976). $\mathsf{PL}_{\mathsf{T}} = \mathsf{GL}$
2. *Solovay's Second theorem* (1976). The set of modal propositional ϕ such that ϕ^* is *true* in the standard model of arithmetic, for all interpretations $*$ in T (here, for any *sound* extension T of $\mathsf{I}\Delta_0 + \mathsf{exp}$) are those provable in $\mathsf{GL}^- + \Box\phi \to \phi$, where GL^- is GL minus the necessitation rule.
3. *Montagna's uniform theorem* (1979). There is an intepretation $*$ in T such that for all ϕ,

$$\mathsf{T} \vdash \phi^* \text{ iff } \mathsf{GL} \vdash \phi$$

Other proofs of the uniform theorem were obtained independently using different methods in the same year or in the years immediately following by logicians as Artemov, Visser, Boolos, and Avron. See, for example, Boolos (2008) 132–136. However, Montagna's proof is the most original and has an algebraic flavour, as does Solovay's original proof. What is actually proved is the following equivalent statement: if T is a theory as above, and has infinite characteristic (see below for a definition of this concept), then the free Magari algebra of countably many generators is embeddable in the provability algebra of T. See Beklemishev and Flaminio (2016) for a detailed discussion.

The logic of provability GL actually has several semantics, including Kripke semantic and a fundamental arithmetic interpretation, which we will focus on in particular. With regard to Kripke semantics, we recall that a Kripke structure is a pair $\langle W, R\rangle$, where W is a nonempty set and R is a binary relation on it; an evaluation in a Kripkean structure is a function that associates each propositional variable p with a set $V(p) \subseteq W$ that can be seen as the set of states in which p is considered true. By a *model* we mean the triple $\mathcal{M} = \langle W, R, V\rangle$ on which we define a relation $x \Vdash_{\mathcal{M}} \phi$ to be read as: "ϕ is true in state x in model $\mathcal{M}$", inductively, as follows:

(a) $x \Vdash_{\mathcal{M}} p$ if and only if $x \in V(p)$

(b) $x \Vdash_{\mathcal{M}} \neg\phi$ if and only if it is not true that $x \Vdash \phi$

(c) $x \Vdash_{\mathcal{M}} \phi \wedge \psi$ if and only if $x \Vdash_{\mathcal{M}} \phi$ and $x \Vdash_{\mathcal{M}} \psi$

(d) $x \Vdash_{\mathcal{M}} \phi \vee \psi$ if and only if $x \Vdash_{\mathcal{M}} \phi$ or $x \Vdash_{\mathcal{M}} \psi$

(e) $x \Vdash_{\mathcal{M}} \phi \to \psi$ if and only if, if $x \Vdash_{\mathcal{M}} \phi$, then $x \Vdash_{\mathcal{M}} \psi$

(f) $x \Vdash_{\mathcal{M}} \Box\phi$ if and only if for all $y \in W$, if xRy, then $y \Vdash_{\mathcal{M}} \phi$.

Actually, a Kripke model for GL is such that R is a converse well-founded (hence irreflexive) strict partial ordering on W. We assume that every model has a *root*. A formula is valid in a model, if it is forced at the root. In Segerberg (1971) was demonstrated that GL is weakly complete with respect to the class of structures that are finite trees, namely that a formula is provable in GL if and only if it is valid in all finite treelike models, a property that we will use in Solovay's theorem[1].

Solovay's proof for classical theories. Solovay's technique consists in defining a function F, which constitutes an immersion of a Kripke model into arithmetic: the behaviour of this function is often described, at a heuristic level, as that of a refugee who moves from country to country, obtaining permission to cross the border on condition that he promises not to settle permanently in the country of arrival: if the refugee is not allowed to return twice to the same country, there must nevertheless be a country where he or she can settle permanently. Therefore, if the refugee is honest, he or she will never have to leave the country of origin. The story is told in Artemov and Beklemishev (2005), which we also follow for its version of the proof.

Solovay's function F (containing an obvious circularity) can be defined by appealing to the formalised principle of recursion (Kleene), and it is introduced in Solovay's original work, even if a series of simplifications can be applied to the original definition.

Suppose that $\mathcal{M} = \langle W, R, r, \Vdash \rangle$ is a finite model, where without loss of generality we assume that $W = \{1, 2, 3, ..., n\}$ and that $r = 1$, where R does not necessarily coincide with the natural order of $\mathbb{N}$; for purely technical reasons, another node 0 is generally added so that $0Ri$, for every $i \leq n$, without assuming anything about the forcing in it.

Solovay's (partially recursive) function $F : \mathbb{N} \longrightarrow W \cup \{0\}$ is informally defined as follows:

(a) $F(0) = 0$

(b) At step $x + 1$, suppose that $F(x)$ has already been defined: check whether $x + 1$ is the code for a proof of the fact that $\lim_{k \longrightarrow \infty} F(k) \neq z$, for some z accessible to $F(x)$: if YES, then $F(x + 1) = z$; if NO $F(x + 1) = F(x)$.

where $\lim_{k \to \infty} F(k) = z$ is an abbreviation for $\exists x \forall y > x \psi_F(x, y)$, and $\psi_F(x, y)$ is the $\Sigma_1^0 - graph$ of F.

Let us remember as we defined the hierarchy of theories $\mathsf{S}_0, \mathsf{S}_1, \mathsf{S}_3$... obtained by iterated addition of consistency statements of the previous chapter. The *characteristic* of $\mathsf{S} = \mathsf{S}_0$ is the last number n such that S_n is inconsistent. If such a number does not exist, we say that it has an *infinite characteristic*: for instance, if S is Σ_1-sound, then it has an *infinite characteristic*. We establish Solovay's result for axiomatizable extensions S of *Elementary Arithmetic* $\mathsf{I}\Delta_0 + \mathsf{exp}$ with infinite characteristic.

Let us abbreviate the expression "$\exists m \forall n > m (h(n) = z)$" as "$L_z$".

Lemma 32. *The following properties of the function F are provable in the Elementary Arithmetic $\mathsf{I}\Delta_0 + \mathsf{exp}$:*

(a) $\bigvee_{z \in W \cup \{0\}} L_{\overline{z}}$

(b) $\forall u \forall v (u \neq v \rightarrow \neg (L_u \wedge L_v))$

(c) $L_{\overline{m}} \wedge \overline{m} R \overline{s} \rightarrow \neg Pr(\overline{\ulcorner \neg L_{\overline{s}} \urcorner})$

(d) $L_{\overline{m}} \wedge \overline{m} \neq \overline{0} \rightarrow Pr(\overline{\ulcorner \bigvee_{mRw} L_{\overline{w}} \urcorner})$

[1] Recall that *strong* completeness does not hold, because this logic is not compact.

Proof. (see Artemov and Beklemishev (2005) 481-482). QED

The idea behind the proof in Solovay (1976) is to now provide an arithmetic simulation of Kripke's model; let us therefore define a *Solovay interpretation* $*$, for a model $\langle\{0,1,2,3...n\},1,\Vdash\rangle$, the one in which, in particular, propositional variables are interpreted as follows:

$$p^* = \bigvee\{L_x | x \Vdash p \ \ and \ \ 0 \leq x \leq n\}$$

where the statements L_x assume the role of the nodes x of the model, and where $*$, in our case, is *an arithmetical interpretation in the theory* S. The empty disjunction is $\overline{0} = \overline{1}$. The fundamental step in proving the arithmetical completeness theorem consists in demonstrating the *faithfulness* of the interpretation, i.e. this result:

Lemma 33. *For $1 \leq x \leq n$ and $*$ as above, the following apply:*

(a) If $x \Vdash \psi$ then $\mathsf{I\Delta_0} + \mathsf{exp} \vdash L_{\overline{x}} \rightarrow \psi^*$

(b) If $x \nVdash \psi$ then $\mathsf{I\Delta_0} + \mathsf{exp} \vdash L_{\overline{x}} \rightarrow \neg\psi^*$

Before proving this lemma, note that arithmetic completeness follows directly from it: if ϕ is not a theorem of GL, hence it is false at the root of a finite treelike model $W = \{1,2,3,...,n\}$. We add a new root 0 such that $0Rm$ for all $m \in W$ (forcing on 0 is arbitrarily defined). From the above lemma, if $1 \nVdash \phi$, then $\mathsf{I\Delta_0} + \mathsf{exp} \vdash L_{\overline{1}} \rightarrow \neg\phi^*$. Now, if $\mathsf{S} \vdash \phi^*$ we have $\mathsf{S} \vdash \neg L_{\overline{1}}$ and by the point c) of Lemma 32 we also obtain $\neg L_{\overline{0}}$. It follows that $L_{\overline{m}}$ must be true, for some $m \in W$. Observe that $m \Vdash \Box^{d(m)+1}\bot$, where $d(m) = sup\{d(s)+1 | mRs\}$ and $\Box^{d(m)+1}$ denotes a block of $d(m)+1-$ boxes. Hence $\mathsf{I\Delta_0} + \mathsf{exp} \vdash L_{\overline{m}} \rightarrow (\Box^{d(m)+1}\bot)^*$ and therefore that $(\Box^{d(m)+1}\bot)^*$ is true, against the hypothesis that S had *infinite* characteristic.

Let us now prove the lemma:

Proof. Induction on the complexity of ϕ. If $\phi = p$, the result follows from the definition; we leave the Boolean cases as an exercise and come to the fundamental case, where $\phi = \Box\psi$; note that $x \Vdash \Box\psi$ implies that for every y, if xRy, then $y \Vdash \psi$, from which for every y, if xRy then $\mathsf{I\Delta_0} + \mathsf{exp} \vdash L_{\overline{y}} \rightarrow \psi^*$ by inductive hypothesis and therefore:

$$\mathsf{I\Delta_0} + \mathsf{exp} \vdash \bigvee\{L_{\overline{y}} \rightarrow \psi^* | xRy\}$$

Hence $\mathsf{I\Delta_0} + \mathsf{exp} \vdash Pr(\overline{\ulcorner\bigvee\{L_{\overline{y}}|xRy\}\urcorner}) \rightarrow Pr(\overline{\ulcorner\psi^*\urcorner})$ by logic and by the derivability conditions. Lastly $\mathsf{I\Delta_0} + \mathsf{exp} \vdash L_{\overline{x}} \rightarrow Pr(\overline{\ulcorner\psi^*\urcorner})$, by the point d) of the previous lemma. If on the contrary $x \nVdash \Box\psi$, then there exists y such that xRy, but $y \nVdash \psi$It follows that exists y such that xRy and $\mathsf{I\Delta_0} + \mathsf{exp} \vdash L_{\overline{y}} \rightarrow \neg\psi^*$, or, equivalently, by contrnominal $\mathsf{I\Delta_0} + \mathsf{exp} \vdash \psi^* \rightarrow \neg L_{\overline{y}}$; hence, for some y, xRy and $\mathsf{I\Delta_0} + \mathsf{exp} \vdash Pr(\overline{\ulcorner\psi^*\urcorner}) \rightarrow Pr(\overline{\ulcorner\neg L_{\overline{y}}\urcorner}))$. Taking the contrapositive of this implication, from point c) of the previous lemma, it follows that:

$$\mathsf{I\Delta_0} + \mathsf{exp} \vdash L_{\overline{x}} \rightarrow \neg Pr(\overline{\ulcorner\psi^*\urcorner})$$

QED

This concludes the proof of Solovay's *first* theorem, and from it we draw the conclusion that GL axiomatises the modal axiom schemes *provable* in S: but what about the *true* schemes in the standard model? The answer is provided by the modal system GL^-, obtained by removing the necessitation rule from GL and adding the schema $\Box\alpha \rightarrow \alpha$. Note incidentally that if we did not remove the necessitation rule, from $\Box\bot \rightarrow \bot$, we could derive $\Box(\Box\bot \rightarrow \bot)$ and by Löb $\Box\bot$ and finally $\bot$. The *second Solovay theorem* follows from the fact that these propositions are equivalent:

(a) $\mathsf{GL}^{-} \vdash \alpha$

(b) $\mathsf{GL} \vdash \bigwedge_{\Box\beta \in Sub(\alpha)}(\Box\beta \to \beta) \to \alpha$, where $Sub(\alpha)$ is the set of subformulas of α.

(c) α is true at the root of every α-sound model (i.e., whose root forces $\Box\beta \to \beta$, for every $\Box\beta \in Sub(\alpha)$).

(d) $\mathbb{N} \models \alpha^*$, for every arithmetic interpretation $*$.

For further details and information, see Boolos (2008), ch.9, Artemov and Beklemishev (2005), ch.3 and De Jongh and Japaridze (1998), ch.3. Among the improvements that have been made we like to remember this: although the above proof employed the recursion theorem, in De Jongh, Jumelet and Montagna (1991) it is shown that the use of this theorem actually is not necessary. Using the recursion theorem makes the procedure more intuitive, but, according to these logicians, it "adds to the mystery of the proof". The alternative proof proposed is closer to the spirit of modal logic, replaces the recursion theorem by Gödel's diagonalization lemma and and is also applicable to another system of modal logic, yielding the arithmetical completeness of the so-called Rosser logic of Gauspari-Solovay with respect to extensions of the *Elementary Arithmetic*. It is still not know what is the provability logic of of the important weak fragment S_2^1, discussed in section 7.3.

We conclude the general introduction here and ask ourselves: what happen if we consider *intuitionistic* theories of arithmetic rather than classical theories? We like to point up that already at the end of the 60s, a particular kind of Solovay style theorem had been shown in De Jongh (1969). It was the first theorem of this kind and concerned the intuitionistic version HA of Peano Arithmetic. de Jongh's theorem provides an answer concerning the *ordinary* propositional logic of the intuitionistic arithmetical theory HA, i.e. the box-free part of the provability logic of HA, and it is the first kind of "arithmetical completeness" theorem discovered. We state here this theorem for IPC (the intuitionistic propositional calculus).

Theorem 88. (De Jongh 1970) *The following statements are equivalent, for all propositional formulas $\phi(p_0, ..., p_n)$:*

(a) $\mathsf{IPC} \vdash \phi(p_0, ..., p_n)$

(b) $\mathsf{HA} \vdash \phi(B_0, ..., B_n)$*, for all arithmetical sentences $B_0, ..., B_n$.*

A proof can be found in Smorynski (1973). In particular Since an arithmetical interpretation is in fact a substitution of the above kind, this means that $\phi(p_0, ..., p_n)$ is in the provability logic of HA. Refinements were given by several logicians. In particular Friedman (1973) proved that the choice of $B_0, ..., B_n$ can be actually made uniformly.

Theorem 89. (Friedman 1973) *There exists a computable sequence $B_0, ..., B_n$ of arithmetical sentences, such that for all $\phi(p_0, ..., p_n)$:*

$$\mathsf{IPC} \vdash \phi(p_0, ..., p_n) \;\; if \;\; only \; if \;\; \mathsf{HA} \vdash \phi(B_0, ..., B_n)$$

Further extensions of these results are obtained by adding to HA other principles e.g. accepted by russian constructivistis (see Trolestra and Van Dalen (1988)), but considered problematic by other constructivists. Tipically, the *Extended Church Thesis* ECT_0 and *Markov's principle* MP. Smorynski proved that the propositional logic of $\mathsf{HA} + \mathsf{MP}$ is still IPC, where MP is Markov's principle:

$$\forall x(A(x) \vee \neg A(x)) \wedge \neg\forall x \neg A(x) \to \exists x A(x)$$

Gavrilenko proved that if we replace MP with ECT_0 we still get IPC, where ECT_0 is the extended Church thesis:

$$\forall x(A(x) \to \exists y B(x, y)) \to \exists z \forall x(A(x) \to \exists u(T(z, x, u) \wedge B(x, U(u)))$$

where $T(z,x,u)$ is Kleene's predicate of recursion theory and $U(u)$ the result of computation u, and $A(x)$ does not contain $\vee$, and $\exists$ only in front. But it is not known what is the propositional logic of *Markov's Arithmetic*, i.e. $\mathsf{HA}+\mathsf{MP}+\mathsf{ECT}_0$. However it is known that it is *not* IPC. For instance, if $\alpha = \neg p \vee \neg q$, then the following:

$$((\neg\neg\alpha \to \alpha) \to (\neg\neg\alpha \vee \neg\alpha)) \to (\neg\neg\alpha \vee \neg\alpha)$$

is in the propositional provability logic of Markov's arithmetic, but is not a thoerem of IPC. Therefore we come to the intermediate logics. The interesting purpose of De Jongh, Verbrugge and Visser (2011) is instead that of strengthen the propositional logic, rather than the arithmetical theory, and then to consider a large class of intermediate logics. The intermediate (or *superintuitionistic*) logics are logics between IPC and CPC. The authors conjecture that, if L is an intermediate logic and HA_L is obtained by adding this logic to the intuitionistic arithmetic, then L is the propositional logic (in the sense of de Jongh's theorem) of HA_L. Actually they proved the conjecture only for logics satisfying the so called *finite frame property.*

Another line of research investigates the consequences of strengthening the *propositional logic* (while remaining in the constructive field), rather than the arithmetical theory. Coming back to the problem of *Provability Logic for Intuitionistic Arithmetic*, the problem is to find an intuitionistic modal logic I such that $\mathsf{I} = \mathsf{PL}_\mathsf{HA}$, namely, a modal formula ϕ is derivable in I if anf only if ϕ^* is derivable in HA, for all arithmetical interpretation $*$. We remark that Provability logics in general are *not monotone*, i.e. if a theory T extends a theory V, this *is not generally true* for the relative provability logics. For this reason Solovay's results in the classical case (e.g. the first theorem still hold for all Σ^0_1 and r.e. theories extending, or intepreting, *Elementary Arithmetic*) are very surprising: Solovay's theorems concerning the classical logic are very *stable*. On the contrary the situation for constructive arithmetical theories is different. Different constructive theories may have different logic. What principles belong to PL_HA? Since the standard proof predicate for HA and the relative Löb derivability conditions are derivable already in the intuitionistic version of *Elementary Arithmetic*, we conclude that the provability logic for HA, must *at least* contain the modal axioms of Löb's logic GL to IPC. The provability logic of Heyting arithmetic contains in general *principles that the provability logic of Peano Arithmetic does not share* and therefore is not a sublogic of that of Peano Arithmetic (non-monotonicity). For example, Leivant's principle (see below). On the other hand, there are classical provability principles that cannot be accepted by the intuitionists, e.g $\Box(p \vee \neg p)$. What do we know until today? Some examples:

(a) The formalization of the so-called *Markov's rule* (an intuitionistically *admissible* rule) is a principle of the Provability logic of HA:

$$\Box\neg\neg\Box\phi \to \Box\Box\phi$$

(b) Since, on the contrary, the formalization of the intuitionistic disjunction property:

$$\Box(\phi \vee \psi) \to (\Box\phi \vee \Box\psi)$$

is *not* provable in HA (H. Friedman), ones can add in its place a (provable in HA) weak form of this principle, due to Leivant:

$$\Box(\phi \vee \psi) \to \Box(\Box\phi \vee \psi)$$

(c) However Leivant's axiom is *not* a theorem of GL: in classical framework it allows to derive the iterated inconsistency statement $\Box\Box\bot$.

After Visser had found partial results concerning the letterless fragment (i.e. based only on constants $\bot, \top$), in particular Ardeshir and Mojtahedi (2014) provided an answer to the problem for the Σ^0_1 provabilty logic of HA, while Zoethout and Visser (2019) provided an alternative route to this problem.

Definition 43. *Let us call an arithmetical interpretation* $*$ *a* Σ^0_1*-interpretation, if and only if for all propositional variables* p_i*, we have that* p_i^* *is* Σ^0_1*, i.e. has the form* $\exists x\theta$*, where* θ *is atomic, or contains only connectives and bounded quantifiers.*

In 1990 Visser had previously introduced an algorithm to associate to propositional ϕ a particular form ϕ^+, called NNIL-form (No Nested Implications to the Left, as for instance $(p \to (q \to \bot)) \vee (q \to p)$). In some sense ϕ^+ is the "best approximation from below" of ϕ, in the sense of this theorem.

Theorem 90. $\mathsf{IPC} \vdash \phi^+ \to \phi$*. Moreover, if* $\mathsf{IPC} \vdash \alpha \to \phi$*, then also* $\mathsf{IPC} \vdash \alpha \to \phi^+$*.*

This class has been studied independently of our problem, due to its interesting properties in Visser, de Jongh, Van Benthem and Renardel de Lavalette (1995). This algorithm has been extended to modal formulasin Ardeshir and Mojtahedi (2014), that introduced the class TNNIL: "no $\to$ in the left side of an implication, except those in the scope of a $\Box$" (example, $\Box(p \to q) \to q$ is in this class, but $(p \to q) \to q$ is not). The result is the following:

Theorem 91. (Ardeshir and Mojtahedi 2014) *A modal formula* ϕ *is in the* Σ^0_1-*provability logic of* HA*, if and only if* $\mathsf{IGLC} \vdash \phi^+$*, where* IGLC *is the intuitionist version of the Löb logic* GL*, plus the completeness principle* $\psi \to \Box\psi$*.*

Other directions of the research involve the notion of relative intepretability; actually, all known principles of $\mathsf{PL_{HA}}$ are derivable in a bi-modal system of intepretability logic introduced by Albert Visser and studied in in particular in Iemhoff (2003). This logic is based on a binary modal operator $\alpha \rhd \beta$ (where $\Box\alpha = \top \rhd \alpha$) where an arithmetical interpretation $*$ of it is the formalization in the language of arithmetic of the fact that if $\mathsf{HA} \vdash \sigma \to \alpha^*$, then $\mathsf{HA} \vdash \sigma \to \beta^*$, for all σ belonging to Σ_1. Subsequent work by logicians such as Visser and de Jongh led to an axiomatisation of this modal logic named IPH and the demonstration of the arithmetical soundness of this axiomatisation. Iemhoff conjectured that it was also arithmetically complete, i.e. that $\mathsf{IPH} = \mathsf{PL_{HA}}$, but this is still open.

As far as we know, a definitive solution to the problem raised by Franco Montagna in the mentioned conference has finally been achieved recently, largely relying on the concepts and results we have briefly summarised. Mojtahedi's theorem is contained in Mojtahedi (2022). It is a not yet published work of frightening complexity. Axioms in particular are rather complex. Rougly speaking, $\mathsf{PL_{HA}}$ consists in IGL, i.e. Löb's logic of provability on an intuitionistic basis formulated in a non standard way, plus all formulas $\Box\alpha \to \Box\beta$ for α and β that satisfy a condition inspired by Iemhoff's definition of intuitionistic interpretability: $\mathsf{IGL} \vdash \sigma \to \alpha$, then $\mathsf{IGL} \vdash \sigma \to \beta$, for all σ belonging to a set of formulas that can be *projected to a* NNIL *formula*, a rather tecnical notion coming from the theory of intuitionistic unification studied in Ghilardi (1999).

5.4. First-order Provability Logic for classical arithmetic

After the results of arithmetic completeness of the logic of provability at the classical propositional level that we have briefly illustrated, it seemed natural to raise the question of the characterisation of the logic of *predicative* provability. We thought it is interesting to provide some information on a more advanced but not very well known topic, namely the first-order quantified version of the Provability Logic, which is in fact a first-order modal logic, subject around which there have been several philosophical controversies. The famous logician Ruth Barcan Marcus (1921-2012) published her first paper on Quantified Modal Logic (QML) in 1946; in 1947 she extended QML to the second order. In these works for the first time is investigated the relationship among modal operators $\Box, \Diamond$ ("is necessary", "is possible") and quantifiers $\forall, \exists$, and between *de re* modalities $\forall x \Box\phi$ ("for all x, necessarily ϕ") and *de dicto* modalities $\Box\forall x\phi$ ("is necessary that for all x, ϕ"). In she placed among the axioms of her system, the schema $\forall x \Box\psi(x) \to \Box\forall x\psi(x)$ that connect those modalities.

The main opponent of QML was Quine, that rejected it since, in his opinion, combination of quantifiers and modalities produced "unintelligible" results. Quine's objections are logical arguments based on failure of substitution: the sentence '8 is necessarily greater than 7' is true, while the sentence "the number of planets is necessarily greater than 7" is false: but the latter was obtained from the first by substitutipn of the coreferential term "the number of planets" in place of term '8'. It follows that such occurrences of singular terms are not "purely referential", and therefore quantification into modal context is unintelligible. Quine had also a metaphysical objection: quantification in modal contexts commits us to accepting essentialism, that he refused as indefensible. Leaving this debate in the background (and not addressing the issue of propositional quantifiers in Provability Logic), we simply highlight the role in this context of the debated Barcan Marcus axioms:

> The converse Barcan formula says that, as we move to an alternative situation, nothing passes out of existence. The Barcan formula says that, under the same circumstances, nothing comes into existence. The two toghether say the same things exist no matter what situation (Fitting and Mendelsohn (1998) 114).

The language of QGL, the first-order quantified version of Gödel-Löb logic, adds the symbol $\Box$ to first order logic without identity, constants or functional symbols. Rules and axioms are those of GL plus those of predicate calculus, for all QGL formulas.

Definition 44. *An arithmetical interpretation for the first-order case is given as follows:*

(a) *For all atomic formula $P(x_0, ..., x_n)$ of the language of* QGL*, the formula $(P(x_0, ..., x_n))^*$ is a formula of the language of arithmetic with the same free variable.*

(b) $(\phi \to \psi)^* = (\phi^* \to \psi^*)$ *(analogously for $\vee$ and $\wedge$).*

(c) $(\neg\phi)^* = \neg(\phi^*)$.

(d) $(\exists x\psi)^* = \exists x(\psi^*)$ *(analogous for $\forall$).*

(e) $(\Box\psi)^* = \exists y \mathsf{Prf}_{PA}(y, \overline{\ulcorner \psi^* \urcorner})$.

We say that a formula ψ of the language of QGL is *always provable*, iff for all arithmetical interpretations $*$, ψ^* is provable in PA. We say that a formula ψ of the language of QGL is always true, iff for all arithmetical interpretations $*$, ψ^* is true in the standard model $\mathbb{N}$. As far as propositional theories GL and GLS (the theory axiomatized by all theorems of GL and that replace the necessitation rule with the schema $\Box A \to A$) the situation is trublefree:

(a) GL axiomatizes the class of *always provable* sentences.

(b) GLS axiomatizes the class of *always true* sentences.

Both theories are decidables and therefore so are the above-mentioned classes. What is the status of the Barcan Marcus schema BS, $\forall x \Box\psi(x) \to \Box\forall x\psi(x)$? Actually this formula is not *always true* (neither *always provable*), but its converse CBS:

$$\exists x \Diamond \psi \to \Diamond \exists x \psi$$

is *always provable*. The following is a correct argument (see Smorynski (1987)). Recall that $RFN_{\Sigma^0_n}(\mathsf{T})$ is the following reflection schema:

$$\forall x_0, ..., \forall x_n (\mathsf{Pr}_{\mathsf{T}}(\overline{\ulcorner \phi(\dot{x_0}, ..., \dot{x_n}) \urcorner}) \to \phi(x_0, ..., x_n))$$

where $\phi \in \Sigma^0_n$ (analogously for $\phi \in \Pi^0_n$). For $n \geq 1$, we show that $RFN_{\Sigma^0_n}(\mathsf{T})$ is equivalent to $RFN_{\Pi^0_{n+1}}(\mathsf{T})$ by means of CBS: let $\phi(x,y) \in \Sigma^0_n$ and suppose that $\mathsf{Pr}_{\mathsf{T}}(\overline{\ulcorner \forall y \phi(\dot{x},y) \urcorner})$; hence, since we have CBS, it follows $\forall y \mathsf{Pr}_{\mathsf{T}}(\overline{\ulcorner \phi(\dot{x},\dot{y}) \urcorner})$. For $Rfn_{\Sigma^0_n}(\mathsf{T})$ and the predicate calculus we conclude $\forall y \phi(x,y)$.

$$\begin{aligned} \mathsf{QGL} \quad &\vdash \Box(\forall x \alpha(x) \to \alpha(u)) \\ &\vdash \forall u \Box(\forall x \alpha(x) \to \alpha(u)) \\ &\vdash \forall u(\Box \forall x \alpha(x) \to \Box \alpha(u)) \\ &\vdash (\Box \forall x \alpha(x) \to \forall u \Box \alpha(u)) \end{aligned} \tag{1}$$

Analogously $\mathsf{QGL} \vdash \exists u \Box \alpha(u) \to \Box \exists u \alpha(u)$. From these result it follows that a kind of formalized completeness $\alpha \to \Box \alpha$ holds in $\mathsf{QGL} + \mathsf{BS}$ for *any* quantified formula α.

Theorem 92. *The following are equivalent in* PA*:*

(a) $\neg Con(\mathsf{PA})$

(b) $\phi \to \mathsf{Pr}_{\mathsf{PA}}(\overline{\ulcorner \phi \urcorner})$*, for closed* ϕ*.*

(c) $\mathsf{Pr}_{\mathsf{PA}}(\overline{\ulcorner \phi \urcorner}) \vee \neg \mathsf{Pr}_{\mathsf{PA}}(\overline{\ulcorner \phi \urcorner})$

(d) $\forall x \mathsf{Pr}_{\mathsf{PA}}(\overline{\ulcorner \phi(\dot{x}) \urcorner}) \to \mathsf{Pr}_{\mathsf{PA}}(\overline{\ulcorner \forall x \phi(x) \urcorner})$

where $Con(\mathsf{PA})$ stands as usual for $\neg \exists y \mathsf{Prf}_{\mathsf{PA}}(y, \overline{\ulcorner 1 = 0 \urcorner})$. The (BS) schema is in this context known as "$\omega-$completeness schema". Recall that according to the Second Gödel's theorem, under the assumption of consistency of the theory, both, $Con(\mathsf{PA})$ and $\neg Con(\mathsf{PA})$ are unprovable. Ergo: none of the above principles is provable! For example, let us see that $\mathsf{PA} + (\mathsf{BS})^* \vdash \neg Con(\mathsf{PA})$. But we have also, for all ψ, $\mathsf{PA} \vdash \mathsf{Pr}_{\mathsf{PA}}(\overline{\ulcorner \mathsf{Prf}_{\mathsf{PA}}(\dot{y}, \overline{\ulcorner \psi(\dot{x}) \urcorner}) \to \psi(\dot{x}) \urcorner})$, from which follows in particular:

$$\mathsf{PA} \vdash \mathsf{Pr}_{\mathsf{PA}}(\overline{\ulcorner \neg(1=0) \to \neg \mathsf{Prf}_{PA}(\dot{y}, \overline{\overline{\ulcorner 1 = 0 \urcorner}}) \urcorner})$$

Since $\mathsf{PA} \vdash \mathsf{Pr}_{\mathsf{PA}}(\overline{\ulcorner \neg(1=0) \urcorner})$, by Löb conditions and manipulation of quantifiers we get $PA \vdash \forall y \mathsf{Pr}_{\mathsf{PA}}(\overline{\ulcorner \neg \mathsf{Prf}_{\mathsf{PA}}(\dot{y}, \overline{\ulcorner 1 = 0 \urcorner}) \urcorner})$. If $(\mathsf{BS})^*$ holds, it follows:

$$\mathsf{PA} \vdash \mathsf{Pr}_{\mathsf{PA}}(\overline{\ulcorner \forall y \neg \mathsf{Prf}_{PA}(y, \overline{\ulcorner \neg(1=0) \urcorner}) \urcorner})$$

Hence $\mathsf{PA} \vdash \mathsf{Pr}_{\mathsf{PA}}(\overline{\ulcorner \exists y \mathsf{Prf}_{\mathsf{PA}}(y, \overline{\ulcorner (1=0) \urcorner}) \to 1 = 0 \urcorner})$, in a few steps, from which $\mathsf{PA} \vdash \mathsf{Pr}_{\mathsf{PA}}(\ulcorner 1 = 0 \urcorner)$, namely $\neg Con(\mathsf{PA})$, by the formalized Löb theorem, according to which, for all ψ:

$$\mathsf{PA} \vdash \mathsf{Pr}_{\mathsf{PA}}(\overline{\ulcorner \mathsf{Pr}_{\mathsf{PA}}(\overline{\ulcorner \psi \urcorner}) \to \psi \urcorner}) \to \mathsf{Pr}_{\mathsf{PA}}(\overline{\ulcorner \psi \urcorner})$$

It is now interesting to understand if the collection of always true or the always provable sentences can be characterized axiomatically. If we denote $\mathsf{QPL}_{\mathsf{PA}}$ the first-order provability logic of PA, we will see that actually $\mathsf{QGL} \subset \mathsf{QPL}_{\mathsf{PA}}$. From Vardanyan's theorem it follows that cannot have not even a recursive axiomatization.

Theorem 93. *The following hold:*

(a) (Vardanyan 1986) *The class of all predicate modal formulas always provable is* $\Pi^0_2 -$ *complete.*

(b) (Boolos and McGee 1987) *The class of all sentences of predicate modal formulas always true is* $\Pi^0_1 - complete\ in\ Th(\mathbb{N})$.

Proof. (See Boolos (2008) pp. 233-41) QED

Contrast with the propositional case: GL axiomatizes the class of always provable sentencese of propositional Provability Logic; GLS axiomatizes the class of always true sentences. Such classes are *decidable*. At the opposite, in respect to QGL, these classes are *not even recursively enumerable*; since the collection of theorems of an axiomatizable theory is r.e. (namely Σ^0_1), it follows that the class of always provable formulas of QGL *is not axiomatizable*.

Now observe that if X is an arithmetically definable set, then $X \leq_T Th(\mathbb{N})$: indeed, if some $\phi \in \Sigma^0_n$ defines X, then $n \in X$ iff $\ulcorner \phi(\overline{n}) \urcorner \in Th(\mathbb{N})$; hence $X \leq_m Th(\mathbb{N})$ (and *a fortiori* Turing reducible) via the funtion $n \xrightarrow{\phi(\dot{x})} \ulcorner \phi(\overline{n}) \urcorner$. But if X is Π^0_1-complete in $Th(\mathbb{N})$, then X is not Σ^0_1 in $Th(\mathbb{N})$, namely is not r.e. in $Th(\mathbb{N})$ and a fortiori, not recursive in it. Since by Boolos and McGee's result the class of *always true* sentences of QGL is Π^0_1-complete in $Th(\mathbb{N})$, it follows that such a class is not arithmetically definable.

5.5. Semantic for first-order Modal Logic

We start by introducing a Kripke-style semantic for first-order Modal Logic.

Definition 45. *A constant domain Kripke frame is a structure* $\mathcal{M} = \langle W, R, D, \{V_w\}_{w \in W} \rangle$ *where for all* $w \in W$, $V_w(P(x_1, ..., x_n))$ *is an* $n-$*ary relation on* D *and if* σ *is an assignement in* D *to the variables and we denote* $\sigma(a/x)$ *its* $x-$*th variant, the forcing relation is given by the following:*

(a) $w \Vdash^{\mathcal{M}}_{\sigma} P(x_1, ..., x_n)$ *iff* $\langle \sigma(x_1), ..., \sigma(x_n) \rangle \in V_w(P(x_1, ..., x_n))$.

(b) $w \Vdash^{\mathcal{M}}_{\sigma} x = y$ *iff* $\sigma(x) = \sigma(y)$

(c) analogous to the propositional case, for connectives.

(d) $w \Vdash^{\mathcal{M}}_{\sigma} \Box\phi$ *for all* v, *if* wRv, *then* $v \Vdash^{\mathcal{M}}_{\sigma} \phi$.

(e) $w \Vdash^{\mathcal{M}}_{\sigma} \exists x\phi$ *iff for some* $x-$*variant* $\sigma(a/x)$, $w \Vdash^{\mathcal{M}}_{\sigma(a/x)} \phi$.

A frame with *monotone variable domain* is obtained by adding to the frame $\langle W, R \rangle$ a collection of sets $\{D_w\}_{w \in W}$ where for all $w \in W$, it holds that if wRv, then $D_w \subseteq D_v$ (we call it *antimonotone* if at the opposite $D_w \supseteq D_v$).

A counterexample to the vailidity of (BS) in frames with variable monotone domain is easily given , taking a model where $W = \{w, v\}$, $R = \{\langle w, v \rangle\}$, $D_w = \{a\}$, $D_v = \{a, b\}$, defining $V_v(P(x)) = \{b\}$ and by considering an assignement σ and its $x-$variant $\sigma(b/x)$. Actually (BS) and (CBS) are both valid in a semantic of *costant domains*, but as long as variable domains are considered, (BS) holds in a variable domain frame, iff such a frame satisfies *antimonotonicity*, and (CBS) holds in a variable domain frame, iff it satisies *monotonicity*:

Let us call QK he system given adding to the First Order classical logic with identity these axioms:

(a) $\Box(\phi \to \psi) \to (\Box\phi \to \Box\psi)$

(b) Necessitation rule $\phi / \Box\phi$

(c) $x \neq y \to \Box(x \neq y)$

(d) Barcan Marcus schema.

A completeness result holds for this semantic: Every valid formula in constant domains frames is a theorem of QK (and *viceversa*). The system obtained removing (BS) is complete with respect to the increasing domain semantic: note that in this system the converse of Barcan Marcus schema is derivable.

Forcing $w \Vdash A$ is defined on closed formulas α with parameters in D_w. A Kripke model $\Re = \langle W, R, \{D_x\}_{x \in W}, \Vdash \rangle$ for QGL satisfies the following conditions:

(a) $W \neq \emptyset$

(b) If xRy, then $D_x \subseteq D_y$

(c) The forcing $\Vdash$ is defined as in the propositional case, but with a further clause:

$$x \Vdash \exists u \alpha(u) \text{ iff there exists } a \in D_x, x \Vdash \alpha(a)$$

Theorem 94. QGL *is complete with respect to the class of models transitive and such that for any closed formula* α*, if* $x \in W$ *and* $x \Vdash \alpha$*, then there exists* $y \in W$ *such that* $y \Vdash \alpha$*, and if* yRz*, then* $z \Vdash \neg\alpha$*.*

Recall that a theory S is *valid within a class of frames* C, if for all frames $\Re \in C$, $\Re \models \phi$ (namely if for all models $\mathcal{M}$ based on that structure, $\mathcal{M} \models \phi$), for all theorems ϕ of S; a theory is (weakly) complete with respect to a class of frames C, iff when $\Re \models \phi$, for all $\Re \in C$, then ϕ is a theorem of S.

S is *strongly complete* with respect to C, if $\Gamma \models_C \phi$ implies $\Gamma \vdash_S \phi$. Strong completeness does not hold for GL because compactness fails: one can device an infinite set of formulas Δ that is not satisfable, but such that each of its finite subset is satisfable. Hence, if Γ is a finite subset of Δ, then $\Gamma \nvdash \bot$, and a fortiori $\Delta \nvdash \bot$, since otherwise $\bot$ would be provable from some finite subset of Δ. However, being Δ unsatisfable, we have also $\Delta \models \bot$. But QGL is incomplete with respect to *any class of frame*. Indeed, this is the situation:

(a) QGL is *valid in a frame*, iff R is transitive and R^{-1} is well-founded, but:

(b) QGL is *not complete* with respect to this class of frames. It follows that:

(c) QGL is not complete with respect to *any class of frames*

For instance, the sentence:

$$\neg(\exists u \Diamond P(u) \wedge \forall v \exists w \Box(P(v) \rightarrow \Diamond P(w)))$$

is valid in all transitive and conversely well founded frames, but is not a theorem of QGL.

We will now illustrate the proof of two negative results namely that due to Artemov and that due to Montagna and mentioned above, respectively. This proof of the arithmetical incompleteness of QGL was provided by Montagna (1984).

Theorem 95. QGL *is not arithmetically complete w.r.t.* PA*.*

Proof. Let T a finitely axiomatizable and consistent theory (e.g. the theory NBG) such that $\mathsf{T} \vdash Con(\mathsf{T}) \rightarrow Con(\mathsf{PA} + \mathsf{Con}(\mathsf{PA}))$. Let $\bigwedge \mathsf{T}$ be the conjunction of all axioms of T and let us define $\alpha = \Diamond \bigwedge \mathsf{T} \rightarrow \Diamond\Diamond\top$.

We claim that α is valid in PA, but not provable in QGL. Indeed, let us suppose that $*$ is an arithmetical interpretation and let us consider $\bigwedge \mathsf{T}^*$ (assume w.l.o.g. that QGL contains the language of T). Hence α^* is provably equivalent to the sentence $Con(\mathsf{PA} + \bigwedge \mathsf{T}^*) \rightarrow$

$Con(\mathsf{PA} + \mathsf{Con}(\mathsf{PA}))$ and is provable in PA: let $B_1, ..., B_n$ be a proof of B_n in T. Hence $B_1^*, ..., B_n^*$ is proof of B_n^* in T^* and by a formalization of this, for all C not containing $\Box$, we have $\mathsf{PA} \vdash \mathsf{Pr}_{\mathsf{T}}(\ulcorner C \urcorner) \rightarrow \mathsf{Pr}_{\mathsf{T}^*}(\ulcorner C^* \urcorner)$ from which follows $\mathsf{PA} \vdash Con(\mathsf{T}^*) \rightarrow Con(\mathsf{T})$ and finally (since $\mathsf{PA} \vdash Con(\mathsf{PA} + \bigwedge \mathsf{T}^*) \rightarrow Con(\mathsf{T}^*)$), we obtain $Con(\mathsf{PA} + Con(\mathsf{PA}))$. Hence α is PA-valid. However it is not provable in GL: let $\Re$ be a model of T and let us consider the model $\Im = \langle W, R, \{D_x\}_{x \in W}, \Vdash \rangle$, where:

(a) $W = \{0, 1\}$

(b) xRy iff $x = 0$ and $y = 1$

(c) $D_0 = D_1 = \mathbb{N}$

(d) $i \Vdash B(a_0, ..., a_n)$ iff $\Re \vDash B(a_0, ..., a_n)$, for $a_0, ..., a_n \in \mathbb{N}$ and $B(x_0, ..., x_n)$ atomic formula of the language of T.

Since $1 \Vdash \bigwedge \mathsf{T}$ and $0R1$, we have $0 \Vdash \Diamond \bigwedge \mathsf{T}$; but $0 \Vdash \Box\Box\bot$, and therefore $0 \Vdash \neg\alpha$. Clearly R is transitive and conversely well founded, so that $\Im \models \mathsf{GL}$.

QED

Another strong negative results is Artemov's theorem. Recall that according to *Tarski-Post* theorem, no single formula Σ_n^0 of the Arithmetical Hierarchy can define $Th(\mathbb{N})$. We have seen that from a computational point of view, $Th(\mathbb{N}) \equiv_T \emptyset^\omega$. Moreover, according to *Tennenbaum's theorem* $+, \times, <$ are not recursive in (countable) non-standard model of PA, and many subtheories of it.

Theorem 96. (Artemov 1985) *Let* T *be an r.e. theory. Then, for all choice of a proof predicate* $\mathsf{Pr}_{\mathsf{T}}(x)$*, the set of predicate modal formulas always true is not arithmetical.*

Proof. (see De Jongh and Japaridze (1998)) Let us consider a purely relational arithmetical language with three predicates $E(x, y), A(x, y, z)$ and $M(x, y, z)$, to be interpreted in standard model respectively as $x = y, x + y = z, x \times y = z$. Let us consider the following version of Tennenbaum's theorem: there exists an arithmetical sentence θ such that:

(a) θ is true in the standard model.

(b) Every countable model of θ where $E(x, y)$ is identity and $A(x, y, z)$, $M(x, y, z)$ are recursive, is isomorphic to the standard model.

The second point actually implies that every countable model where $E(x, y)$ is identity and $A(x, y, z), M(x, y, z)$ are recursive, is *elementary equivalent* to the standard model. Let now C be the conjunction of the following sentences:

(a) $\forall x \forall y(\Box E(x, y) \vee \Box\neg E(x, y))$

(b) $\forall x \forall y \forall z(\Box A(x, y, z) \vee \Box\neg A(x, y, z))$

(c) $\forall x \forall y \forall z(\Box M(x, y, z) \vee \Box\neg M(x, y, z))$

Claim For any arithmetical formula ϕ, ϕ is true iff for all arithmetical interpretations $*$, $((\theta \wedge C) \rightarrow \phi)^*$ is true.

$\Rightarrow$ Actually if ϕ is true, $*$ is an interpretation and $\theta^* \wedge C^*$ is true and T is r.e., the truth of C^* implies that in the standard model E^*, A^*, M^* are recursive. Hence we define a countable model $\Re$ such that:

(a) $\Re \models E(k, m)$ iff $E^*(k, m)$ is true.

(b) $\Re \models A(k,m,n)$ iff $A^*(k,m,n)$ is true.

(c) $\Re \models M(k,m,n)$ iff $M^*(k,m,n)$ is true.

Hence we have in general that $\Re \models \phi$ iff ϕ^* is true, and in particular this holds for θ. Since by hyp. θ^* is true, also $\Re \models \theta$, and by our version of Tennenbaum's theorem we conclude that $\Re \equiv \mathbb{N}$. But by hyp. ϕ is true; it follows that $\Re \models \phi$ and once more ϕ^*. $\Leftarrow$ Let ϕ be false and let $*$ the trivial interpretation $E^* = E, A^* = A, M^* = M$. Hence $\theta^* = \theta$ and $\phi^* = \phi$. We have to check that $\theta \wedge C^* \rightarrow \phi$ is false, namely that $\theta \wedge C^*$ is *true* (since ϕ is false by hypothesis). But θ is true by hyp. and from decidability in T of $x = y, x + y = z, x \times y = z$ it follows that C^* is true. It follows that the set V of *always true* formulas of QGL is not aritmetical. Indeed, we know that $Th(\mathbb{N})$ is not aritmetical and the previous lemma provide an m-reduction of $Th(\mathbb{N})$ to V.

Let $f(\ulcorner \phi \urcorner) = \ulcorner (\theta \wedge C) \rightarrow \phi \urcorner$; therefore $\ulcorner \phi \urcorner \in Th(\mathbb{N})$ iff $f(\ulcorner \phi \urcorner) \in V$, namely $Th(\mathbb{N}) \leq_m V$. But if $X \leq_m V$ and $V \in \Sigma^0_n$, also $X \in \Sigma^0_n$ (contradiction). QED

As alreday mentioned in the former section, a basic result of propositional Provability Logic GL is the Sambin-de Jongh *fixed point theorem* and we wonder whether an analogous result holds in the predicative case. In other words, we ask whether is possible to prove in QGL the "Diagonalization theorem" of Gödel-Carnap. Let us consider the language of QGL extended with a countable amount of variable for formulas $p_0, p_1, p_2...$ and the axioms extended to this language. The question is the following:

> Let $\alpha(p_i)$ be a modalized formula of the above language. It is asked: is there a formula β of the original language, with individual variables identical to those of $\alpha(p_i)$, such that $\mathsf{QGL} \vdash \beta \leftrightarrow \alpha(\beta)$?

The answer is *no*.

Theorem 97. *No provable fixed point exists in* QGL *to the formula:*

$$\forall u \exists v \Box (p_i \rightarrow A(u,v))$$

Proof. Let us consider the Kripke model $\Re = \langle \mathbb{N}, R, \{D_w\}_{w \in \mathbb{N}}, \Vdash \rangle$ where:

(a) xRy iff $y < x$

(b) $D_w = \{y \in \mathbb{N} | w \leq y\}$

(c) If B is a predicate letter:

i. If B is not A, then $w \Vdash B(a_0,..,a_n)$, for all $a_0,..,a_n \in D_w$ and $w \in \mathbb{N}$.

ii. If B is A, then $w \Vdash A(a,b)$ iff either $b = w+1$ and $a \neq w+1$, or $a < b$ and $a, b \neq w+1$ $(a, b \in D_w)$.

The order induced by $A(a,b)$ is the following:

$$\begin{array}{lll} \vdots & \vdots & \vdots \\ 2 & D_2 & 2,3,4...3 \\ \downarrow & \cap & \\ 1 & D_1 & 1,2,3...2 \\ \downarrow & \cap & \\ 0 & D_0 & 0,1,2...1 \end{array} \tag{2}$$

Since R is transitive and R^{-1} well founded, this is a QGL model. Now let us observe that the forcing $\Vdash$ is definable in the structure $\langle \mathbb{N}, <, = \rangle$, namely for any formula $B(v_0, v_2, v_4 ... v_{2n})$, there exists a formula $B^*(v_1, v_0, v_2, v_4 ... v_{2n})$ such that for all $a_0, .., a_n \in D_w$ and $w \in \mathbb{N}$:

$$w \Vdash B(a_0, a_2, a_4 ... a_{2n}) \; iff \; \langle \mathbb{N}, <, = \rangle \models B^*(a_1, a_0, a_2, a_4 ... a_{2n})$$

Let us define B^* by induction:

(a) For $B(v_0, v_2, v_4 ... v_{2n})$ atomic, different to A, let:

$$B^*(v_1, v_0, v_2, v_4 ... v_{2n}) = \bigwedge_{i \leq n} v_1 \leq v_{2i}$$

(b) For $B(v_0, v_2, v_4 ... v_{2n}) = A(v_0, v_2)$, let:

$$B^*(v_1, v_0, v_2, v_4 ... v_{2n}) = A^*(v_1, v_0, v_2) =$$

$$v_1 \leq v_0 \wedge v_1 \leq v_2 \wedge (v_2 = v_1 + 1 \wedge \neg(v_0 = v_1 + 1)) \vee$$
$$\vee (\neg(v_2 = v_1 + 1) \wedge \neg(v_0 = v_1 + 1) \wedge v_0 < v_2)$$

(c) for $B(v_0, v_2, v_4 ... v_{2n}) = \neg C(v_0, v_2, ..., v_{2n})$, let:

$$B^*(v_1, v_0, v_2, v_4 ... v_{2n}) = \bigwedge_{i \leq n} v_1 \leq v_{2i} \wedge \neg(C^*)(v_1, v_0, v_2, ..., v_{2n})$$

(d) $*$ commutes with $\exists$ and $\vee$.

(e) for $B(v_0, v_2, v_4 ... v_{2n}) = \Box C(v_0, v_2, ..., v_{2n})$, let:

$$B^*(v_1, v_0, v_2, v_4 ... v_{2n}) = \forall w < v_1 (C^*)(v_1, v_0, v_2, ..., v_{2n})$$

where w is an odd variable not occurring in C^*.

It follows that $\{x \in \mathbb{N} | x \Vdash B\}$ is definable in $\langle \mathbb{N}, <, = \rangle$; but it is well known that every set definable in such a structure is *finite or cofinite* (hence, for instance, neither $2\mathbb{N}$, nor $(2\mathbb{N} + 1)$ is definable). Let us suppose now that:

$$\mathsf{QGL} \vdash B \leftrightarrow \forall u \exists v \Box (B \rightarrow A(u, v))$$

hence, since $\Re$ is a model of QGL we will have that for all $x \in \mathbb{N}$, $x \Vdash B \leftrightarrow \forall u \exists v \Box (B \rightarrow A(u, v))$. Since $0 \Vdash \Box \bot$, also $0 \Vdash B$. Moreover, if $v \in D_1$, $0 \Vdash \neg A(1, v)$ and therefore $0 \Vdash B \wedge \neg A(1, v)$, from which follows $1 \Vdash \exists u \forall v \Diamond (B \wedge \neg A(u, v))$ and $1 \Vdash \neg B$. Note that $0 \Vdash B$ *and* $1 \Vdash \neg B$. We want generalize this and we make an inductive argument by considering this result as the basis of an induction and we claim that B is forced at *even* nodes, whereas $\neg B$ is forced at *odd* nodes.Now let us suppose that for all $i \leq n, 2i \Vdash B$ and $2i + 1 \Vdash \neg B$. Hence $2n + 1 \Vdash \neg B$; moreover if $j \leq 2n$, $j \Vdash A(u, u + 1)$, for $u \in D_{2n+2} = 2n + 2, 2n + 3.....$; so, for every $u \in D_{2n+2}$ there is a $v \in D_{2n+2}$ such that if $(2n + 2)Rj$, then $j \Vdash \neg B$ or $j \Vdash A(u, v)$. It follows that $2n + 2 \Vdash \forall u \exists v \Box (B \rightarrow A(u, v))$. Hence $2n + 2 \Vdash B$; moreover, if $v \in D_{2n+3}$ then $2n + 2 \Vdash B \wedge \neg A(2n + 3, v)$, from which follows $2n + 3 \Vdash \neg \forall u \exists v \Box (B \rightarrow A(u, v))$ that implies $2n + 3 \Vdash \neg B$. This concludes the induction step and proves the claim, namely that $\{x \in \mathbb{N} | x \Vdash B\}$ coincides with the set of even numbers: *but this is not finite, neither cofinite* (contradiction).

QED

Which steps of the Sambin-de Jongh theorem fails in predicative case? The fixed point theorem for GL is based on the following substitution lemma:

Let $A(p), B, C, D$ formulas of GL. If $\mathsf{GL} \vdash D \to (B \leftrightarrow C)$, then:

(a) $\mathsf{GL} \vdash D \wedge \Box D \to (A(B) \leftrightarrow A(C))$

(b) Moreover, if p is modalized on A, then it holds that $\mathsf{GL} \vdash \Box D \to (A(B) \leftrightarrow A(C))$

This *does not extend to* QGL. However a weaker version is provable, sufficient to prove *uniqueness* of fixed point: for all modalized $A(p)$ and all B, C, if $\mathsf{QGL} \vdash A(B) \leftrightarrow B$ and $\mathsf{QGL} \vdash A(C) \leftrightarrow C$, then $\mathsf{QGL} \vdash B \leftrightarrow C$.

Question What happens if we add the Barcan Marcus schema? If we add the Barcan Marcus schema, actually some counterexamples to the fixed point theorem in QGL, fail to be counterexamples. For instance, if $A(p)$ is $\forall x \exists y \Box(p \to A(x, y))$ and D is $\forall u \exists v \Box A(u, v)$, then $\mathsf{QGL} + \mathsf{BS} \vdash D \leftrightarrow A(D)$. On the other hand, modifying Montagna's proof (and considering Kripke models with fixed domain), it is possible to show there are other formulas, as for example $\forall u(\Box\Box P(u) \to \Box(p \to P(u))$, that do not have a fixed point, not even in $\mathsf{QGL} + \mathsf{BS}$. Moreover this is also an arithmetical counterexample, in the sense that there is no sentence of the language of QGL, such that for all arithmetical interpretations $*$, $\mathsf{PA} \vdash D^* \leftrightarrow A(D)^*$.

We conclude by reporting some important recent developments and pointing to recent lines of research. The above mentioned Vardanyan's Theorem, i.e. the result which establishes the Π_2-completeness of the quantified modal logic of PA, is a benchmark and a barrier in this field of research, excluding the possibility of recursive axiomatisation. However, to paraphrase Visser and De Jonge (2006), we can say that we can glimpse an "escape" from it, although precisely this work dramatically generalises this result, leading to the conclusion that it is impossible to recursively axiomatise the quantified modal logic of a vast class of arithmetic theories. In fact, despite this negative result, some *positive* results still hold: on the one hand the result according to which the existence of a Kripke countermodel implies arithmetic nonvalidity, can be actually extended to the predicate level (see De Jongh and Japaridze (1998), 533-38); on the other hand Artemov and Japaridze (1990) proved that arithmetical completeness still holds, if we restrict ourselves to formulas with just one individual variable. Opening up a further glimmer of light in this gloomy outlook, in Yavorski (2002), Hao and Tourlakis (2021) and De Almeida Borges and Joosten (2013), some positive results are achieved. In the last of these works, in particular, a strictly positive first order modal calculus, named $\mathsf{QRC_1}$ (*Quantified Reflection Calculus with one modality*), is introduced, whose signature is composed by relational symbols and constants, among which $\top$, and logical operators are restricted to $\wedge$, $\forall$ and $\Diamond$. This logic is based on very basic rules that manipulates judgements of the form $\phi \vdash \psi$:

(a) $\phi \vdash \top$ and $\phi \vdash \phi$

(b) $\phi \wedge \psi \vdash \phi$ and $\phi \wedge \psi \vdash \psi$

(c) $\dfrac{\phi \vdash \psi \quad \phi \vdash \chi}{\phi \vdash \psi \wedge \chi}$

(d) $\dfrac{\phi \vdash \psi \quad \psi \vdash \chi}{\phi \vdash \chi}$

(e) $\dfrac{\phi \vdash \psi}{\Diamond\phi \vdash \Diamond\psi}$

(f) $\Diamond\Diamond\phi \vdash \Diamond\phi$

(g) $\dfrac{\phi \vdash \psi}{\phi[t/x] \vdash \psi[t/x]}$

(h) $\dfrac{\phi[c/x] \vdash \psi[c/x]}{\phi \vdash \psi}$

(i) $\dfrac{\phi \vdash \psi}{\phi \vdash \forall x\psi}$ (x not free in ϕ)

(j) $\dfrac{\phi[t/x] \vdash \psi}{\phi \vdash \forall x\psi}$ (t free for x)

A modal completeness (and decidability) was shown with respect to finite irreflexive and constant domain Kripke models. A theorem of arithmetical completeness is also demonstrated, providing a more complex notion of arithmetical interpretation and we would like to highlight the peculiar aspects of this interpretation. First a provability predicate *à la* Feferman Prf_τ is considered (see on p.122) for a Σ_1-definition of axioms τ of the theory T. The basic idea

is that relational symbols $S(x,c)$ of the modal language are intepreted as Σ_1 arithmetical formulas $S(x,c)^* = \sigma(u, v_x, v_c)$ that define set of axioms of theories and starting from such an intepretation $*$, then this is extended to an intepretation τ^* of all formulas as follows, where $\tau(u)$ is added to the definition of the axioms, to ensure that we are dealing with extensions of T:

(a) $\top^{\tau^*} = \tau(u)$.

(b) $S(x,c)^{\tau^*} = S(x,c)^* \vee \tau(u)$.

(c) $(\psi \wedge \delta)^{\tau^*} = \psi^{\tau^*} \vee \delta^{\tau^*}$.

(d) $\Diamond\psi^{\tau^*} = \tau(u) \vee (u = \ulcorner Cons_{\psi^{\tau^*}} \urcorner)$

(e) $(\forall x\psi)^{\tau^*} = \exists y \psi^{\tau^*}$

where conjunction is intepreted as the union of corresponding sets of axioms and the universal quantifier as an infinite union (see Beklemishev (2014)). For example:

$$(\Diamond S(s,c))^{\tau^*} = \tau(u) \vee (u = \ulcorner Cons_{\sigma(u,v_x,v_c)\vee\tau(u)} \urcorner)$$

A highly sophisticated proof in Solovay's style concludes that QRC_1 coincides with the set of judgements $\phi(x,c) \vdash \psi(x,c)$ such that for all arithmetical interpretations $*$ the theory T proves that for all sentences θ, the formula $\forall x \forall y(\mathsf{Pr}_{\psi^{\tau^*}}(\overline{\ulcorner\theta\urcorner}) \rightarrow \mathsf{Pr}_{\phi^{\tau^*}}(\overline{\ulcorner\theta\urcorner}))$ holds true, where ψ^{τ^*} and ϕ^{τ^*} generally depend on y, z, for all recursively enumerable and sound T extending $I\Sigma_1$. To close the circle with the previous section, the logic QRC_1 turns out to be arithmetically sound even with respect to HA, but the arithmetical completeness problem is still open.

5.6. The quest for consistency proofs: proposal for further study

By Gödel's theorem, a proof of consistency of arithmetic must necessarily go beyond the means of arithmetic itself, since no consistency proof of a sufficiently strong consistent arithmetical theory can use methods that can be formalized in the theory itself. A proof of the consistency of arithmetic can be given simply by showing that the standard model verifies all its axioms. This in fact is a proof in some set theory like ZFC, or in second-order arithmetic, quite far from the idea of a constructive proof. Or it can be given through Gödel's "Dialectica" interpretation using higher type functionals, which we briefly mentioned when discussing the extensions of simply typed lambda calculus. Having to transcend Hilbert's finitistic level with the aim of not straying too far from it, we emphasised how Gödel distinguished between finitistic reasoning and constructive reasoning. Or, finally, such a consistency proof can be given using transfinite induction up to $\varepsilon_0 = sup\{\omega, \omega^\omega, \omega^{\omega^\omega}...\}$. Recall from set theory that a set is called well-ordered if every non-empty subset of it has a least element. Ordinals (see section 5.2) are transitive sets, well ordered by the appartenance. The transfinite induction principle is a generalization of the complete induction principle on natural numbers, to more general well ordered sets. So, for instance, what we call the principle of transfinite induction *up the ordinal* α can be expressed as:

$$\forall x \in \alpha((\forall y < x P(y)) \rightarrow P(x)) \rightarrow \forall x \in \alpha P(x)$$

Well orderings are, in particular, linear ordering and actually holds true that any linearly ordered set enjoys the principle of complete induction if and only if it is well-ordered. Recall, by the way, that if we restrict ourselves to countable ordinal numbers, such as ε_0, these ordinals are in fact just different orderings of the natural numbers. This brings us to Gentzen's method, based on the well-foundedness of ordinal notations up to ε_0. The aim, in this case too,

is to remain as close as possible to finitistically acceptable reasoning. Introducing the sequent calculus we will see what difficulties arise in proving cut-elimination for theories, rather than for pure logic. In Chapter 7, we will demonstrate that the consistency of first-order logic follows from the cut-elimination theorem for its formalization in the sequent calculus because a consequence is that the empty sequent (the formalisation, in this calculus, of the contradiction) cannot be proven. We will also examine the difficulties and some partial solutions for fragments to the difficulties that arise regarding cut elimination for theories (i.e., in the presence of specific axioms or additional rules, such as induction, restricted to some classes of formulas). However cut elimination fails dramatically for *Peano Arithmetic* and therefore this method is not applicable. Gentzen, in alternative, devised a method for assigning an ordinal number smaller than ε_0 to each finite proof and showed how to effectively transform a proof in *Peano Arithmetic* of the empty sequent, into another proof of the empty sequent such that the latter receives a smaller ordinal than the former. Each reduction step is assigned an ordinal number, and then it is shown that the ordinal number decreases with every step. Actually it would be better to talk about *ordinal notations* and well ordering of these notations, since ordinals are given in the so-called *Cantor normal form* $\alpha = \omega^{\beta_0} + ... + \omega^{\beta_n}$ with exponents $\beta_0 \geq ... \geq \beta_n$. These "ordinals" indeed, are syntactic objects, rather than transfinite ordinals, strings of symbols introduced in a purely combinatorial manner, which can be equipped with a well-order and an algebra of elementary operations. In elaborating Gentzen's proof, Takeuti (1987) 92-100 proposes an effective method for demonstrating that these ordinal notations are well-ordered, based on so-called "eliminators", which are methods for showing that each strictly decreasing sequence starting from any ordinal notation is finite. This well-ordering property is called *accessibility* and constitutes an attempts, to a certain extent, to fill the gap with finitary mathematics. Starting from the assumption that, by contradiction, the empty sequent is provable, since the ordinal ε_0 is *accessible*, the proofs of the empty sequent can only be reduced a finite number of times, and this leads to a contradiction, by using the induction on transfinite ordinal numbers up to ε_0 in the form of an "infinite descent" argument (see Kleene (1952) 12-13). According to Takeuti (1987), the method based on "eliminators" is still acceptable from a finitist point of view, although modified in a more liberal sense to admit "*Gedankenexperimente* on (concrete) operations" (Takeuti (1987) 100-101).

In Gentzen's proof every step except the well-ordering of the ordering of type ε_0 can be performed in Skolem's *Primitive Recursive Arithmetic* PRA. If we agree, following Tait (1981), that what Takeuti calls Hilbert's "purely finitist standpoint" coincides with Skolem's primitive recursive arithmetic PRA, we can formally express this argument as follows:

$$\mathsf{PRA} + TI(\varepsilon_0) \vdash Con(\mathsf{PA})$$

where $TI(\varepsilon_0)$ is the formalization of the transfinite induction principle up to ε_0. We can also say that, in a certain sense, the ordinal ε_0 *measures the strength* of PA. However, to formalise the principle of transfinite induction in the language of arithmetic, we must find a way to represent the ordinals up to ε_0 in this language and the representation system chosen is not irrelevant to the result. What constitutes a good "natural" order is a much-debated question. Pathological, although *ad hoc*, examples are available. However in Rathjen (1999) (work that we also recommend for a more extensive discussion and bibliography on the subject, primarily by the author himself) is described how "natural" systems typically arise: the ordinals α smaller than ε_0 are represented in *Cantor normal form*, whose exponents themselves have Cantor normal forms with even smaller exponents. Since the process must end, this ensures that they can be encoded with natural numbers. It follows that we can devise a coding system $\ulcorner x \urcorner$ such that for every $\alpha < \epsilon_0$ the code $\ulcorner \alpha \urcorner$ is a natural numbers that denotes the ordinal α and the two structures, ε_0 with the operations $+, \cdot, \omega^\alpha$ and the relation $<$, on the one hand, and on the other hand the set of these codes, the primitive recursive functions $\ulcorner \alpha \urcorner \hat{+} \ulcorner \beta \urcorner = \ulcorner \alpha + \beta \urcorner$, $\ulcorner \alpha \urcorner \hat{\cdot} \ulcorner \beta \urcorner = \ulcorner \alpha \cdot \beta \urcorner$, $\hat{\omega}^{\ulcorner \alpha \urcorner} = \ulcorner \omega^\alpha \urcorner$ and the primitive recursive relation $\ulcorner \alpha \urcorner \prec \ulcorner \beta \urcorner$ if and only if $\alpha < \beta$ turn out to be isomorphic. Therefore, the principle of transfinite induction can actually be expressed with a formula in the language of primitive

recursive arithmetic:

$$\forall x(\forall y \prec xP(x)) \rightarrow \forall xP(x)$$

where $P(x)$ is a primitive recursive predicate. In Gentzen (1943) it is showed that PA proves the transfinite induction up to α, for each $\alpha < \varepsilon_0$, so we can say that his consistency result is optimal.

Between 1934 and 1943, Gentzen gave four proofs of consistency (three of which were published, according to the the historical reconstruction in Von Plato (2014)). Nowadays, when talking about "Gentzen's consistency proof for arithmetic", one usually refers to Gentzen (1938). Some *Proof Theory* textbooks report variations of this method (see for instance Takeuti (1987) and Mancosu, Galvan and Zach (2021)), trying to shed light on its darkest aspects.

The proof in Takeuti (1987) is long and complex and here we will limit ourselves to providing a taste, illustrating a crucial point and showing how induction is replaced by cut, that highlights the method. Again, we refer to Chapter 7 for notation regarding the sequent calculus and for the formalization in this calculus of the arithmetical theories (in particular, of the induction rule). On the contrary, the complexity measures are peculiar to this proof. In the mechanism for assigning ordinals less than ε_0 to sequents S within a proof π (notation $o(S;\pi)$, or simply $o(S)$), we proceed inductively, looking at the rule by which S was obtained. Hence in particular, if S is the lower sequent of a cut inference, where the upper sequents were assigned respectively the ordinals μ and ν, then $o(S) = \omega_{k-l}(\mu \# \nu)$ (where $\omega_0(x) = x$ and $\omega_{n+1}(x) = \omega^{\omega_n(x)}$), where k and l are the *hights* (where in this proof, the hight of a sequent is the maximum complexity, according to a measure of complexity called *grade*, of the cut-formulas and formulas to which induction under that sequent applies) of the upper sequents and of S respectively and $\#$ is the so-called "natural (or Hessenberg) sum": $\omega^{\mu_0} + ... +, \omega^{\mu_m} \# \omega^{\upsilon_0} + ... +, \omega^{\nu_n} = \omega^{\xi_0} + ... +, \omega^{\xi_{m+n}}$, where $\xi_0, ..., \xi_{m+n}$ are the exponents $\mu_0, ..., \mu_m, \nu_0, ..., \nu_n$ sorted in nonincreasing order. If S is the lower sequent of an induction inference, where the upper sequent was assigned the ordinal (in Cantor normal form) $\mu = \omega^{\mu_0} + ... +, \omega^{\mu_n}$, then $o(S) = \omega_{l-k+1}(\mu_0 + 1)$, where l and k are the hights of the upper sequent and of the lower sequent respectively.

The lemma according to which an hypotetical proof of the empty sequent $\Longrightarrow$ in a sequent calculus formalization of PA can be transformed into another proof of the same sequent, but with a lower ordinal associated with it, is obtained by analysing the so-called *end-piece* of a hypothetical proof of the this sequent, i.e., if the end-sequent is the empty sequent, that part of the proof selected ascending each thread starting from the final sequent, until a logical rule is encountered, and stopping in each branch at the lower sequent of such a logical rule. A crucial step is that of replacing all inductions appearing in such *end-piece* with a sequence of cuts, obataining a bound to the proof, strictly smaller than the previous one. Hence consider the end-piece of a proof π of the empty sequent and take the lowermost induction inference, i.e. an inference (see Chapter 7.) of this form:

$$\frac{\psi(\overline{x}), \Gamma \Longrightarrow \Delta, \psi(S(x))}{\psi(\overline{0}), \Gamma \Longrightarrow \Delta, \psi(s)}$$

Let respectively l and k the hights of (both) upper sequents and the height of the lower sequent. By defiition, the ordinal assigned to the latter is $o(S) = \omega_{l-k+1}(\mu_0 + 1)$, where $\mu = \omega^{\mu_0} + ... +, \omega^{\mu_n}$ is the one assigned to the upper sequent. It can be shown that in the end-piece any variable that is not an *Eigenvariable* can be replaced by a constant, so we can assume that s is a closed term and that therefore there is a proof for a certain number m (its value). that in the final part any variable that is not an Eigenvariable can be replaced by a constant, so we can assume that s is a closed term and that therefore, as a consequence, there is a proof of $\psi(\overline{m}) \Longrightarrow \psi(s)$, for some number m (the value of s). Inductive inference will be replaced by a sequence of cuts:

$$\dfrac{\dfrac{\dfrac{\dfrac{\psi(\overline{0}),\Gamma \Longrightarrow \Delta,\psi(\overline{1}) \quad \psi(\overline{1}),\Gamma \Longrightarrow \Delta,\psi(\overline{2})}{\psi(\overline{0}),\Gamma \Longrightarrow \Delta,\psi(\overline{2})} \quad \psi(\overline{2}),\Gamma \Longrightarrow \Delta,\psi(\overline{3})}{\psi(\overline{0}),\Gamma \Longrightarrow \Delta,\psi(\overline{3})}}{\vdots}}{\psi(\overline{0}),\Gamma \Longrightarrow \Delta,\psi(\overline{m})} \qquad \psi(\overline{m}),\Gamma \Longrightarrow \Delta,\psi(s)}{\psi(\overline{0}),\Gamma \Longrightarrow \Delta,\psi(s)}$$

Having all $\psi(\overline{n})$ the same complexity (the *grade* count the number of logical symbols), to all sequents $\psi(\overline{n}),\Gamma \Longrightarrow \Delta,\psi(S(\overline{n}))$ is assigned the same ordinal μ and by definition each $\psi(\overline{0}),\Gamma \Longrightarrow \Delta,\psi(n)$ is assigned $\mu\#...\#\mu$ (n-times). But in the proof of $\psi(\overline{m}),\Gamma \Longrightarrow \Delta,\psi(s)$ no induction or cut on complex formulas is required and in this case the assignement system assigns to it a finite number, say q. So, actually, the ordinal assigned to the final sequent $\psi(\overline{0}),\Gamma \Longrightarrow \Delta,\psi(s)$ of this subderivation is $\omega_{l-k}(\mu \cdot m + q)$, which is less than the original $\omega_{l-k+1}(\mu_1 + 1)$.

The mechanism for assigning ordinals appeared to some to be too ad hoc. Since it is largely the induction rule that causes problems, it seemed considerably preferable to admit infinitary proof systems as a formalization of PA, with infinitary rules replacing the induction rule, and obtaining a more transparent proof. In the wake of Schütte (1960), it is today quite common to overcome both the obstacle of eliminating the cuts and the alleged lack of transparency in the assignment of ordinal numbers to proofs by following a different method by admitting rules with *infinite premises*. This logic is frequently proposed in a certain version of the one-side sequents calculus (i.e. where the antecedent of all sequents is empty), introduced in Tait (1968). The so-called ω rule, in the two-side calculus, that we use in these lectures, consists instead of two types of infinitary inference, a possibility already explored by Hilbert, which Gentzen did not admit, however:

$$\dfrac{\Gamma \Longrightarrow \Delta,\phi(\overline{n}) \ \ (for\ all\ n \in \mathbb{N})}{\Gamma \Longrightarrow \Delta,\forall x\phi(x)} \qquad \dfrac{\phi(\overline{n}),\Gamma \Longrightarrow \Delta \ \ (for\ all\ n \in \mathbb{N})}{\exists x\psi(x),\Gamma \Longrightarrow \Delta}$$

(see ch.7. for notations) replacing the right universal quantifier rule and the left existential quantifier rule in sequent calculus. The version of PA with these rules instead of the induction rule, named PA_ω, enjoys cut elimination. It is proved that if the empty sequent has a proof in PA, then it has a cut-free proof in PA_ω of hight less than ε_0, but this is impossible, beacause, as we will see, there can be no cut-free derivation of the empty sequent. A proof of the consistency theorem along these lines is given, for example, in Schwichtenberg (1977), in Girard (1987) pp. 347-416, and in Troelstra and Schwichtenberg (2000) pp. 259-79.

Let us give at least an idea of how a consistency proof using the above infinitary rules can work. First, note that the derivations now are well-founded trees, which are generally *infinite*. Still arguing infomally, we define the *cut-rank* k of a derivation be the length of the longest cut-formula, and the *height* α of a derivation the supremum of the heights plus one of all its subderivations (where axioms have height 0). We write $\mathsf{PA}_\omega \vdash_k^\alpha \Gamma \Longrightarrow \Delta$ to mean that there exists a derivation of that sequent of height α and cut-rank k. For example, a PA_ω-derivation of the principle of induction now has this form (see Girard (1987) p. 355). Let $\Gamma = \psi(\overline{0}),\forall x(\psi(x) \to \psi(S(x))$. We can derive the sequents $\Gamma \Longrightarrow \psi(\overline{n})$ as follows:

(i) for $n = 0$, by weakening from $\psi(\overline{0}) \Longrightarrow \psi(\overline{0})$.

(b) Suppose that each $\Gamma \Longrightarrow \psi(\overline{n})$ has been proved. Hence build this derivation:

$$\dfrac{\dfrac{\dfrac{\Gamma \Longrightarrow \psi(\overline{n}) \quad \psi(\overline{n+1}) \Longrightarrow \psi(\overline{n+1})}{\Gamma, \psi(\overline{n}) \to \psi(\overline{n+1}) \Longrightarrow \psi(\overline{n+1})}}{\Gamma, \forall x(\psi(x) \to \psi(S(x)) \Longrightarrow \psi(\overline{n+1})}}{\Gamma \Longrightarrow \psi(\overline{n+1})}$$

(c) Now conclude as follows:

$$\dfrac{\dfrac{\dfrac{\dfrac{\Gamma \Longrightarrow \psi(\overline{0}), \Gamma \Longrightarrow \psi(\overline{1}), \Gamma \Longrightarrow \psi(\overline{2})....}{\Gamma \Longrightarrow \forall x \psi(x)}}{\psi(\overline{0}) \Longrightarrow (\forall x(\psi(x) \to \psi(S(x))) \to \forall x \psi(x))}}{\Longrightarrow \psi(\overline{0}) \to (\forall x(\psi(x) \to \psi(S(x)) \to \forall x \psi(x))}$$

This is a proof of the induction axiom which is a well founded tree *of height* $\omega + 2$. The crucial result (see Rathjen (2006)) of cut-elimination for this formalization of Peano Arithmetic with infinitary rules establishes that if $\mathsf{PA}_\omega \vdash^{\alpha}_{k+1} \Gamma \Longrightarrow \Delta$, then $\mathsf{PA}_\omega \vdash^{\omega^\alpha}_{k} \Gamma \Longrightarrow \Delta$ and therefore, by iterating this result we can obtain a cut-free derivation ($k = 0$) at the price of ingreasing the length of the derivation as:

$$\omega^{\omega^{\cdot^{\cdot^{\cdot^{\alpha}}}}}$$

for $k-$iterations, and therefore the height of such a derivation is bounded by ε_0. But it is provable that if $\mathsf{PA} \vdash \Gamma \Longrightarrow \Delta$, then it is also provable $\mathsf{PA}_\omega \vdash^{\omega+s}_{k} \Gamma \Longrightarrow \Delta$ for some $s, k < \omega$ and from the above, there exists a *cut-free* derivation of this sequent in PA_ω. Now, if we apply this argument to the empty sequent $\Longrightarrow$, we would obtain a cut-free derivation of it, which is impossible (indeed, it is an immediate application of the subformula property, see ch.7).

Based on these methods, the analysis of the strenght of mathematical theories by means of transfinite induction, now called *ordinal analysis*, has been extended to other theories. For example, sticking to the theories mentioned in this volume, the ordinal associated with Q is ω; the one associated with $\mathsf{I}\Delta_0$ is ω^2, while the one associated with PRA is ω^ω (see Sommer (1990) and Sommer (1995)). Finally, to cite a fragment of the second order, to which ordinal analysis has been most applied, the proof-theoretic ordinal associated with ATR_0, a theory which we mentioned in section 6.4., and of Feferman's *Predicative Analysis* is Γ_0, the so-called "Feferman-Schütte ordinal" , a countable ordinal that is the least ordinal "unreachable" by predicative means that we mentioned on p.131. However, analysing stronger and second-order theories goes beyond the purpose of these lecture notes. Rather, remaining within the realm of weak theories, we conclude by pointing to a line of research into the provability of well-foundedness of ordinal notations in weak theories of *Bounded Arithmetic* develped in Sommer (1995) and Beckmann, Pollett and Buss (2003). For instance, the latter authors define a notion of *well-foundedness on bounded domains* and show that the theories T^1_2 and S^2_2, that we introduce in section 7.3, can prove the well-foundedness on bounded domains of the ordinal notations below ε_0 and Γ_0.

Part III Proof Theory, mathematics and complexity

6. Independent sentences of mathematical character

6.1. Skepticism about Gödel's results

It has been pointed out (see for instance Grattan-Guinness (2011)) that the reception of incompleteness results within the *mathematical* community was very slow. One reason for the large underestimation of gödelians' results in part of the community of mathematicians, which helped to brake their assimilation, was linked to the *metamathematical* character of the statement "I am not provable" used in the constructive proof. The perceived distance from the concrete mathematical work is perhaps behind the most striking case of the silence in this regard: that of the the French group 'Bourbaki' of formalists mathematicians that began its activity in 1935. Actually still in 1948 Dieudonné, in David Hilbert's obituary, could write rather vaguely: "it seems that Hilbert's intuition ... has resulted in hopes a little exaggerated". Moreover, judging the presence of an undecidable proposition to be irrelevant, in Dieudonné (1987) the French mathematician complains that the undecidable statement of Gödel's first theorem is too artificial and unrelated to number theory. Nor were the mathematicians of the Bourbaki group too concerned about the problem of consistency, that considered just an empirical fact.

On the one hand, the formalization of mathematical reasoning, and on the other hand, the self-referential arithmetical statement, seemed to many mathematicians to be builded ad hoc for the purpose of obtaining the incompleteness result and very different from those occurring naturally in the mathematical research. In other words, the working mathematicians of the 1930s, although impressed by the wide-ranging philosophical implications of Gödel's limitative result, need not themselves have felt particularly limited by them, and could continue in their research for the most part as before.

A question that naturally arised, was therefore that about the pervasivity of Gödel's results, its impact on mathematical practice and if there were statements coming from the concrete mathematical practice, who shared the same fate of the bizarre statement invented by Gödel, but without resorting to diagonalization and other tricks. To a certain extent the underestimation of Gödelian achievements conceals a misunderstanding. Kurt Gödel himself, as some sources report, reacted to the objection of the *logical*, rather than *mathematical*, character of his results by saying that nothing is more mathematical than a Diophantine equation (see Kripke (2021) and the discussion on Matiyasevich's theorem on p.46). The tools introduced by the incompleteness theorems are perhaps better defined as "strikingly original mathematics, with something of the charm of Cantor's first work in set theory", as remarks Macintyre (2011), although, according to this prominent British model-theorist, the discovery of the phenomenon of incompleteness had little effect on current mathematics.

For this reason, in an attempt to address these objections, some scholars have devoted themselves to the search for independent statements of mathematical content, with proofs that did not rely on the typical Gödelian toolbox. These statements were actually found systematically, starting from Paris (1978), both of combinatorial as Paris and Harrington (1977), or numeric character, as for example Kirby and Paris (1982), i.e. expressing natural

Duccio Pianigiani, University of Siena, Italy, duccio.pianigiani@unisi.it, 0000-0001-9441-7226

Referee List (DOI 10.36253/fup_referee_list)
FUP Best Practice in Scholarly Publishing (DOI 10.36253/fup_best_practice)

Book References DOI 10.36253/979-12-215-0778-2.references

properties of integers. The distinguished logician John Barwise, editor of the famous Barwise and Keisler (1977), points out at p. 1133 that as early as 1931, i.e. the year of the publication of the incompleteness theorem, the mathematical world was already calling for similar results concerning statements of a clear and simple mathematical nature. Barwise points to the Paris and Harrington theorem as the first example of this kind (in the authors' own words: "a reasonably natural theorem of finitary combinatorics, a simple extension of the Finite Ramsey Theorem"). The other popular example is Paris and Kirby's independence result concerning Goodstein sequences and of the related 'Hydra game'. We give an account of both in this chapter. Goodstein (1944) proved the rather surprising result that eventually these sequences reach 0. Studying the correspondence between Bernays and Goodstein, Rathjen (2015) shows how close Goodstein came to proving the independence result later obtained in Kirby and Paris (1982). We will see that this last paper actually showed that the power of Peano's Arithmetic is not enough to prove this result also for a kind of sequences so-called *special* Goodstein sequences. Indeed, it has been emphasised by various logicians that the first strictly mathematical result of independence was actually the famous result of independence of the induction principle up to ε_0 of Gentzen (1936). In particular Kripke (2021) puts the question in these terms: Gentzen gave the first such result, that was restated by Goodstein in a number-theoretic form.

6.2. Ramsey's theorems and the Paris-Harringhton theorem

An important example of combinatorial statement formalized in first order Peano arithmetic, but independent of it, arises in the context of the mathematical study of combinatorial objects known as *Ramsey theory*. In a concise manner, we are concerned with problems like that posed in *The American Mathematical Monthly* in 1958 (N. 65, vol. 6, Problem E 1321):

> "Prove that in a meeting of any six persons, about three of them are either mutual acquaintances or complete strangers to each other".

The party acquaintances problem. So, what is the minimum number of guests such that it is sure to find among them three people who do not know each other, or three people who know each other? For example, let us suppose that there are five individuals: observe the complete graph K5 below, where the red lines connect individuals who do not know each others and the blue lines individuals who they know to persuade you that there cannot be three people who know each other or three people who do not know:

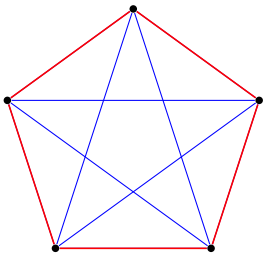


Figure 5. Graph K5.

But if we add an individual (the complete graph K6 below), then six individuals are sufficient to determine a *clique* of three individuals who either know or do not know each other:

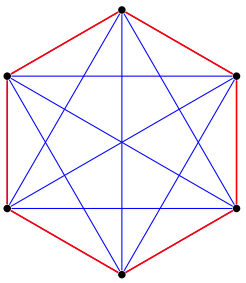


Figure 6. Graph K6.

Notice that in this case each node can belong to at least three red edges, or at least three blue edges. For each node x there are 5 edges incident to it. Take for example three of these edges (x, a), (x, b) and (x, c) of the same color, say blue. Now, if any of the edges $(a, b), (a, c)$ and (b, c) is blue, then we're done. If not, then all these last three edges must be red and we are done as well.

The general problem is: find the minimum number $R(k, h)$ of guests which is necessary to invite to a party in such a way that at least k of them know each other, or at least h guests do not know each other. In our example, $R(3, 3) = 6$.

Ramsey's number $R(h, k)$ in general is definde as the minimum n such that for any coloring in two colors B and R of the edges of the complete graph (i.e. in which each pair of nodes is connected with an arc) with n-nodes, the graph contains, either an R-subgraph with k-nodes, or a B-subgraph with h-nodes, where a coloration is a function $f : \{\langle x, y\rangle | x, y \leq n\} \to \{B, R\}$.

The *Ramsey theorem for graphs* states that for all k, h there is a n such that each complete graph of at least n nodes and colored with two colors $Blue, Red$, must contain a $Blue$ monocromatic subgraph with k nodes, or a Red monocromatic subgraph with h nodes. This result holds also for a finite number of colors. The *infinite* version states that every countably infinite complete graph must contain a monochromatic complete infinite graph. In view of the results we wish to present, we observe that the theorem can be given in a more general version. First let $[A]^k = \{X \subseteq A || X| = k\}$, where with the notation "$|X|$" we mean the cardinality of X; notice that a graph $G = \langle V, E\rangle$ is given by a set of nodes V and a set of edges E, namely of unordered pairs of nodes, that can be seen as a subset of $[V]^2$ (a complete graph is one such that $E = [V]^2$). E.g. $A = \{a, b, c, d, e\}$, then a three-color (say red, blue and green) coloring of $[A]^4$ might be the following:

(r) $\{a, b, c, d\}$, $\{a, b, d, e\}$

(b) $\{a, c, d, e\}$

(g) $\{a, b, c, e\}$, $\{b, c, d, e\}$.

Ramsey's theorem for sets. Given $X \subseteq \mathbb{N}$, we write $[X]^k$ to indicate the $Y \subseteq X$ such that $|Y| = k$. Moreover, by identifiyng a number with the set of its predecessors $m = \{0, 1, 2, ..., m - 1\}$, a surjective function $f : [X]^k \longrightarrow m$ will be called *coloring*, or *partition*. A subset $Y \subseteq X$ will be called *omogeneous*, or *monocromatic*, if all elements of $[Y]^k$ receive the same color by f, i.e. if $f \restriction_Y$ is costant. Lastly, with:

$$n \longrightarrow (h)^k_m$$

we mean that for all set X of cardinality $\geq n$, for all coloring $f : [X]^k \longrightarrow m$, there is a $Y \subseteq X$ of cardinality $\geq h$ such that Y is homogeneous with respect to f.

Theorem 98. (Infinite Ramsey theorem) *Let n, k be positive integers and let $X \subseteq \mathbb{N}$ be infinite. If $[X]^n$ is colored with k colors, then X has a monocromatic subset. In symbols:*

$$\aleph_0 \longrightarrow (\aleph_0)^n_k$$

Proof. We follow Marker (2002) Chapter 5.1. Induction on n:

(a) $n = 1$. Notice that $[X]^1 = \{\{x\}|x \in X\}$. Hence, if $f : [X]^1 \longrightarrow K$, then there is an infinite $Y \subseteq X$ such that f is constant on $[Y]^1$, i.e. there exists $i < k$ such that $f^{-1}(i)$ is infinite. This follows from the "Pigeon hole principle": if we place an infinite number of objects in a finite number of boxes, at least one box will contain infinite objects.

(b) $n > 1$. Without loss of generality, set $X = \mathbb{N}$. Let $f : [\mathbb{N}]^n \longrightarrow k$. Suppose the theorem holds for $n - 1$ by inductive hypothesis. Let us define a sequence:

$$\mathbb{N} = X_0 \supset X_1 \supset X_2 \supset ...$$

and a sequence $a_0 < a_1 < a_2 < ...$ where $a_i = min(X_i)$. Suppose we have defined a_s and X_s and let:

$$f_{a_s} : [X_s \setminus \{a_s\}]^{n-1} \longrightarrow k$$

defined as:

$$f_{a_s}(A) = f(A \cup \{a_s\})$$

Hence $A \cup \{a_s\} = \{a_s, a_{i_0}, ..., a_{i_{n-2}}\} \subseteq X_s$. By the induction hypothesis exists $Z \subseteq X_s \setminus \{a_s\}$ homogeneous with respect to f_{a_s}. Hence we let $X_{s+1} = Z$.

Let now C_{a_s} be the color assigned by f_{a_s} to all elements of $[Z]^{n-1}$. Notice tha colors are *finite*, while the a_s are *infinite*, hence for an infinite $Y = \{a_{m_0}, a_{m_1}, a_{m_2}...\}$ we will have $C_{a_{m_0}} = C_{a_{m_1}} = C_{a_{m_2}} = ... = j$. Then, for all $\{a_{m_{j_0}}, ..., a_{m_{j_{n-1}}}\} \in [Y]^n$, $f(\{a_{m_{j_0}}, ..., a_{m_{j_{n-1}}}\}) = j$.

Indeed, by definition $f(\{a_{m_{j_0}}, ..., a_{m_{j_{n-1}}}\}) = f_{a_{m_{j_0}}}(\{a_{m_{j_1}}, ..., a_{m_{j_{n-1}}}\} = j$, where $\{a_{m_{j_1}}, ..., a_{m_{j_{n-1}}}\} \subseteq X_{m_{j_0+1}}$. QED

Corollary 20. (Finite Ramsey theorem) *For all $k, n, m \in \mathbb{N}$ there exists $h \in \mathbb{N}$ such that:*

$$h \longrightarrow (m)^n_k$$

Proof. Suppose this is not true. Hence for all $h \in \mathbb{N}$, let T_h the set of colorations $f : [h]^n \longrightarrow k$ such that there is no $X \subseteq h$, of cardinality bigger or equal to m, X monocromatic for f. Notice that if $f \in T_{h+1}$, then there is a unique $g \in T_h$ such that $g \subset f$. Let therefore $T = \bigcup_h T_h$;it is a finitely branching tree but infinite and therefore by König's lemma it has an infinite branch $f_0 \subset f_1 \subset f_2 \subset ...$, where $f_i \in T_i$. Let therefore $f = \bigcup_i f_i$; hence $f : [\mathbb{N}]^n \longrightarrow k$. For the infinite Ramsey theorem exists an infinite set $X <$ homogeneous for f. Let therefore $x_0, .., x_{m-1}$ the first m elements of X and let $s > x_{m-1}$. Then $\{x_0, .., x_{m-1}\}$ will be homogeneous for f_s (contradiction). QED

We would now like to show that changing the premises of the the finite Ramsey theorem in an apparently harmless manner, as the fact that the proof is almost the same invites one to think (but that is not the only one for it!), has unexpected consequences. Notice that the proof that uses the *infinite version* is similar to that of the finite Ramsey theorem.

Theorem 99. *For all $k, n, m \in \mathbb{N}$ exists $h \in \mathbb{N}$ such that, if $f : [h]^n \longrightarrow k$ is a coloring, then there exists $Y \subseteq h$ homogeneous for f and such that:*

(a) $|Y| \geq m$

(b) $|Y| \geq min(Y)$ *("Y is relatively big").*

In symbols $h \longrightarrow_* (m)^n_k$.

Proof. Suppose that not; then for each $h \in \mathbb{N}$, let T_h the set of colorings $f : [h]^n \longrightarrow k$ such that there is no $Y \subseteq h$, of cardinality bigger or equal to m and to $min(Y)$, Y monocromatic for f. Hence let us consider $T = \bigcup_h T_h$; it is an infinite, finitely branching tree and therefore the König's lemma has an infinite branch $f_0 \subset f_1 \subset f_2 \subset ...$ where $f_i \in T_i$. Let therefore $f = \bigcup_i f_i$; hence $f : [\mathbb{N}]^n \longrightarrow k$. For the infinite theorem of Ramsey exists an infinite set X homogeneous for f. Let $x_0 = min(X)$; let $x_0, ..., x_{e-1}$ the first e elements of X and let $s \geq e \geq x_0, m$. Then $Y = \{x_0, ..., x_{e-1}\}$ is homogeneous for f_s, $|Y| \geq m, min(Y)$ (contradiction). QED

Nevertheless, there are considerable differences between these statements:

(a) The infinite Ramsey's theorem is formalized and provable in *Peano Arithmetic* of the second order PA_2.

(b) The finite Ramsey's theorem is provable in first order *Peano Arithmetic* PA.

(c) The Paris-Harrington's theorem is formalized in PA and true in the standard model, but unprovable in PA.

Following Marker (2002), pp. 175-202, we prove point 3. in a manner that is quite usual today, namely by proving the independence of the Kanamori-McAloon principle. This theorem is actually a consequence of Paris and Harrington's statement.

Definition 46. *Let us say that:*

(a) if $f : [X]^n \longrightarrow \mathbb{N}$ is a coloring, it is regressive when $f(A) < min(A)$ for all $A \in [X]^n$

(b) $Y \subseteq X$ is called min-homogeneous for f, if when $A, B \in [Y]^n$ and $min(A) = min(B)$, then $f(A) = f(B)$.

For instance $f : [\{1, 2, 3\}]^2 \longrightarrow 2$ where $f(\{1, 2\}) = f(\{1, 3\}) = 0$ and $f(\{2, 3\}) = 1$ is regressive and the set $\{1, 2, 3\}$ is min-homogeneous for f, because $min(\{1, 2\}) = min(\{1, 3\})$ and $f(\{1, 2\}) = f(\{1, 3\})$.

Theorem 100. *For all $c, m, n, k \in \mathbb{N}$ exists d such that, if $f_1, ..., f_k : [d]^n \longrightarrow d$ are regressive, then exists a subset $Y \subseteq [c, d)$ such that:*

(a) $|Y| \geq m$

(b) Y is min-homogeneous for all $f_1, ..., f_k$.

Although formalizable in the language of PA *and true, this statement is not provable in* PA.

Definition 47. *Let:*

$$\Gamma = \{\phi_1(u_1, ..., u_m, v_1, ..., v_n), ..., \phi_e(u_1, ..., u_m, v_1, ..., v_n)\}$$

a set of formulas and let $\mathcal{M}$ be a model of PA. *Let us call $I \subseteq M$ a set of indiscernible elements for Γ, if for all set of elements $x_0 < x_1 < ... < x_n$ e $x_0 < y_1 < ... < y_n$ in I, for all $a_1, ..., a_m < x_0$ and all $\phi_i \in \Gamma$, we have*[1]

From Kanamori-McAloon follows the existence of a set of indiscernibles in the *standard* model.

Main Lemma. For all $e, m, n \in \mathbb{N}$ and formulas;

$$\phi_1(u_1, ..., u_k, v_1, ..., v_n), ..., \phi_e(u_1, ..., u_k, v_1, ..., v_n)$$

there exists a set of indiscernibles I, such that $|I| \geq m$.

[1] To improve readability, we use this simplified notation $\mathcal{M} \models \psi(a)$, where a is an element of the model (not a constant or a numeral), meaning that a satisfies $\phi(x)$ in that model.

Proof. Let $m > 2n$. We will use two facts:

(a) From the finite version of Ramsey theorem, let us take w such that:

$$w \longrightarrow (m+n)^{2n+1}_{e+1}$$

(b) From Kanamori and McAloon, given $w, 2n+1$, there exists s such that for all regressive functions $f_1, ..., f_k : [s]^{2n+1} \longrightarrow s$ there exist a subset $Y \subseteq s$ such that $|Y| \geq w$ and Y is min-homogeneous for all $f_1, ..., f_k$.

Hence let us define *regressive functions* $f_1, ..., f_k$ and a coloring g in this way. Let $X = \{x_0, ..., x_{2n}\}$ where $x_0 < ... < x_{2n} < s$. Then:

(a) if for all $0 < i \leq e$ and $a_1, ..., a_k < x_0$ we have:

$$\phi_i(a_1, ..., a_k, x_1, ..., x_n) \leftrightarrow \phi_i(a_1, ..., a_k, x_{n+1}, ..., x_{2n})$$

then let $f_j(X) = 0$, for all $0 < j \leq k$ and $g(X) = 0$.

(b) If there are instead $i \leq e$ and $a_1, ..., a_k < x_0$ such that:

$$\neg(\phi_i(a_1, ..., a_k, x_1, ..., x_n) \leftrightarrow \phi_i(a_1, ..., a_k, x_{n+1}, ..., x_{2n}))$$

then let $f_j(X) = a_j$, for all $0 < j \leq k$, and $g(X) = i$. In this way:

$$\neg(\phi_{g(X)}(f_1(X), ..., f_k(X), x_1, ..., x_n) \leftrightarrow \phi_{g(X)}(f_1(X), ..., f_k(X), x_{n+1}, ..., x_{2n}))$$

Note that $f_j(X) < min(X)$, indeed, either $f_j(X) = 0$, or $f_j(X) = a_j < x_0 = min(X)$, namely the $f_j : [s]^{2n+1} \to s$ are *regressive*. Therefore we can apply Kanamori - McAloon to find $Y \subseteq s$ min-homogeneous, $|Y| \geq w$; for Ramsey and the choice of w there exist $Z \subseteq Y$ and $i < e+1$ such that $g(A) = i$, for all $A \in [Z]^{2n+1}$, $|Z| > m+n$ (recall that $w \longrightarrow (m+n)^{2n+1}_{e+1}$ and that $g : [s]^{2n+1} \longrightarrow e+1$).

Recall that $m > 2n$ and then $|Z| > 3n$, so that we can write Z as $x_0 < x_1 < ... < x_{3n} < ....$ Note that Z is min-homogeneous for each f_j.

Claim. We show now that $g(A) = i = 0$ for all $A \in [Z]^{2n+1}$. Let therefore:

(a) $\alpha = x_0, x_1, ..., x_n, x_{n+1}, ..., x_{2n}, \overline{x_{2n+1}, ..., x_{3n}}$

(b) $\beta = x_0, x_1, ..., x_n, \overline{x_{n+1}, ..., x_{2n}}, x_{2n+1}, ..., x_{3n}$

(c) $\gamma = x_0, \overline{x_1, ..., x_n}, x_{n+1}, ..., x_{2n}, x_{2n+1}, ..., x_{3n}$

subsets of Z of cardinality $2n+1$, where we highligth the omitted elements.

Suppose by contradiction that $i > 0$. Note that α, β, γ are subsets of Z whose cardinality is $2n+1$. Let now $a_j = f_j(\alpha) = f_j(\beta) = f_j(\gamma)$ observing that from the min-homogeneity the sequences α, β, γ of elements of Z give the same result for each f_j and, as observed before, this is less than x_0. For Ramsey's theorem (being Z homogeneous for g) $g(\alpha) = g(\beta) = g(\gamma) = i > 0$. Condition (2) says that:

(a) $\neg\phi_i(a_1, ..., a_k, x_1, ..., x_n) \leftrightarrow \phi_i(a_1, ..., a_k, x_{n+1}, ..., x_{2n})$

(b) $\neg\phi_i(a_1, ..., a_k, x_1, ..., x_n) \leftrightarrow \phi_i(a_1, ..., a_k, x_{2n+1}, ..., x_{3n})$

(c) $\neg\phi_i(a_1, ..., a_k, x_{n+1}, ..., x_{2n}) \leftrightarrow \phi_i(a_1, ..., a_k, x_{2n+1}, ..., x_{3n})$

But the conjunction of these conditions implies a contradiction. Hence $i = 0$

Take now the last n-elements of Z, say $z_1 < z_2 < ... < z_n$. Let therefore $I = Z \setminus \{z_1, ..., z_n\}$. Taking two sequences $x_0 < x_1 < ... < x_n$ and $x_0 < y_1 < ... < y_n$ of elements in I, for $a_1, ..., a_k < x_0$ we will have:

(a) $\phi_i(a_1, ..., a_k, x_1, ..., x_n) \leftrightarrow \phi_i(a_1, ..., a_k, z_1, ..., z_n)$

(b) $\phi_i(a_1, ..., a_k, y_1, ..., y_n) \leftrightarrow \phi_i(a_1, ..., a_k, z_1, ..., z_n)$, from which follows

(c) $\phi_i(a_1, ..., a_k, x_1, ..., x_n) \leftrightarrow \phi_i(a_1, ..., a_k, y_1, ..., y_n)$

Namely, I is *a set of indiscernibles*. QED

The Main Lemma gives us indiscernibles for a finite Γ; at the opposite, in the following we need it for the *infinite* class Δ_0. Notice that the *Main Lemma* is formalizable in any theory (say PA) which has a truth predicate for the formulas concerned. Let therefore $\mathcal{M}$ a non standard model of PA and suppose that Kanamori-McAloon holds in it. Since the lemma of existence of indiscernibles holds in it for all standard numbers e, m, n, by overspill holds for some non-standard c and for what "look" to $\mathcal{M}$ like the first c formulas from Δ_0, having $2c$ variables and therefore we note that holds for all infinite Δ_0 standard formulas.

Recall that a general theorem of *absoluteness* holds.

Theorem 101. *Let $\mathcal{M} \models$ PA and let $J \subseteq \mathcal{M}$ an initial segment of it (i.e. if $a \in \mathcal{M}$, $b \in J$ and $a < b$, then $a \in J$). Let $\phi(x) \in \Delta_0$ and let $c \in J$. Then $\mathcal{M} \models \phi(c)$ if and only if $J \models \phi(c)$.*

Let us consider now $\Gamma = \Delta_0$. We show how to use indiscernibles to find initial segments that are models of PA.

Theorem 102. *Let $\mathcal{M}$ be a model of* PA *and let $x_1 < x_2 < x_3 < ...$ indiscernibles for the class Δ_0 of formulas. Let $J = \{y \in M | \exists i(y < x_i)\}$. Then also J is a model of* PA.

Proof. First we prove the closure under the operations:

(a) suppose that $i < j < k < e$ and $a < x_i$; if $a + x_j \geq x_k$, then there exists $b \leq a (b + x_j = x_k)$; but from indiscernibility also we have $b + x_j = x_e$, from which $x_k = x_e$, against the hypothesis that $x_k < x_e$. Hence $a + x_j < x_k$ and since x_k is un indiscernible also we have $a + x_j \in J$. In particular $x_i + x_j \leq x_k$. It follows that J is closed under addition.

(b) Under the same conditions we also have $a \cdot x_j < x_k$. Otherwise we would have, for some minimal $a = b + 1$, that $bx_j < x_k \leq (b + 1)x_j$. By indiscernibility, also $x_e \leq (b + 1)x_j$. Moreover, $(b + 1)x_j = bx_j + x_j < x_k + x_j$. But from the previous point $x_k + x_j \leq x_e$, from which $x_e \leq (b + 1)x_j < x_e$ (contradiction). Hence $ax_j < x_k$. It follows the closure under multiplication.

Now we show the closure under induction. We remark therefore that truth of a formula in J can be reduced to truth of a Δ_0 formula in $\mathcal{M}$. For instance, let us consider $\exists v_1 \forall v_2 \exists v_3 \psi(w, v_1, v_2, v_3)$. Let $a < x_i$. But $J \models \exists v_1 \forall v_2 \exists v_3 \psi(a, v_1, v_2, v_3)$ iff there exists $i_1 > i$ such that for all $i_2 > i_1$, exists $i_3 > i_2$, such that:

$$J \models \exists v_1 < x_{i_1} \forall v_2 < x_{i_2} \exists v_3 < x_{i_3} \psi(a, v_1, v_2, v_3)$$

But truth of Δ_0 formulas is preserved in extensions of models and then these remains true in $\mathcal{M}$. Being the x_j indiscernibles, the above formulas can be reduced to a unique formula, i.e. it will be true in $\mathcal{M}$ that $\exists v_1 < x_{i+1} \forall v_2 < x_{i+2} \exists v_3 < x_{i+3} \psi(a, v_1, v_2, v_3)$.

Hence $J \models \exists v_1 \forall v_2 \exists v_3 \psi(a, v_1, v_2, v_3)$ if and only if:

$$M \models \exists v_1 < x_{i+1} \forall v_2 < x_{i+2} \exists v_3 < x_{i+3} \psi(a, v_1, v_2, v_3)$$

But in $\mathcal{M}$ will be true the least number principle (equivalent to the induction). Hence there is a *minimum* $a^* < x_i$ such that the formula holds in the model, i.e.:

$$M \models \exists v_1 < x_{i+1} \forall v_2 < x_{i+2} \exists v_3 < x_{i+3} \psi(a^*, v_1, v_2, v_3)$$

but from this follows $J \models \exists v_1 \forall v_2 \exists v_3 \psi(a^*, v_1, v_2)$, and the the least number principle also holds in J. QED

Independence of the Kanamori-McAloon principle. Let $\mathcal{M}$ a non-standard model of PA and let c a non-standard element.

(A) Recall that PA proves the finite "Ramsey". Hence take the *minimum* w such that in $\mathcal{M}$ is true that $w \longrightarrow (3c+1)^{2c+1}_c$, namely, that for all $f : [w]^{2c+1} \to c$, there is Y of cardinality bigger than $3c+1$ homogeneous.

(B) Let us suppose, by contradiction, that PA proves Kanamori-McAloon's principle. Therefore the above $\mathcal{M}$ is a nonstandard model of Peano Arithmetic where this principle holds. Given the above w, let therefore d the *minimum* such that, if $f_0, ..., f_c : [d]^{2c+1} \longrightarrow d$ are regressive, then exists $Y \subseteq [c, d)$ such that $|Y| \geq w$ and Y is min-homogeneous for $f_0, ..., f_c$.

Following the steps of the proof of the *Main Lemma* inside the model, we can indeed obtain a set $c \leq I < d$ of cardinality bigger or equal than c, that according to $\mathcal{M}$ is a set of indiscernibles for Δ_0-formulas coded with code at most c (and then in particular *all* Δ_0-formulas coded by standard numbers). Recall that in PA we can define *partial* truth predicates, in particular for the Δ_0 formulas. This, toghether with the usual coding apparatus, is all what is needed to formalize the proof of the result about the existence of indiscernibles and rebuild it inside the model $\mathcal{M}$. Following the steps of the proof of the *Main Lemma* inside the model, we can obtain a set $I \subseteq Y$ of cardinality bigger than c (where $Y \subseteq [c, d)$ is the above mentioned set), that according to $\mathcal{M}$ is a set of indiscernibles for Δ_0-formulas coded with code at most c (and then in particular *all* Δ_0-formulas coded by standard numbers).

Let therefore $x_0 < x_1 < x_2 < ...$ an initial segment of I and let $J = \{y \in M | y < x_i, for\ some\ i\}$. For the previous theorem J is a model of PA. Moreover $c \in J$, but $d \notin J$. Note that:

(A') For finite Ramsey's theorem (true in J) there is $v \in J$ such that in J is true (A), i.e. for all functions $f : [v]^{2c+1} \longrightarrow c$ there exists Z homogeneous of cardinality bigger or equal to $3c+1$. Since all functions and sets needed for this statement are included in J, this holds also in $\mathcal{M}$. But once fixed $c, 2c$, and $3c+1$, according to (A) the set w was minimal in $\mathcal{M}$ to satisfy this statement, hence $w \leq v$. Therefore $w \in J$.

(B') Analogously, if $h \in J$ and is true in J that if the $f_0, ..., f_c : [h]^{2c+1} \longrightarrow h$ are regressive, by the hypothesis that the Kanamori-McAloon's principle is true in models of PA, there exists Y min-homogeneous for them of cardinality greater or equal to w, and arguing as above, this is also true for $\mathcal{M}$. But according to (B) d was the minimum for which this is true in $\mathcal{M}$. Hence $d \leq h$ and therefore $d \in J$. But we also had $d \notin J$ (contradiction)

It follows that Peano Arithmetic does not prove Kanamori-McAloon statement and consequently does not prove the Paris-Harrington either.

6.3. The Hydra game

Kirby and Paris (1982) proved an extension of Goodstein's original result, mentioned in the introduction: recall that each integer can be written in base $b \geq 2$, in form $b^{n_0} \cdot c_0 + ... + b^{n_k} \cdot c_k$ where $n_0 \geq n_1 \geq ... \geq n_k$. The expontents in turn will be written in base b. For example, if $n = 266$ and $b = 2$, we can write n in base b as:

$$n = 2^{2^{2+1}} + 2^{2+1} + 2^1$$

Let us define $G_n(x)$ by cases as follows: $G_n(m) =$ "the number obtained replacing each n in base n representation of m, with $n+1$, if $m \neq 0$ ($G_n(m) = 0$ otherwise)".

For instance $G_2(266) = 3^{3^{3+1}} + 3^{3+1} + 3^1$; a Goodstein's sequence $n_0, n_1, n_2...$ is such that $n_0 = n$ and $n_{k+1} = G_{k+2}(n_k) - 1$ if $n_k > 0$; in our previous element:

$$\begin{aligned}
&n_0 = n = 266\\
&n_1 = G_2(n_0) - 1 = 3^{3^{3+1}} + 3^{3+1} + 2\\
&n_2 = G_3(n_1) - 1 = 4^{4^{4+1}} + 4^{4+1} + 1\\
&n_3 = G_4(n_2) - 1 = 5^{5^{5+1}} + 5^{5+1}\\
&\vdots
\end{aligned}$$

This is called "the Goodstein sequence for m starting at 2". This definition can be generalized to "the Goodstein sequence for m starting at r", for each $r \geq 2$. The *Goodstein's theorem* (1944) says that each such sequence converges to 0, i.e. for all numbers n, there is a number k such that $n_k = 0$. However this k is of very big dimension, for instance, if $n = 4$, then $k = 3 \cdot 2^{40265321} - 3$.

Well, although the Goodstein's theorem is formalizable in the language of Peano arithmetic at first order PA, this theory (which we assume to be consistent) is not able to prove the corresponding formal statement. It is in fact possible to prove that the statement of the Goodstein's theorem is equivalent to the provability in PA, of the principle $TI(\varepsilon_0)$ of transfinite induction up to ε_0, but it is well known by Gentzen's historical results that this principle is not provable in PA and in fact, it is the principle used by Gentzen to show the consistency of this theory. The game of Hercules versus Hydra of Kirby and Paris is a combinatorial game whose termination is true, although this truth is not provable in Peano Arithmetic. The Hydra is the monster of Graeco-Roman mythology with many heads, which grow multiplying every time they are cut off. Hercules' aim is to cut off all Hydra's heads. Actually, in the game hydras are graphs, in particular they are finite rooted trees, i.e. connected graphs with no cycles and a specific node, the root. A head is an edge (a, b) where b is a leaf of the tree and the node a is its (unique) predecessor, that we call "its father" (sometimes a is also called "neck"). After Hercules cuts off a head, the hydra grows new heads, namely the tree changes according to this rule:

(a) At stage $n \geq 1$ Hercules chop one head, namely an edge (a, b) as above, where b is a leaf.

(b) At this point the Hydra grows n new heads: from the predecessor of a, say s (sometimes called "grandfather of b", or "trunk"), the monster generates n copies rooted in s of what remains of the subtree generated by a after the decapitation, i.e. after removing (a, b).

(c) if one of the two nodes of the edge just chopped off coincides with the root, no new head is grown.

(d) Hercules wins the battle if he reduces in a finite number of moves the monster to its root.

Actually Hercules always kill the Hydra: we know that every recursive strategy is a winning strategy, however, the statement "Every recursive strategy is a winning strategy", although true, is not provable in Peano Arithmetic. Here we illustrate the original proof due to Kirby

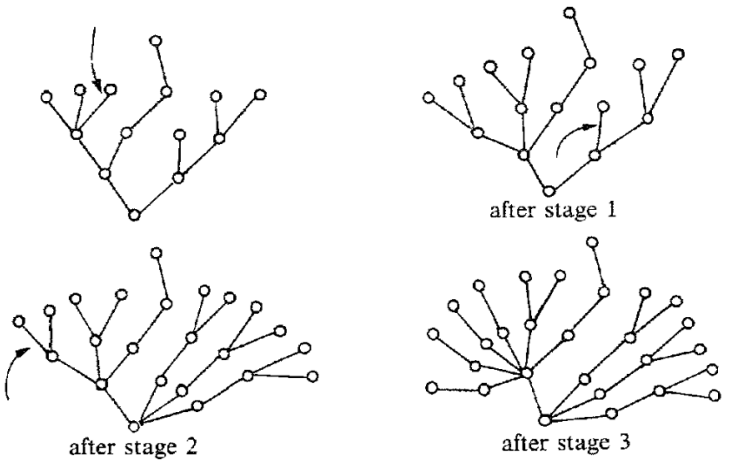


Figure 7. Reproduction scheme of the Hydra, from the paper by Kirby and Paris.

and Paris, which makes use of the method of *indicators* and of the theory of α-*large sets* (where α is an ordinal) introduced by Ketonen and Solovay. The *indicators* method comes from model theory, was introduced by Kirby and Paris in the 1970s and was later used to find proofs of independence. Other, proof theoretic, rather than model-theoretic methods were used for the same purpose. For instance, Carlucci (2003) followed another route, using the sequents calculus and showing that a relation holds between Kirby-Paris *Hydra Game* and Gentzen Reduction Strategy by a natural interpretation of derivations as hydras.

Definition 48. *Suppose that Θ is a set of cuts of a countable non-standard model $\mathcal{M} \models I\Sigma_1$. A Σ_1 formula $Y(x, y, z)$ is called an indicator for Θ in $\mathcal{M}$, if the following hold:*

(a) $\mathcal{M} \models \forall x \forall y \exists! z Y(x, y, z)$*; in other words, this formula defines a function in the model $\mathcal{M}$ (therefore we can write $Y^{\mathcal{M}}(x, y) = z$ in place of $\mathcal{M} \models Y(x, y, z)$).*

(b) For all $a, b \in \mathcal{M}$, $Y^{\mathcal{M}}(a, b) > \mathbb{N}$ (meaning that is bigger of all natural numbers) if and only if there exists $I \in \Theta$ of $\mathcal{M}$, such that $a \in I < b$ (which means that contains a and all its elements are less than b). It is actually provable that there are $2^{\aleph_0}$ such initial segments.

(c) For a, b, c, d elements of $\mathcal{M}$, if $a \leq b$ and $c \leq d$, then $Y^{\mathcal{M}}(b, c) \leq Y^{\mathcal{M}}(a, d)$.

See Paris (1980) and Hájek and Pudlák (1993), pp. 245-260 for a detailed discussion of the subject. For example, using the notation from the previous section, this is an indicator for $\Theta = \{I \subseteq_e \mathcal{M} | I \models \mathsf{PA}\}$ where $\mathcal{M} \models I\Sigma_1$ (we say that it is *an indicator for models of* PA *in* $I\Sigma_1$):

$$Y(a, b) = max\ c\ such\ that\ [a, b] \rightarrow_* (c + 1)^c_c$$

The following theorem shows some important properties of indicators for $\Theta = \{I \subseteq_e \mathcal{M} | I \models \mathsf{T}\}$. Actually, this is the case in which we are most interested and we speak in that case simply of *indicators for models of* T.

Theorem 103. *Let* T *be a recursive extension of $I\Sigma_1$ and $Y(x, y) = z$ an indicator for $\Theta = \{I \subseteq_e \mathcal{M} | I \models \mathsf{T}\}$ in any countable model $\mathcal{M}$ of* T*. Hence the following hold:*

(a) $\mathsf{T} \nvdash \forall x \forall z \exists y Y(x, y) \geq z$

(b) *for all natural number* n, $\mathsf{T} \vdash \forall x \exists y Y(x,y) \geq \overline{n}$

(c) $\mathbb{N} \models \forall x \forall z \exists y Y(x,y) \geq z$

(d) *For any provably total function* $f(x)$ *there exists a natural number* n *such that:*

$$\mathsf{T} \vdash \forall x (f(x) < g_n())$$

where $g_n(x) =$ *least* y *such that* $Y(x,y) \geq n$ *(we also say that* g_n *forms an envelope for* T*-provably total funcions).*

The next tools we need come from the theory of $\alpha-$largeness and of fundamental sequences of ordinals, i.e. of strictly increasing sequence of ordinals whose supremum is a limit ordinal. First of all, we must therefore recall some important notions, such as that of *Cantor's normal form*, relative to the ordinals less than ε_0, i.e. the fact that each ordinal can be written as a finite sum as follows:

(a) 0 is an ordinal.

(b) if $\alpha_0, ..., \alpha_n$ are ordinals and $\alpha_0 \geq ... \geq \alpha_n$, then $\omega^{\alpha_0} + ... + \omega^{\alpha_n}$ is an ordinal (where $\omega^1 = \omega$ and $\omega^0 = 1$ and the ordinal is limit in case $\alpha_n \neq 0$).

(c) $\omega^{\alpha_0} + ... + \omega^{\alpha_n} \geq \omega^{\beta_0} + ... + \omega^{\beta_k}$ if and only if either for some i, $\alpha_i \geq \beta_i$ and for all $j < i, \alpha_j = \beta_j$, or $n > k$ and for all $i \leq k, \alpha_i = \beta_i$.

Then we define the following sort of "predecessor" operation $\{\alpha\}(k)$, due to Ketonen and Solovay.

Definition 49. *Let* $\alpha < \varepsilon_0$ *and* $k < \omega$. *Then:*

(a) $\{0\}(k) = 0$

(b) *If* $\alpha = \omega^{\alpha_0} + ... + \omega^{\alpha_n}$ *and* $\alpha_n = 0$ *(i.e.* α *is a successor), then:*

$$\{\alpha\}(k) = \omega^{\alpha_0} + ... + \omega^{\alpha_{n-1}}$$

(c) *If* $\alpha = \omega^{\alpha_0} + ... + \omega^{\alpha_n}$ *and* $\alpha_n = \beta + 1$ *for some* β, *then*

$$\{\alpha\}(k) = \omega^{\alpha_0} + ... + \omega^{\alpha_{n-1}} + \omega \cdot k$$

(d) *If* $\alpha = \omega^{\alpha_0} + ... + \omega^{\alpha_n}$ *and* α_n *is limit, then:*

$$\{\alpha\}(k) = \omega^{\alpha_0} + ... + \omega^{\alpha_{n-1}} + \omega^{\{\alpha_n\}(k)}$$

This operation gives rise to a sequence:

$$\{\alpha\}(0) < \{\alpha\}(1) < \{\alpha\}(2) < ... < \alpha$$

where $\alpha = sup_{n \in \omega}\{\alpha\}(n)$. Let us see some examples to make this definition intuitively clearer:

(a) $\{\omega\}(k) = \{\omega^1\}(k) = \{\omega^{0+1}\}(k) = \omega^0 \cdot k = k$.

(b) $\{\omega^\omega\}(k) = \omega^{\{\omega\}(k)} = \omega^k$.

(c) $\{\omega^{n+1}\}(k) = \omega^n \cdot k$.

With the writing $\{\alpha\}(n_0, n_1, ..., n_k)$ we abbreviate the iteration:

$$\{...\{\{\alpha\}(n_0)\}(n_1)...\}(n_k)$$

Lastly, we introduce the notion of "$\alpha-$largeness", which generalises the notion of "largeness"that we encountered in the Paris and Harrington theorem. Let $X = n_0 < n_1 < ... < n_k$ be a finite set of numbers. We say that:

(a) X is $1-$large if and only if it contains at least two elements.

(b) X is $\alpha-$large if and only if $X - \{n_0\}$ is $\{\alpha\}(n_1)$-large.

Arguing by induction on α, we see that actually this condition is equivalent to say that $\{\alpha\}(n_1, n_2, ..., n_k) = 0$. Indeed X is $\alpha-$large if and only if $X - \{n_0\}$ is $\{\alpha\}(n_1)-$large, if and only if by induction hypothesis (since $\{\alpha\}(n_1) < \alpha$):

$$\{\{\alpha\}(n_1)\}(n_2, ..., n_k) = 0$$

that is, $\{\alpha\}(n_1, n_2, ..., n_k) = 0$.

For our purposes, it is important to emphasise this result of Ketonen and Solovay (1981).

Theorem 104. *The following is an indicator for models of Peano Arithmetic* PA*:*

$$Y(a, b) = \textit{greatest } c \textit{ such that the interval } [a, b] \textit{ is } \omega_c - \textit{large}$$

where $\omega_0 = \omega$ *and* $\omega_{n+1} = \omega^{\omega_n}$*, although this statement is independent of* PA*:*

$$\forall a \forall c \exists b([a, b] \textit{ is } \omega_c - \textit{large})$$

To continue this exposition, it is necessary to further examine the mechanism of the so-called *fundamental sequences* of ordinals. Hence let us write $\beta \to_n \alpha$ to mean that, for $j_0, ..., j_k \leq n$, we have $\alpha = \{\beta\}(j_0, ..., j_k)$. We will write $\beta \Rightarrow_n \alpha$ in case $j_0 = j_1 = ... = n$.

Lemma 34. *The following basic properties hold:*

(a) if $\beta \Rightarrow_n \alpha$*, for* $n > 0$*, then* $\omega^\beta \Rightarrow_n \omega^\alpha$*.*

(b) If $0 < i < j \leq n$*, then* $\{\beta\}(j) \Rightarrow_n \{\beta\}(i)$*.*

(c) Let us write $\alpha \gg \beta$ *to mean that* $\alpha = \omega^{\gamma_0} + ... + \omega^{\gamma_m}$ *and* $\beta = \omega^{\delta_0} + ... + \omega^{\delta_k}$ *where* $\gamma_0 > \gamma_1 > ... > \gamma_m \geq \delta_0 > \delta_1 > ... > \delta_k$*. Hence if* $\beta > 0$ *and* $\alpha \gg \beta$*, then* $\{\alpha + \beta\}(n) = \alpha + \{\beta\}(n)$*.*

(d) If $\alpha \gg \beta$ *and* $\beta \Rightarrow_n \gamma$*, then* $\alpha + \beta \Rightarrow_n \alpha + \gamma$*.*

(e) If $\alpha < \varepsilon_0$ *and* $n \geq 0$*, then* $\alpha \Rightarrow_n 0$*.*

(f) $\beta \Rightarrow_n \alpha$ *if and only if* $\beta \to_n \alpha$*.*

(g) If $\beta \to_n \alpha$ *and* $0 < n \leq n_0 < ... < n_k$*, then* $\{\beta\}(n_0, ..., n_k) \geq \{\alpha\}(n_0, ..., n_k)$*.*

Proof. We illustrate only a sketch of the proofs of the more complex points:

(a) by induction on β. We can consider just the case $\alpha = \{\beta\}(n)$, since the general case follows arguing once again inductively.

   i. If $\beta = 0$, this is immediate.

ii. If β is limit, then notice that $\{\omega^\beta\}(n) = \omega^{\{\beta\}(n)} = \omega^\alpha$.

iii. If β is a successor, and then $\beta = \alpha + 1$, observe that for $k < l < \omega$, we have $\omega^\beta \cdot l \Rightarrow_n \omega^\beta \cdot k$. Indeed, by point 5. $\omega^\beta \cdot (l-k) \Rightarrow_n 0$; hence $\omega^\beta \cdot l = \omega^\beta \cdot k + \omega^\beta \cdot (l-k) \Rightarrow_n \omega^\beta \cdot k$ by point 4. Lastly:

$$\omega^\beta = \omega^{\alpha+1} \Rightarrow_n \{\omega^{\alpha+1}\}(n) = \omega^\alpha \cdot n \Rightarrow_n \omega^\alpha$$

(b) By transfinite induction, using points 1. and 4.

(c) It follows from the fact that, under the hypotesis of the lemma, the sum $\alpha + \beta$ is just the concatenation of the respective normal forms. Hence, look at the index δ_k.

(d) This is an application of the previous point: notice that if $\gamma = \{\beta\}(n)$, then $\{\alpha+\beta\}(n) \Rightarrow_n \alpha + \{\beta\}(n) = \alpha + \gamma$, hence $\alpha + \beta \Rightarrow_n \alpha + \gamma$.

(e) By induction on α:

i. $\alpha = 0$, immediate.

ii. If $\alpha = \gamma + 1$, then $\{\gamma + 1\}(n) = \gamma$, namely $\gamma + 1 \Rightarrow_n \gamma$. But by the inductive hypothesis $\gamma \Rightarrow_n 0$.

iii. If α is limit, say in Cantor normal form $\alpha = \omega^{\gamma_0} + ... + \omega^{\gamma_m}$; being limit, $\gamma_m \neq 0$. Suppose $\gamma_m = \sigma + 1$; hence:

$$\{\omega^{\gamma_0} + ... + \omega^{\gamma_m}\}(n) = \omega^{\gamma_0} + ... + \omega^{\gamma_{m-1}} + \{\omega^{\delta+1}\}(n)$$

Now apply twice the inductive hypothesis and point 4. The case of γ_m limit is analogous.

(f) Follows from point 2. and is left as an exercise.

(g) Induction on β. If $\beta = 0$ it is clear. Otherwise, assume the result holds below β and observe that if $\beta \to_n \alpha$ and $0 < n \leq n_0 < ... < n_k$, then $\beta \to_{n_0} \alpha \to_{n_0} \{\alpha\}(n_0)$, from which follows $\{\beta\}(n_0) \to_{n_0} \{\alpha\}(n_0)$. By inductive hypothesis:

$$\{\{\beta\}(n_0)\}(n_1, ..., n_k) \to_{n_0} \{\{\alpha\}(n_0)\}(n_1, ..., n_k)$$

which is our desired result.

QED

The next operator we are going to introduce is another "predecessor" function (sometimes called *Goodstein predecessor*) denoted $\langle\alpha\rangle(n)$. It defines a sequence of ordinals converging to α, but faster than $\{\alpha\}(n)$, in the sense that:

$$\langle\alpha\rangle(n) \to_n \{\alpha\}(n)$$

Definition 50. *Let $\alpha < \varepsilon_0$. Then:*

(a) $\langle 0\rangle(n) = 0$.

(b) $\langle\alpha + 1\rangle(n) = \alpha$.

(c) $\langle\omega^\delta(\alpha + 1)\rangle(n) = \omega^\delta \cdot \alpha + \omega^{\langle\delta\rangle(n)} + \langle\omega^{\langle\delta\rangle(n)}\rangle(n)$.

For instance, $\langle 3\rangle(n) = 2$ and $\langle \omega^3(\alpha+1)\rangle(n) = \omega^3 \cdot \alpha + \omega^2 \cdot n + \omega$.

Moreover, with the notation $f_{m,n}(x)$ we will denote the function that takes the representation of m in base n and replaces n with x. In other words, if:

$$n = 2^{2^{2+1}} + 2^{2+1} + 2^1$$

and $x = n+1$, then:

$$f_{m,n}(x) = 3^{3^{3+1}} + 3^{3+1} + 3^1$$

and in general, if $m = a_0 \cdot n^0 + a_1 \cdot n^1 + ... + a_k \cdot n^k$ then:

$$f_{m,n}(x) = a_0 \cdot x^{f_{0,n}(x)} + a_1 \cdot x^{f_{1,n}(x)} + ... + a_k \cdot x^{f_{k,n}(x)}$$

In terms of the previous notation, for $m > 0$, $G_n(m) = f^{m,n}(n+1) - 1$; we call instead $o_m(n)$ the number $f_{m,n}(\omega)$. These notions are related as follows.

Theorem 105. *(a) for $m \geq 0$, $n > 1$, if $o_{n+1}(m) = \alpha$, then $o_{n+1}(m \dot{-} 1) = \langle \alpha \rangle(n)$.*

(b) For $n > 1$, $\langle o_n(m)\rangle(n) = o_{n+1}(G_n(m))$.

Proof. (a) If $m = 0$ this is obviuos, so let us consider the base $n+1$ representation of $m > 0$:

$$m = \sum_{i=0}^{p} a_i (n+1)^{f_{i,n+1}}$$

where $a_i \leq n$. Let j be the minimum index such that $a_j \neq 0$. Notice that if $j = 0$ the result is immediate, hence let us assume thsat $j > 0$ and assume by inductive hypothesis that the result holds for al $s < m$. Observe that $o_{n+1}(m \dot{-} 1)$ is the sum of the following addends:

(i) $(\sum_{i=j+i}^{p} \omega^{f_{i,n+1}(\omega)} a_i) + \omega^{f_{j,n+1}(\omega)}(a_j - 1)$

(ii) $o_{n+1}(n \cdot (n+1)^{f_{j,n+1}(n+1)-1})$

(iii) $o_{n+1}((n+1)^{f_{j,n+1}(n+1)-1} - 1)$

On the other hands, $\langle \alpha \rangle(n)$ is the sum of (i) plus:

(iv) $\omega^{\langle f_{j,n+1}(\omega)\rangle(n)} \cdot n$

(v) $\langle \omega^{\langle f_{j,n+1}(\omega)\rangle(n)} \rangle(n)$

We see that actually (iii)=(v) and (ii)=(iv):

(a) (iii)=(v). By induction hypothesis (iii) is equal to $\langle o_{n+1}((n+1)^{f_{j,n+1}(n+1)-1}\rangle)(n)$ and this is (v) by definition.

(b) (ii)=(iv). By definition (ii) is $\omega^{f_{j,n+1}(\omega)-1} \cdot n$, namely to $\omega^{o_{n+1}(f_{j,n+1}(n+1)-1}$. Still by the inductive hypothesis this is equal to (iv).

(b) The proof is along the same lines and we omit it.

QED

By using this notation, if for instance $n_0, n_1, n_2...$ is the above mentioned Goodstein sequence, then it correspond to a sequence of ordinals in Cantor normal form:

$$o_n(n_0), o_{n+1}(n_1), o_{n+2}(n_2)...$$

where $o_n(n_0) = \omega^{\omega^{\omega+1}} + \omega^{\omega+1} + \omega$, that is:

$$\omega^{\omega^{\omega+1}} + \omega^{\omega+1} + \omega, \omega^{\omega^{\omega+1}} + \omega^{\omega+1} + 2, \omega^{\omega^{\omega+1}} + \omega^{\omega+1} + 1...$$

that can also be written as:
$\omega^{\omega^{\omega+1}} + \omega^{\omega+1} + \omega$, $\langle \omega^{\omega^{\omega+1}} + \omega^{\omega+1} + \omega \rangle(n)$, $\langle \omega^{\omega^{\omega+1}} + \omega^{\omega+1} + \omega \rangle(n, n+1)$, $\langle \omega^{\omega^{\omega+1}} + \omega^{\omega+1} + \omega \rangle(n, n+1, n+2)$...

We therefore come to a central result for our purpose.

Theorem 106. *Let $n_0, n_1, n_2....$ a Goodstein sequence for m starting at n and let k be the minimum such that $n_k = 0$. Then the interval $[n-1, n+k-1]$ is $o_n(m)$-large.*

Proof. Let us consider the cooresponding sequence of ordinals:

$$o_n(m) = o_n(n_0) = \alpha$$

$$o_{n+i}(n_i) = \langle \alpha \rangle(n, n+1, .., n+i-1)$$

$$\vdots$$

$$o_{n+k}(n_k) = \langle \alpha \rangle(n, n+1, .., n+k-1) = 0$$

From the relationships previously shown between the two predecessor operators and the result on p.166, point 6., it can be seen that:

$$\begin{aligned} &\{\alpha\}(n, n+1, ..., n+k-1) \leq \\ &\leq \{\langle \alpha \rangle(n)\}(n+1, .., n+k-1) \\ &\leq \{\langle \alpha \rangle(n, n+1)\}(n+2, .., n+k-1) \\ &\vdots \\ &\leq \langle \alpha \rangle(n, n+1, ..., m+k-1) = 0 \end{aligned}$$

and therefore $[n-1, n+k-1]$ is α-large. QED

We remark that this proof *can be formalised and carried out in* PA. We now want to apply this result to show that the statement (true and formalisable in the language of arithmetic):

> (*) "for each m and each n, the Goodstein sequence for m starting at n eventually hits zero".

is not provable in PA. Suppose by contradiction it is actually provable. As a consequence of the independence result of Ketonen and Solovay (theorem 3) mentioned at the beginning of the paragraph and some mathematics for the 'indicators', we can find a nonstandard model $\mathcal{M}$ of PA and a non standard element c of this model in which the statement:

> (**) "there exists an element b such that the interval $[1, b]$ is ω_c-large".

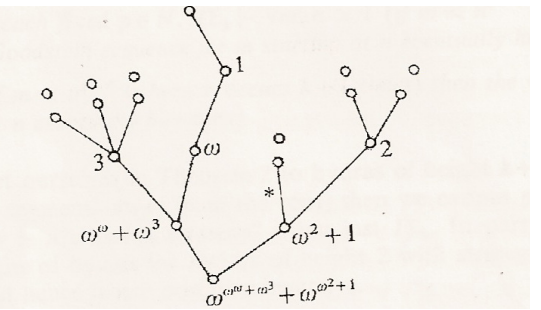


Figure 8. Assignment of ordinals to the Hydra (from Kirby and Paris paper).

is false. Hence take $d = 2^{2^{2}}$ with c iteration of the exponentiation, so that $o_2(d)$ is just ω_c. By (*), that, as a theorem, is true in all models, we can take an element e of the model $\mathcal{M}$ such that $d_e = 0$. By the theorem 106 and the subsequent remark, we have in $\mathcal{M}$ that $[1, 2 + e - 1]$ is ω_c large, but this contradicts (**). Hence (*) is not provable in PA.

Let us finally return to our Hydra and show that the true statement:

(***) "Every recursive strategy for Hercules is a winning strategy"

is not provable in PA. We begin by assigning the Hydra nodes an ordinal less than ε_0 according to this criterion:

(a) To the leafs, assign 0.

(b) To each other node assign $\omega^{\alpha_0} + ... + \omega^{\alpha_n}$ where $\alpha_0 \geq ... \geq \alpha_n$ are the ordinals assigned to the sons of that node. The ordinal of the Hydra is the ordinal assigned to the root.

We now consider the strategy τ based on this algorithm: starting from the root, go to the son (i.e. the immediate successor) labelled with the smallest ordinal among all the sons; continue in the next nodes in the direction of the leaves, always moving towards the immediate successor labelled with the smallest ordinal among the immediate successors, following this criterion until you reach a head. When you reach it, cut it off.

If we denote $\tau(\alpha, n)$ the operator that following the strategy τ, maps the ordinal α at the root of the hydra after the stage $n - 1$, to the ordinal at the root of the Hydra after the stage n, then it is clear that $\tau(\alpha, n) < \alpha$ and therefore Hercules eventually wins.

However, it occurs that this fact cannot be proved in PA. It is in fact the case that:

$$\tau(\alpha, n) = \{\alpha\}(n + 1)$$

Hence a proof of the unprovability of (***) follows the line of the analogous proof for (*). Actually the proof that τ is a winning strategy is equivalent to the ε_0-induction, with $\{\alpha\}(n)$ as predecessor function, that in turn is equivalent to the fact (consequently unprovable in PA) that the function:

$$\lambda x \lambda v.g_v(x) = min.y \geq x \; such \; that \; [x, y] \; is \; \omega_v - large$$

is provably total in PA. A proof-theoretic method for proving incompleteness of PA is indeed based on the classification of provably total functions in PA (see Schwichtenberg, Wainer (2011)).

The proof we have illustrated does indeed appear rather complex, which is why it was felt necessary to find alternative proofs. With this goal, a proof of Goodstein's theorem using

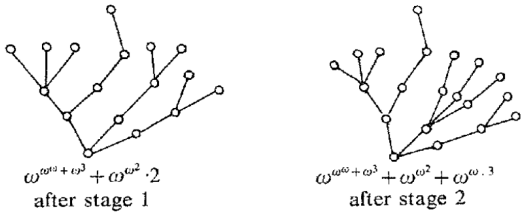


Figure 9. Assignment of ordinals to the Hydra (from the Kirby and Paris paper)

methods from computability theory, rather than model theory, was given for instance in Cichon (1983), which connects the problem of the provability of Goodstein's theorem to well-known results on the recursion theoretic hierarchies of functions, showing that if the function:

$$G(a) = \textit{least } k \textit{ such that } a_k = 0$$

where $a = a_0, a_1.a_2....$ is a Goodstein sequence, were provably total in PA, then so would be the Hardy function H_{ε_0}. Recall that the Hardy hierarchy $\{H_\alpha\}_{\alpha \leq \varepsilon_0}$ is defined as follows:

(a) $H_0(n) = n$

(b) $H_{\alpha+1}(n) = H_\alpha(n+1)$

(c) $H_\lambda(n) = H_{\{\lambda\}(n)}(n)$ for λ limit.

Adding the clauses $\{\varepsilon_0\}(0) = \omega$ and $\{\varepsilon_0\}(n+1) = \omega^{\{\varepsilon_0\}(n)}$ to our previous definition of this operator. Now, it is provable that H_{ε_0} majorizes all functions provably total in PA but it is not itself provably total in this theory. The relationship that exists between the study of the so-called "fast growing functions"and the themes of this chapter has been the subject of extensive in-depth analysis in Buchholz and Wainer (1987) and Ketonen and Solovay (1981).

Many scientific contribution which is worth mentioning are indeed related to the ordinal analysis of theories and its connection to provably total functions. Actually these proof-theoretic works are very difficult and would require further mathematical premises such that their detailed analysis go beyond the scope of this book.

As we have repeatedly recalled in this book, Kreisel (1952) showed that the functions provably total (called also *provably recursive* or *provably computable*) in Peano Arithmetic are those definable by recursion over well-ordering of order-type less than ε_0. At the origin of everything there are Gentzen's 1934 - 1939 consistency proofs for PA, obtained by adding the transfinite induction principle up to ε_0 for primitive recursive (more precisely, the *elementary computable*) predicates to an acceptable finitistic basis, so that we say that the proof-theoretic ordinal of Peano Arithmetic is ε_0.

6.4. Further developments and guide for further study

Beklemishev (2006) proposed a variation of the Hydra game, called 'Worm battle' (see Carlucci (2005) to understand the connection with other works mentioned here). A system of ordinal notation emerges amazingly in this context from a certain system of propositional modal logic. Beklemishev's very original work proposes an algebraic approach to proof-theoretic analysis based on the notion of graded provability algebra, that is, Lindenbaum boolean algebra of a theory enriched by additional modal operators. Another modified version of the Hydra Game is due to Buchholz (1987), that extended Kirby-Paris' Hydra Game by defining a game on labeled finite trees (following a suggestion from Martin Gardner) in which a hydra grew not only in width but also in height and that is independent of a certain strong subsystem

of analysis. A proof of the independence result of a restricted Buchholz style-Hydra Game of certain subsystems of analysis is given also in Hamano and Okada (1998). Both, the Kirby-Paris Hydra as well as the Buchholz type of Hydras have been studied in the context of one of the most important research topic within Proof Theory, namely *ordinal analysis*, which started with Gentzen's often mentioned work on the consistency of arithmetic. In Proof Theory, the so-called "ordinal analysis" assigns tranfinite ordinals to mathematical theories as a measure of their consistency strength or computational power (see Wolfram Pohlers (1993), Rathjen (2006) and Arai (2020)). The proof-theoretic ordinal of a theory can be also defined as the smallest ordinal that this theory cannot prove to be well-founded and can be seen also as a measure of the system's ability to prove the totality of computable functions.

If we say that a Hydra has ordinal strength α if a proof of its termination requires a theory with ordinal strength at least α, then in in this sense, the Kirby-Paris Hydra, has ordinal strength ε_0, while the ordinal strength of the Buchholz Hydra exceeds even all the ordinals expressible by the so-called Buchholz's ψ ordinal functions (see Endrullis, Klop and Overbeek (2021)).

A remarkable characteristic of Carlucci (2003) proof-theoretic approach to the Hydra Game is that it uses a very natural interpretation of the derivations in a system of Peano Arithmetic as hydras.

The results of this work are parallel and, although independent, somewhat related to the work of Hamano and Okada (1998), but the author emphasises that the common 'diagrammatical' flavour of the idea of the proofs is made much more evident in his own approach. In particular, no mention of ordinals nor of transfinite hierarchies is made. Carlucci is able to prove that if D' is obtained by a PA-derivation D in sequent calculus by one step of Gentzen's reduction algorithm in his consistency proof of PA, as described in Takeuti (1987), then it is possible to obtain $H(D')$ from $H(D)$ by a finite number of steps of the Hydra Game.

Other examples of so-called "natural independence phenomena", which are considered by most logicians as more natural than the metamathematical incompleteness results first discovered by Gödel, are the powerful *tree theorem* due to Kruskal, as well as its *finite miniaturization* due to Harvey Friedman. These versions of Kruskal's theorem are remarkable from a proof-theoretic point of view because they are not provable in relatively strong logical systems (see e.g. Simpson (1990) and Gallier (1991)). Kruskal's theorem on trees is a classical result of combinatorics with several applications in computer science. The formal system we are interested in here is the second order system called "Arithmetical Transfinite Recursion" ATR_0 and is one of the "big five" systems well known in the area of *Reverse Mathematics* (see the classic treatise Simpson (1999)).

7. Sequent calculus and complexity theory

7.1. Gentzen's formalism of sequents

We now come to the other formalism introduced by Gentzen, namely the sequents calculus (the name, however, is due to Kleene). We want to introduce here the basic concepts and prove the most important result about it, namely cut-elimination. Actually we will not illustrate here the cut elimination for *pure logic*, for which we refer to Girard (1987) or Takeuti (1987), but rather a proof of the *free-cut elimination theorem*, according to which *some* cuts can be eliminated, and of *partial* cut-elimination, i.e. the elimination of cuts on formulas above a given logical complexity, in fact the only results available for *theories* (i.e. logic plus proper axioms and induction rule): we will see that for them *the full cut-elimination is not valid*). Takeuti (1987) only offers a sketch of the proof. The proof we propose was instead presented in Beckmann and Buss (2011) and involves the modification of the definition of anchored and free formulas, also with respect to an earlier version by the first of the two authors. To date, it does not appear in any manual, to our knowledge, and given the originality of the method employed, it seemed appropriate to refer to it. *Free-cut elimination* has important applications in computational complexity. In particular, Buss (1986) applies this result to obtain his "witnessing theorems"in *Bounded Arithmetic*, that is, the important result of characterization of functions computable in polynomial time that we will discuss in the last chapter: for this, we need only to be able to restrict cut formulas to lie in a given complexity class. Like the proof provided by the first author in Buss (1998) for pure logic, this too differs from the various proofs in the scientific literature, starting with Gentzen's, in that it is of the global kind, that is, it is not based on *local* transformations to a proof to reduce measures of complexity as the depth of cuts, the number of cuts, or the so-called rank of a cut. At the opposite here the depth or number of cuts are reduced by making *global* transformations to a proof.

Unlike the natural deduction by Gentzen-Prawitz, of which there are few variants in the literature (Fitch, sequential rules, generalized elimination rules ...) the the sequent calculus has a wide range of variations, which we will try in this section of illustrate and motivate. While in natural deduction and axiomatic systems rules apply to formulas, in the sequent calculus they apply to *assertions of derivability* of form:

$$\alpha_0, ..., \alpha_n \Longrightarrow \beta_0, ..., \beta_m$$

which must be read: "from $\alpha_0 \wedge ... \wedge \alpha_n$ it is derivable $\beta_0 \vee ... \vee \beta_m$". A sequent is therefore a contruction of the form:

$$\Gamma \Longrightarrow \Delta$$

where Γ, Δ can be, according to different versions, sets, sequences, or multisets (that is, sets that admit repetition, such as $\{A, A, B\}$, that as a set would be equivalent to $\{A, B\}$) of formulas; the choice of which data structure to prefer has one immediate consequences in the formulation of the rules. We use the longest arrow $\Longrightarrow$ to denote this derivability. We remark

Duccio Pianigiani, University of Siena, Italy, duccio.pianigiani@unisi.it, 0000-0001-9441-7226

Referee List (DOI 10.36253/fup_referee_list)
FUP Best Practice in Scholarly Publishing (DOI 10.36253/fup_best_practice)

Book References DOI 10.36253/979-12-215-0778-2.references

that the big arrow is a *metalinguistic* symbol. A sequent $\alpha_0, ..., \alpha_n \Longrightarrow \beta_0, ..., \beta_m$ has to be therefore intended as the formula $\alpha_0 \wedge ... \wedge \alpha_n \rightarrow \beta_0 \vee ... \vee \beta_m$, where:

(a) $\alpha_0, ..., \alpha_n \Longrightarrow$ must be read $\neg(\alpha_0 \wedge ... \wedge \alpha_n)$.

(b) $\Longrightarrow \beta_0, ..., \beta_m$ has to be read as $\beta_0 \vee ... \vee \beta_m$.

(c) "$\Longrightarrow$" has to be read as $\alpha \wedge \neg\alpha$; to prove the consistency of this calculus, therefore, means just to prove the unprovability of "$\Longrightarrow$".

(d) to prove a formula α means to prove the sequent $\Longrightarrow \alpha$.

(e) Unlike of natural deduction, this calculus has no elimination rules and introduction rules, but only introduction rules, right and left: the only one way to delete a connective or a quantifier, is to delete the entire formula where is contained, by means of a rule called CUT.

(f) The calculus is not subject to certainty typical asymmetries of natural deduction for full language, which we can find for example in the rule of elimination of $\vee$.

(g) Moerover it has a further collection of extremely important rules called structural rules; unlike the natural Gentzen-Prawitz deduction, this calculus has axioms.

Let us consider the propositional calculus PK defined in this way (see Buss (1998)). Suppose the sequents are made up of *sequences* of formulas.

Logical axioms $\alpha \Longrightarrow \alpha$ (where α is atomic).

Logical rules

$$\frac{\Gamma \Longrightarrow \Delta, \alpha}{\neg\alpha, \Gamma \Longrightarrow \Delta} \qquad \frac{\alpha, \Gamma \Longrightarrow \Delta}{\Gamma \Longrightarrow \Delta, \neg\alpha}$$

$$\frac{\Gamma \Longrightarrow \Delta, \alpha \quad \beta, \Gamma \Longrightarrow \Delta}{\alpha \rightarrow \beta, \Gamma \Longrightarrow \Delta} \qquad \frac{\Gamma, \alpha \Longrightarrow \Delta, \beta}{\Gamma \Longrightarrow \Delta, \alpha \rightarrow \beta}$$

$$\frac{\Gamma, \alpha, \beta \Longrightarrow \Delta}{\Gamma, \alpha \wedge \beta \Longrightarrow \Delta} \qquad \frac{\Gamma \Longrightarrow \Delta, \alpha \quad \Gamma \Longrightarrow \Delta, \beta}{\Gamma \Longrightarrow \Delta, \alpha \wedge \beta}$$

$$\frac{\Gamma, \alpha \Longrightarrow \Delta \quad \beta, \Gamma \Longrightarrow \Delta}{\alpha \vee \beta, \Gamma \Longrightarrow \Delta} \qquad \frac{\Gamma \Longrightarrow \Delta, \alpha, \beta}{\Gamma \Longrightarrow \Delta, \alpha \vee \beta}$$

The reader who already knows Natural Deduction can think that left rules correspond to elimination rules and the right rules correspond to introduction rules.

To the logical rules, we must add the structural rules:

Exchange

$$\frac{\Gamma, \alpha, \beta, \Pi \Longrightarrow \Delta}{\Gamma, \beta, \alpha, \Pi \Longrightarrow \Delta} \qquad \frac{\Delta \Longrightarrow \Gamma, \alpha, \beta, \Pi}{\Delta \Longrightarrow \Gamma, \beta, \alpha, \Pi}$$

Contraction

$$\frac{\alpha, \alpha, \Gamma \Longrightarrow \Delta}{\alpha, \Gamma \Longrightarrow \Delta} \qquad \frac{\Gamma \Longrightarrow \Delta, \alpha, \alpha}{\Gamma \Longrightarrow \Delta, \alpha}$$

Weakening

$$\frac{\Gamma \Longrightarrow \Delta}{\alpha, \Gamma \Longrightarrow \Delta} \qquad \frac{\Gamma \Longrightarrow \Delta}{\Gamma \Longrightarrow \Delta, \alpha}$$

Observe that if the calculus is formulated in terms of *sets*, then *exchange* and *contraction* are superfluous, since $\{A, A\} = \{A\}$ e $\{A, B\} = \{B, A\}$ are properties of sets; if on the contrary is formulated in terms of multisets $\{A, A\} \neq \{A\}$ and only the exchange is not necessary; lastly, if the axioms are formulated as: $\Theta \Longrightarrow \Psi$, where $\Theta \cap \Psi \neq \emptyset$ (e.g. $\alpha, \Gamma \Longrightarrow \alpha, \Sigma$), then the weakening rule is superfluous.

Lastly we have the (fundamental) rule of CUT:

$$\frac{\Gamma \Longrightarrow \Theta, \alpha \quad \alpha, \Gamma \Longrightarrow \Theta}{\Gamma \Longrightarrow \Theta}$$

The CUT rule can be intuitively interpreted in this way: divide the derivation of Θ from Γ into two *lemmas* which are subsequently reunited. A sequent calculus is *closed for cut*, if for any derivation there is another derivation of the same sequent, which does not make use of CUT. The sequences $\Gamma, \Delta, \Pi, \Theta...$ in the above rules are called *cedents*, whose formulas are called *side formulas*; in the sequent $\Gamma \Longrightarrow \Delta$, the sequence Γ is the *antecedent*, and Δ is the *consequent*. In the rules:

$$\frac{S_0 ... S_n}{S}$$

the sequents $S_0, ..., S_n$ are called *upper* sequents S is the *lower* sequent. In the conclusion of a rule, the formula that does not belong to cedents constitutes the *principal* formula (in axioms $\alpha \Longrightarrow \alpha$, both α are considered principal), while the formulas (not belonging to the cedents) of the premises, from which derives the principal formula, are called *active* (sometimes called *auxiliary* of the principal).

The first order calculus LK is obtained but adding to PK the following rules.

Universal $\forall : left$ and $\forall : right$ rules:

$$\frac{\phi(t), \Gamma \Longrightarrow \Delta}{\forall x \phi(x), \Gamma \Longrightarrow \Delta} \qquad \frac{\Gamma \Longrightarrow \Delta, \phi(y)}{\Gamma \Longrightarrow \Delta, \forall x \phi(x)}$$

Existential $\exists : left$ and $\exists : right$ rules:

$$\frac{\phi(y), \Gamma \Longrightarrow \Delta}{\exists x \phi(x), \Gamma \Longrightarrow \Delta} \qquad \frac{\Gamma \Longrightarrow \Delta, \phi(t)}{\Gamma \Longrightarrow \Delta, \exists x \phi(x)}$$

In $\forall : right$ and $\exists : left$ the *eigenvariable* $y \notin FVar(\Gamma, \Delta)$.

Definition 51. *Ancestors and descendents:*

(a) If ϕ is side and occours in a cedent Γ of an upper sequent, then the immediate descendent *of ϕ is the correspondent occurrence ϕ in the same position of the correspondent cedent Γ in the lower sequent.*

(b) In the exchange rule, say between ϕ and ψ, the immediate descendants of these ϕ and ψ, are still the ϕ and ψ in the lower sequent.

(c) If ϕ is active (auxiliary) in a rule that is not exchange or cut, then the immediate descendent is the principal formula.

(d) We say that ϕ is an immediate ancestor *of ψ if and only if ψ is an immediate descendent of ϕ. Formulas in the initial sequents and of a weakening have non immediate ancestors.*

(e) The cut formula has no descendants in an application of CUT.

(f) ϕ is a descendant *of ψ iff there is a chain of length ≥ 0 (reflexive and transitive closure) of immediate descendant from ψ to ϕ. Analogously we define the ancestor relation as the reflexive and transitive closure of the relation of immediate ancestor.*

(g) ϕ *is a* direct immediate descendant *of* ψ *(analogously* immediate direct ancestor*), iff it is an immediate descendant and* $\psi = \phi$.

We would like to briefly highlight the role and effect of the structural rules. Admitting the weakening rule, actually we admit in fact the *a fortiori* principle:

$$\dfrac{\dfrac{\dfrac{\alpha \Longrightarrow \alpha}{\alpha, \beta \Longrightarrow \alpha}}{\alpha \Longrightarrow (\beta \to \alpha)}}{\Longrightarrow \alpha \to (\beta \to \alpha)}$$

As well as, analogously, admitting the Contraction rule, we admit the law of absorption:

$$\dfrac{\dfrac{\alpha \Longrightarrow \alpha \quad \dfrac{\alpha \Longrightarrow \alpha \quad \beta \Longrightarrow \beta}{\alpha \to \beta, \alpha \Longrightarrow \beta}}{(\alpha \to (\alpha \to \beta)), \alpha, \alpha \Longrightarrow \to \beta}}{(\alpha \to (\alpha \to \beta)), \alpha \Longrightarrow \beta}}{\Longrightarrow (\alpha \to (\alpha \to \beta)) \to (\alpha \to \beta)}$$

Remember that these principles are *not* accepted in some non classical logic. For example the *a fortiori* is not accepted by the Relevant Logic, absorption is not accepted in the infinite-valent logics of Łukasiewicz etc. Therefore many formalizations in terms of sequent of these logics avoid some or all of the structural rules. This introduces to the topic of *substructural* logics (i.e. with limitation or absence of structural rules) and of Linear Logic. In the classical *propositional* calculus (unlike the intuitionist one), the contraction rule is actually redundant. Ketonen and Solovay (1981) showed instead that the classical *predicates* calculus without the contraction rule is decidable.

There are various reasons for formulating the rules as in PK, for example in this form they are invertible. A rule is called invertible in a sequent calculus system of a proof of its conclusion implies the existence of proofs of each of its premises. Conversely, to make those on quantifiers invertible, they must be formulated as follows:

Universal

$$\dfrac{\phi(t), \forall x \phi(x), \Gamma \Longrightarrow \Delta}{\forall x \phi(x), \Gamma \Longrightarrow \Delta} \qquad \dfrac{\Gamma \Longrightarrow \Delta, \phi(y)}{\Gamma \Longrightarrow \Delta, \forall x \phi(x)}$$

In the right rule, $y \notin FVar(\Gamma, \Delta)$ and if $y \neq x$, $y \notin FVar(\phi)$.

Existential

$$\dfrac{\phi(y), \Gamma \Longrightarrow \Delta}{\exists x \phi(x), \Gamma \Longrightarrow \Delta} \qquad \dfrac{\Gamma \Longrightarrow \Delta, \exists x \phi(x), \phi(t)}{\Gamma \Longrightarrow \Delta, \exists x \phi(x)}$$

In the left rule, $y \notin FVar(\Gamma, \Delta)$ e, se $y \neq x$, $y \notin FVar(\phi)$.

Note that in the rules above, the main formula has been repeated in the upper sequent, on the same side. The systems G3c and G3i invertible and closed for cut and contraction are essentially based on this idea, although in the intuitionistic case G3i the left-hand rule of implication will also have to undergo a similar modification (see Troelstra and Schwichtenberg (2000)).

Another distinction that needs to be made, which does not coincide with the previous one, is that between additive and multiplicative rules: in the presence of structural rules these two formulations are equivalent, but in their *absence* they introduce, on the contrary, connectives with distinct meaning. For example (see Girard, Lafont and Taylor (1989)), regarding

conjunction and disjunction, we have a pair of *multiplicative* connectives $\otimes$ (times) and $\wp$ (par) with rules:

$$\frac{\Gamma \Longrightarrow \Theta, \alpha \quad \Sigma \Longrightarrow \Pi, \beta}{\Gamma, \Sigma \Longrightarrow \Theta, \Pi, \alpha \otimes \beta} \qquad \frac{\Gamma, \alpha, \beta \Longrightarrow \Delta}{\Gamma, \alpha \otimes \beta \Longrightarrow \Delta}$$

$$\frac{\Gamma, \alpha \Longrightarrow \Theta \quad \Sigma, \beta \Longrightarrow \Pi}{\Gamma, \Sigma, \alpha \wp \beta \Longrightarrow \Theta, \Pi} \qquad \frac{\Gamma \Longrightarrow \Delta, \alpha, \beta}{\Gamma \Longrightarrow \Delta, \alpha \wp \beta}$$

and *additive* connectives & (and) e $\oplus$ (plus):

$$\frac{\Gamma \Longrightarrow \Theta, \alpha \quad \Gamma \Longrightarrow \Theta, \beta}{\Gamma \Longrightarrow \Theta, \alpha \& \beta} \qquad \frac{\Gamma, \alpha \Longrightarrow \Delta}{\Gamma, \alpha \& \beta \Longrightarrow \Delta} \qquad \frac{\Gamma, \beta \Longrightarrow \Delta}{\Gamma, \alpha \& \beta \Longrightarrow \Delta}$$

$$\frac{\Gamma, \alpha \Longrightarrow \Theta \quad \Gamma, \beta \Longrightarrow \Gamma}{\Gamma, \alpha \oplus \beta \Longrightarrow \Theta} \qquad \frac{\Gamma \Longrightarrow \Delta, \alpha}{\Gamma \Longrightarrow \Delta, \alpha \oplus \beta} \qquad \frac{\Gamma \Longrightarrow \Delta, \beta}{\Gamma \Longrightarrow \Delta, \alpha \oplus \beta}$$

In the presence of the structural rules, the two previous formulations are equivalent. In the presence of only the *Weakening*, $\otimes$ is stronger than &:

$$\frac{\dfrac{\dfrac{\alpha \Longrightarrow \alpha}{\alpha, \beta \Longrightarrow \alpha} \quad \dfrac{\beta \Longrightarrow \beta}{\alpha, \beta \Longrightarrow \beta}}{\alpha, \beta \Longrightarrow \alpha \& \beta}}{\alpha \otimes \beta \Longrightarrow \alpha \& \beta}$$

To show the equivalence we need *Contraction*[1].

Gentzen's original LK for classical logic is actually a little bit different from ours: also in that case sequents are sequences of formulas, axioms are of the form $\alpha \Longrightarrow \alpha$ (α atomic), all structural rules and CUT (in a moltiplicative form) are included, but the right rule for $\vee$ and left rule for $\wedge$ are in the *additive* version.

Sequent calculus for intuitionistic logic. The intuitionistic calculus LJ is obtained by imposing to LK a restriction on the form of sequents, namely that in $\Gamma \Longrightarrow \Delta$, the Δ is authorised to have at most oneformula. From the restriction on the shape of the sequents it follows automatically others on the form of the rules: for example *Exchange* right and *Contraction* right will be deleted and *Weakening* rigth will have the form:

$$\frac{\Gamma \Longrightarrow}{\Gamma \Longrightarrow \phi}$$

It should be noted that the additive intuitionist rules, in the presence of the permitted structural rules, are not equivalent to the corresponding multiplicatives, being implied by the latter, but not implying. The intuitionist structural rules allow to demonstrate the equivalence between multiplicative formulation and additive of the rules for the $\wedge$, but not for $\vee$. Ultimately, the calculs LJ has for $\vee$ the additive form and for $\rightarrow$ the multiplicative form.

Let's see a consequence of these restriction, considering a sequent $\Longrightarrow \alpha \vee \beta$ obtained without CUT: what is the last rule applied? It cannot be *Weakening*, because otherwise we would have previously had a derivation of $\Longrightarrow$, the empty sequence, i.e. a contradiction; but, as we shall see, this is not possible, since from the cut-elimination theorem it follows the consistency of the calculus, and in particular of LJ. It cannot be a right *Contraction* because otherwise we had $\Longrightarrow \alpha \vee \beta, \alpha \vee \beta$ as upper sequent, which is not an intuitionistic sequent. Then the

[1] Another important variant for the classical or for linear calculus which should be mentioned is the *one-side* version, based on the following argument: in a sequent $\alpha_0, ..., \alpha_n \Longrightarrow \beta_0, ..., \beta_m$ bringing all the formulas on the right (see the rules on negation) we obtain a sequent of the form: $\Longrightarrow \neg\alpha_0, ..., \neg\alpha_n \beta_0, ..., \beta_m$ that can be written simply as $\neg\alpha_0, ..., \neg\alpha_n, \beta_0, ..., \beta_m$. Negation cannot be a primitive symbol: rather, to each propositional variable p is associated its complement $\overline{p}$ and the inductively $\neg p = \overline{p}$,$\neg(p \wedge q) = (\neg p \vee \neg q)$ etc.

last ruel must be a right logical rule. It follows that in LJ, if $\Longrightarrow \alpha \vee \beta$ is derivable, then either $\Longrightarrow \alpha$, or $\Longrightarrow \beta$ is derivable. In particular it is not derivable $\Longrightarrow \alpha \vee \neg\alpha$: if α is atomic then we would have either $\Longrightarrow \alpha$, or $\Longrightarrow \neg\alpha$ but neither is, for α atomic. For the same reasons, if $\Longrightarrow \exists x\phi(x)$ is derivable in LJ, then $\Longrightarrow \phi(t)$ also is. A correspondence with natural deduction as in 74 is given by considering that proofs in LJ can be (not 1-1) translated in natural deduction derivations in which right rules correspond to introduction and left rules to elimination (see Girard, Lafont and Taylor (1989) pp. 43-49).

We finally arrive at the *Cut elimination.* The so called *Hauptsatz* is a weak normalization theorem, i.e. a *strategy* of normalization (for a discussion on strong normalization results see e.g. Urban and Bierman (2001)). Famous achievements due to Statman and Orevkov say that this algorithm for classical predicate calculus cannot be, in general, efficient. The methods of cut-elimination most frequently traceable in the scientific literature (with significant variants), are in general derived, either from the original one in Gentzen (1935), introduced with the aim of giving a consistency proof for Peano arithmetic (see e.g. Takeuti (1987)), or from Tait (1968). The *Hauptsatz* theorem for pure logic has important consequences, in first place the principle of *subformula*, where this notion is specified as follows:

Definition 52. *this is Gentzen's notion of a dottoformula:*

(a) if $\phi = p$, then the unique subformula of ϕ is p.

(b) If $\phi = \phi_0 \wedge \phi_1$ then the subformulas of ϕ are ϕ the subformulas of ϕ_0 and of ϕ_1.

(c) Analogously for $\phi_0 \vee \phi_1$, $\phi_0 \to \phi_1$.

(d) If $\phi = \neg\phi_0$, the subformulas of ϕ are ϕ and the subformulas of ϕ_0.

(e) If $\phi = \forall x\phi_0$, the subformulas of ϕ are ϕ and the subformulas of $\phi_0(t)$ for all terms t. Analogously for $\exists x\phi_0$.

As for the quantified formulas, notice that, since there are infinite variables, this formulas have infinite subformulas.

Corollary 21. (Principle of the subformula) *In a* cut-free *proof of $\Gamma \Longrightarrow \Delta$, all sequents consist of subformulas of formulas in Γ, Δ.*

Proof. Simply observe that, in the absence of *CUT*, in the proof in question will only need logical rules and structural rules; but in both cases, in the logical rules and in the structural ones, the upper sequents are made up of subformulas of the lower sequents. Remember that only the *CUT* rule eliminates formulas QED

However the cut-elimination theorem *does not hold* (in its general form) if there are specific axioms: a simple counterexample (see Girard (1987)) is the following. Let $\Longrightarrow \phi$ e $\Longrightarrow (\gamma \to \delta)$ sequents that represent proper axioms and consider the derivation:

$$\dfrac{\Longrightarrow \gamma \qquad \dfrac{\Longrightarrow (\gamma \to \delta) \qquad \dfrac{\gamma \Longrightarrow \gamma \quad \delta \Longrightarrow \delta}{\gamma, \gamma \to \delta \Longrightarrow \delta}}{\gamma \to \delta}}{\Longrightarrow \delta}$$

How to obtain a cut-free proof of $\Longrightarrow \delta$? However, there is one "partial" form of this result, also dating back to Gentzen, which we will state. To what extent is it possible to eliminate cuts in $\mathsf{LK} + \Psi$ proofs, where Ψ is a set of initial sequents closed under substitution? For instance, the first order logic with equality is obtained by adding the following sequents:

(a) $\Longrightarrow s = s$

(b) $s = t, t = r \Longrightarrow s = r$

(c) $s = t \Longrightarrow t = s$

(d) $s_0 = t_0, ..., s_k = t_k \Longrightarrow f(s_0, ..., s_k) = f(t_0, ..., t_k)$

(e) $s_0 = t_0, ..., s_k = t_k, P(s_0, ..., s_k) \Longrightarrow P(t_0, ..., t_k)$

The elementary axioms of arithmetic P^- can be expressd by sequents:

(a) $\Longrightarrow x + \overline{0} = x$

(b) $\Longrightarrow x + s(u) = S(x + u)$

(c) $\Longrightarrow x \cdot \overline{0} = \overline{0}$

(d) $\Longrightarrow x \cdot S(u) = x \cdot u + x$

(e) $x < 0 \Longrightarrow$

(f) $u < x \Longrightarrow u < S(x)$

(g) $u = x \Longrightarrow u < S(x)$

(h) $u < S(x) \Longrightarrow u = x, u < x$

(i) $\Longrightarrow u < x, u = x, x < u$

(j) $S(x) = 0 \Longrightarrow$

(k) $S(x) = S(u) \Longrightarrow x = u$

A set of sequents Ψ is *closed under substitution*, iff for all $\Gamma(x) \Longrightarrow \Delta(x)$ in Ψ and all terms t, also $\Gamma[t/x] \Longrightarrow \Delta[t/x]$ is in Ψ. The account offered in Buss (1998) of the free-cut elimination theorem proceeds as follows.

Definition 53. *Let π a proof $\mathsf{LK} + \Psi$; say that a formula of π is* anchored *at a sequent in Ψ, iff it is a direct descendant of a formula occurring in an initial sequent in Ψ. A cut inference is* anchored *iff at least one of the two occurrences of the cut formulas is anchored, and is* free *iff both of these occurrences of the cut formulas in the upper sequents are not anchored (i.e. are* free*). A proof is called* free-cut-free *if does not contain free cuts (i.e. all cuts are anchored).*

Theorem 107. *If Ψ is a set of sequents closed under substitution and there is a proof of $\Gamma \Longrightarrow \Delta$ in $\mathsf{LK} + \Psi$, then there exists a proof free-cut-free of the same sequent in $\mathsf{LK} + \Psi$.*

To conclude, in the applications of sequent calculus to formal arithmetic and to the fragments of arithmetic, it is customary to add an induction rule, that *in this form*:

$$\Phi - \mathsf{IND} = \frac{\phi(x), \Gamma \Longrightarrow \Delta, \phi(x+1)}{\phi(0), \Gamma \Longrightarrow \Delta, \phi(t)}$$

where $\phi \in \Phi$, for a class of formulas Φ, is equivalent to the induction axiom. Let us call $\phi(0)$ e $\phi(t)$ the *principal formulas* of this inference.

If $\mathsf{T} = \mathsf{LK} + \Psi + \Phi - \mathsf{IND}$ is an arithmetical theory formalized in sequent calculus (where Ψ are the specific axioms and both Ψ, Φ are closed under substitution), then an occurrence of a formula in a derivation in T is called *anchored* iff it is *direct descendant* of a formula occourring in a sequent of Ψ, *or a direct descendant of a principal formula of an induction.*

Theorem 108. *The following hold:*

(a) if $T = \mathsf{LK} + \Psi + \Phi - IND$ is a theory of formal arithmetic and Ψ, Φ are closed under substitution and $\Gamma \Longrightarrow \Delta$ follows from T, then exists a free-cut-free proof of the same sequent in T.

(b) If Φ, the class of formulas on which induction is allowed is closed under substitution and subformulas (e.g. $\Sigma_n \cup \Pi_n$) and all all sequents in Ψ is made of formulas from Φ and all formulas in $\Gamma \Longrightarrow \Delta$ are in Φ , then each formula occurring in the proof is in Φ.

7.2. Free-cut elimination: a more recent proof

What we briefly summarised was the version proposed in Buss (1998). We present here the detailed proof in Beckmann and Buss (2011) of the free-cut elimination theorem, according to which any provable sequent can be proved using only cuts in which at least one cut-formula was *anchored*. The aim of this paper was to correct and strengthen the previous upper bounds, after an inaccuracy was noted in the estimates of the size of free cut. This refinement actually involved a slight modification of the previous definition of *anchored* and *free* formulas, as well as the definition of the *depth* of a cut formula. In the above version, a formula was *anchored* if at least one of the places it is introduced is an anchor; cuts in which neither cut formula was anchored were called *free* and and it was shown that that any provable sequent is provable by a proof in which no cuts are free. In the most recent improvement we are showing, every place the formula is introduced is considered to be an anchor. A generic set $\Im$ of axioms or inference rules is actually added to which to anchor the cuts.

Definition 54. *A skeleton consists of a rule with k-hypothesis, for $k \geq 0$:*

$$\frac{\Psi_1, C_1 \Longrightarrow D_1, \Xi_1, ..., \Psi_k, C_k \Longrightarrow D_k, \Xi_k}{\Psi, C \Longrightarrow D, \Xi}$$

where cedents Ψ, Ξ contain the principal formulas and the cedents Ψ_i, Ξ_i contain the auxiliary formulas, C, C_i, D, D_i are metavariables for cedents that contain the side formulas. If $k = 0$, there are no upper sequents.

Moreover we have:

(a) side formulas indicators $s_1, ..., s_k \in \{0, 1\}$ indicating which hypothesis have side formulas.

(b) Lastly, we possibly have eigenvariables $a_1, ..., a_k$, that may appear each in exactly one upper sequent.

An *instance* of a skeleton is obtained as follows: let Γ, Δ be any cedents not containing *eigenvariables*; if $C = \Gamma$ and $D = \Delta$, then for each $i \leq k$, if $s_i = 1$, then $C_i = \Gamma$ and $D_i = \Delta$; if $s_i = 0$, then C_i, D_i are empty.

A set of inferences $\Im$ is *acceptable*, provided it is the union of all instances of some set of skeletons.

Some examples of acceptable sets of inferences are the following.

(a) Set of non logical axioms (with $k = 0$).

(b) *Induction*, for each $\phi(x)$ arithmetic (or belonging to a specific class as Σ_k), there is a skeleton with (necessarily!) $s_1 = 1$ and $k = 1$:

$$\frac{\phi(b), C_1 \Longrightarrow D_1, \phi(A(b))}{\phi(0), C \Longrightarrow D, \phi(t)}$$

(c) Negri-Von Plato quantifier -free axioms:

$$\frac{q_1, C_1 \Longrightarrow D_1, ..., q_k, C_k \Longrightarrow D_k}{p_1, ..., p_m, C \Longrightarrow D}$$

with q_i, p_j atomic.

(d) Logical rules (see below).

Moreover, the following complexity measures are adopted:

(a) Size of a proof $|\pi|$=total number of non structural inferences, without considering initial sequents.

(b) Height of a proof $h(\pi)$= maximum number of non-structural inferences on any branch , without considering initial sequents.

Ancestors. Let C, C' be two occurrences of the same formula in a proof π. We say that C' is an *immediate direct ancestor* of C, where C' appears in an upper sequent and C in the lower sequent of a logical or $\Im$ inference, if:

(a) C, C' occupy the same position in Γ, Δ of the respective sequents, or

(b) in contraction, they are occurrences of the contracted formula, or

(c) in exchange of ϕ, ψ occurrences C and C' are both ψ or both ϕ.

(d) Principal formulas of weakening, or of logical inferences, or of logical axioms, or formulas in Ψ, Ξ of $\Im$ do not have immediate direct ancestors.

The $\Im$-*depth* of an occurrence C of a formula is defined as follows:

(a) formulas in Ψ, Ξ of $\Im$ have $depth(C) = 0$.

(b) If C is in a logical axiom, $depth(C) = 1$.

(c) If C is in the lower sequent of a structural rule, or it is a side formula of a non structural rule, then:

$$depth(C) = max\{depth(C') | C' \ immediate \ direct \ ancestor \ of \ C\}$$

If a set is empty, then its maximum is defined as $-\infty$.

(d) If C is principal in a non $\Im$ rule and non structural rule, then:

$$depth(C) = 1 + max\{depth(C') | C' \ is \ auxiliary\}$$

(e) The depth of a CUT is the *minimum* of the depths of the cut formulas.

(f) The depth of a proof is the maximum of the depth of its CUT rules.

Anchored cuts. A CUT is *anchored*, if one occurrence C of its cut-formulas has $depth(C) = 0$. A CUT is *free* if either one occurrence of its cut formulas has depth=$-\infty$, or the cut formulas are atoms and one occurrence has depth=1, or it is not anchored.

Note that a *non free cut has depth* 0. A proof is *free cut-free*, if it has no *free cuts*: we are going to prove a *free-cut elimination* theorem.

Definition 55. *Let $\pi' \preccurlyeq \pi$ means that the proofs π, π' have the same endsequent, and each formula occurring in it has depth in π' less or equal than in π.*

Theorem 109. *For each proof π there is a proof π' of the same sequent with no depth $-\infty$ cuts, such that $|\pi'| \leq |\pi|$, and $h(\pi') \leq h(\pi)$ and the depth of π' is less or equal to the depth of π. Furthemore $\pi' \preccurlyeq \pi$.*

We prove a refined form of the theorem: remove an arbitrary set of formulas of depth $-\infty$ from the endsequent of π; then you can get a proof π' of what is left that has no cuts of depth $-\infty$ and such that $|\pi'| \leq |\pi|$ and $h(\pi') \leq h(\pi)$ and the depth of π' is less or equal of the depth of π and $\pi' \preccurlyeq \pi$.

Proof. Induction on $|\pi|$. Let us see the case in which the last rule is:

$$\frac{\Gamma \Longrightarrow \Delta, \alpha \quad \Gamma \Longrightarrow \Delta, \beta}{\Gamma \Longrightarrow \Delta, \alpha \wedge \beta}$$

Claim. Find a proof of $\Gamma' \Longrightarrow \Delta', (\alpha \wedge \beta)'$, the lower sequent of the above inference where some formulas of depth $-\infty$ has been removed and where $(\alpha \wedge \beta)'$ means that this formula has depth $-\infty$ and has been deleted, or is not among the formulas of this depth that have been deleted, or is just $\alpha \wedge \beta$ of depth $\neq -\infty$. QED

(a) In case $(\alpha \wedge \beta)'$, and this formula has depth $-\infty$ and has been deleted and we would give a proof of $\Gamma' \Longrightarrow \Delta'$. Since by convention $1 + (-\infty) = -\infty$, the formulas α and β in the upper sequents have depth $-\infty$. Just apply (IH) to the subproofs π_1 and π_2 of the upper sequents and from any of the resulting proofs you can get π_i' of $\Gamma' \Longrightarrow \Delta'$.

(b) In the other cases, just apply (IH) to the subproofs of the upper sequents to give a proof of $\Gamma' \Longrightarrow \Delta', \alpha \wedge \beta$

Note that in transformations 1. and 2. the depth of the formulas in the endsequent has not been increased. This follows from the definition of "depth" and by (IH).

Il the last rule is CUT:

$$\frac{\Gamma \Longrightarrow \Delta, \alpha \quad \alpha, \Gamma \Longrightarrow \Delta}{\Gamma \Longrightarrow \Delta}$$

By (IH) there are subproofs π_1', π_2' respectively of $\Gamma' \Longrightarrow \Delta', \alpha$ and $\alpha, \Gamma' \Longrightarrow \Delta'$ as required. In case in one of these two subproofs the formula α has depth $-\infty$, just apply once more (IH) to it and obtain a proof of $\Gamma' \Longrightarrow \Delta'$. Otherwise apply CUT to $\Gamma' \Longrightarrow \Delta', \alpha$ and $\alpha, \Gamma' \Longrightarrow \Delta'$ and note that the cut inference has depth $> -\infty$.

If the last rule is a $\Im$ rule, note that only the side formulas in that case may have depth $-\infty$. Hence just apply (IH).

If the last rule is structural, the proof is trivial.

Theorem 110. (Free-cut elimination) *Let π be a proof of depth $\leq d$ for $d \geq 0$. Then another proof π' exists of the same endsequent which contains no free cuts. Moreover:*

(a) $h(\pi') < 2^{h(\pi)+1}_{d+1}$.

(b) $|\pi'| < c^{2^{|\pi|+1}_{d+1}}$, *where c is the maximum of 2 and the maximum arity of $\Im$ inferences in π.*

where $2^k_0 = k$ and $2^k_{m+1} = 2^{2^k_m}$.

The theorem follows from this Lemma.

Lemma 35. *Suppose π ends with a free cut of depth $d \geq 0$ and all other free cuts above have depth $< d$. Then another proof π' exists of the same endsequent, such that all free cuts in π' have depth $< d$.*

Moreover:

(a) $h(\pi') \leq 2 \cdot h(\pi)$ *and* $\pi' \preccurlyeq \pi$.

(b) If the cut is not atomic $|\pi'| \leq |\pi|^2$; *otherwise* $|\pi'| \leq (c-1)|\pi|^2$.

Proof. By induction on the size $|\pi|$. Suppose π ends with the free cut inference:

$$\dfrac{\dfrac{\pi_1}{\Gamma \Longrightarrow \Delta, \alpha} \quad \dfrac{\pi_2}{\alpha, \Gamma \Longrightarrow \Delta}}{\Gamma \Longrightarrow \Delta}$$

where one occurrence of the cut-formula has depth d and the other has depth $\geq d$.

(a) If α is not atomic. We see the possible cases.

(a) $\alpha = \neg\beta$. Notice that, being $d \geq 0 > -\infty$, the formula not being atomic and the cut being free, this by definition means that the cut is actually *not anchored*, and this implies that for both cut formulas actually $d \geq 1$. Find all direct ancestors of $\neg\beta$ in π_1 that have no immediate direct ancestors (the points where this formula originates). These can be:

(i) The principal formula of a Weakening.

(ii) An $\Im$-inference.

(iii) A right rule:

$$\dfrac{\beta, \Pi \Longrightarrow \Lambda}{\Pi \Longrightarrow \Lambda, \neg\beta}$$

In case (iii), if $\neg\beta$ as a cut formula has depth d, in this inference has therefore depth $\leq d$ and β has depth $< d$.

This kind of inference will be replaced by:

$$\dfrac{\dfrac{\beta, \Pi \Longrightarrow \Lambda}{weak + exchange}}{\Pi, \beta \Longrightarrow \Lambda, \neg\beta}$$

Note that $\neg\beta$, being introduced by weakening, has here depth $-\infty$.

By means of structural rules the β in the antecedent of the lower sequent propagates as a side formula down in the proof, using weakening to add it when necessary, so that we get a proof π_1' of:

$$\Gamma, \beta \Longrightarrow \Delta, \neg\beta$$

where, after this transformation, direct ancestors of $\neg\beta$ can originate only from weakening (i) or $\Im-$inferences (ii), since logical axioms are atomic and therefore $\neg\beta$ cannot occur in a logical axiom. Hence in the final sequent of such a proof, by definition of "depth", $\neg\beta$ has depth ≤ 0.

(a) if $depth(\neg\beta) = -\infty$ in the endsequent of π_1', then use the first theorem to get a proof π_1'' of $\Gamma, \beta \Longrightarrow \Delta$.

(b) if $depth(\neg\beta) = 0$ in the endsequent of π_1', then obtain π_1'' as follows:

$$\dfrac{\dfrac{\pi_1'}{\Gamma, \beta \Longrightarrow \Delta, \neg\beta} \quad \dfrac{\neg\beta, \Gamma \Longrightarrow \Delta}{\neg\beta, \Gamma, \beta \Longrightarrow \Delta}}{\Gamma, \beta \Longrightarrow \Delta}$$

The cut formula in the right upper sequent has depth≥ 1 by the hypothesis of the theorem, whereas the left-hand occurrence has depth 0. Thus the cut has depth 0 and in anchored (not free!).

Now make a similar contruction on the right upper sequent of the initial cut inference, transforming π_2 in π_2'', a proof of $\Gamma \Longrightarrow \beta, \Delta$ and observe that if $\neg\beta$ had depth d in the endsequent of π_2, then β will have depth $< d$ in the endsequent of π_2''.

Hence we can cross π_1'' and π_2'' making a cut on β of depth $< d$ and obtaining the desired proof π' of $\Gamma \Longrightarrow \Delta$.

Now observe that:

(a) $h(\pi') = max\{h(\pi_1'') + 1, h(\pi_2'') + 1\}$
$\leq max\{h(\pi_1') + 2, h(\pi_2') + 2, h(\pi_1) + 2, h(\pi_2) + 2\}$
$\leq max\{h(\pi_1) + 2, h(\pi_1) + 2\} = h(\pi) + 1 < 2 \cdot h(\pi)$

(b) $|\pi'| \leq (|\pi_1'| + |\pi_2| + 1) + (|\pi_2'| + |\pi_1| + 1) + 1 \leq 2 \cdot |\pi_1| + 2 \cdot |\pi_2| + 1 < |\pi|^2$

Notice that $|\pi_i'| < |\pi_i|$, due to the removal of at least an introduction of the negation from π_i. From the construction it follow that $\pi' \preccurlyeq \pi$.

(b) $\alpha = \beta \vee \gamma$. How can this formula be generated in π_1 as a principal formula (i.e. as a first ancestor of a sequence of occurrences)? By a logical rule (i) (right introduction of $\vee$):

$$\frac{\Pi \Longrightarrow \Lambda, \beta, \gamma}{\Pi \Longrightarrow \Lambda, \beta \vee \gamma}$$

or by a $\Im-$inference (ii) or by a weakening (iii). First replace all the inferences of case (i.) by:

$$\frac{\dfrac{\Pi \Longrightarrow \Lambda, \beta, \gamma}{weakening + exchange}}{\Pi \Longrightarrow \beta, \gamma, \Lambda, \beta \vee \gamma}$$

and add structural rules to propagate β, γ down in the proof and obtain a proof π_1' of $\Gamma \Longrightarrow \beta, \gamma, \Delta, \beta \vee \gamma$.

Note that if $\beta \vee \gamma$ had depth d in the endsequent of π_1, then the occurrences of β, γ in the endsequent of π_1' have depth $< d$, while the depth of $\beta \vee \gamma$ in the endsequent of π_1' has depth ≤ 0 (i.e. 0 or $-\infty$). Now the depth of $\beta \vee \gamma$ in the endsequent of π_1' is 0 or $-\infty$:

(a) If the depth is $-\infty$ (case (iii.)), then use once more the first theorem to find a proof π_1'' of $\Gamma \Longrightarrow \beta, \gamma, \Delta$.

(b) If this depth is 0 (case ii.), form π_1'' as follows:

$$\frac{\dfrac{\pi_1'}{\Gamma \Longrightarrow \beta, \gamma, \Delta, \beta \vee \gamma} \quad \dfrac{\dfrac{\beta \vee \gamma, \Gamma \Longrightarrow \Delta}{structural\ rules}}{\beta \vee \gamma, \Gamma \Longrightarrow \beta, \gamma, \Delta}}{\Gamma \Longrightarrow \beta, \gamma, \Delta}$$

Note that cut has depth 0.

Now, let us consider π_2: once more, the inferences that originate the first direct ancestors of $\beta \vee \gamma$ in π_2 can be (i) a logical left-introduction of $\vee$, or (ii) an $\Im-$inference, or (iii) a weakening. In case (i) the upper sequents have the form $\beta, \Pi \Longrightarrow \Lambda$ and $\gamma, \Pi \Longrightarrow \Lambda$.

(a) Take the first kind of sequents and obtain by weakening $\beta \vee \gamma, \Pi, \beta \Longrightarrow \Lambda$, then from this, a proof π_2^B of $\beta \vee \gamma, \Gamma, \beta \Longrightarrow \Delta$.

(b) Perform the same transformation taking the second kind of sequents, obtaining a proof π_2^C of $\beta \vee \gamma, \Gamma, \gamma \Longrightarrow \Delta$.

(a) If $\beta \vee \gamma$ in the endsequent of π_2^B has depth 0, obtain $\pi_2'^B$ of $\Gamma, \beta \Longrightarrow \Delta$ by crossing it with π_1 making a cut of depth 0:

$$\dfrac{\dfrac{\dfrac{\Gamma \Longrightarrow \Delta, \beta \vee \gamma}{structural}}{\Gamma, \beta \Longrightarrow \Delta, \beta \vee \gamma} \quad \dfrac{\pi_2^B}{\beta \vee \gamma, \Gamma, \beta \Longrightarrow \Delta}}{\Gamma, \beta \Longrightarrow \Delta}$$

(b) If in the endsequent of π_2^B the formula $\beta \vee \gamma$ has depth $-\infty$, use the first theorem to get a proof $\pi_2'^B$ of $\Gamma, \beta \Longrightarrow \Delta$ with the claimed properties.

Do the same with π_2^C to obtain $\pi_2'^C$ of $\Gamma, \gamma \Longrightarrow \Delta$.

Now, if $\beta \vee \gamma$ in the endsequent of π_2 had depth d, then β has depth $< d$ in $\pi_2'^B$. Analogously with γ in $\pi_2'^C$.

Make therefore two cuts of depth $< d$:

$$\dfrac{\dfrac{\dfrac{\pi_1''}{\Gamma \Longrightarrow \beta, \gamma, \Delta} \quad \dfrac{\dfrac{\pi_2'^B}{\Gamma, \beta, \Longrightarrow \Delta}}{\Gamma, \beta, \Longrightarrow \Delta, \gamma}}{\Gamma \Longrightarrow \gamma, \Delta} \quad \dfrac{\pi_2'^C}{\Gamma, \gamma \Longrightarrow \Delta}}{\Gamma \Longrightarrow \Delta}$$

(Before the last cut, make some exchange). The following hold:

(a) $h(\pi') \leq max\{h(\pi') + 3, h(\pi_2) + 3, h(\pi_2^B) + 3, h(\pi_1) + 3, h(\pi_2^C) + 2\}$
$\leq max\{h(\pi_1) + 3, h(\pi_2) + 3\}$
$\leq h(\pi) + 2 \leq 2 \cdot h(\pi)$.

(b) $|\pi'| \leq |\pi_1'| + |\pi_2| + |\pi_2^B| + |\pi_2^C| + 2 \cdot |\pi_1| + 5$
$\leq (|\pi_1| - 1) + |\pi_2| + 2 \cdot (|\pi_2| - 1) + 2 \cdot |\pi_1| + 5$
$\leq 2 \cdot (|\pi_1| + |\pi_2|) + 2 < (|\pi_1| + |\pi_2| + 1)^2 = |\pi|^2$

(c) $\alpha = \forall x \beta(x)$. In π_1 once more the first ancestor of this formula can be originated by weakening (i.), or by a $\Im$–inference (ii.), or by a right introduction rule (iii.):

$$\dfrac{\Pi \Longrightarrow \Lambda, \beta(a)}{\Pi \Longrightarrow \Lambda, \forall x \beta(x)}$$

where a is an *eigenvariable* different from one inference to another. In case (iii.) take a fresh variable c and replace each of the above inferences with:

$$\dfrac{\Pi \Longrightarrow \Lambda, \beta(c)}{\Pi \Longrightarrow \beta(c), \Lambda, \forall x \beta(x)}$$

obtaining a proof π_1' of $\Gamma \Longrightarrow \beta(c), \Delta, \forall x \beta(x)$.

Get a proof π_1'' of $\Gamma \Longrightarrow \beta(c), \Delta$ as follows:

(a) If the formula $\forall x\beta(x)$ has depth $-\infty$ in the endsequent of π_1', then once more use the first theorem to obtain a proof π_1'' of $\Gamma \Longrightarrow \beta(c), \Delta$.

(b) If the formula $\forall x\beta(x)$ has depth 0 in the endsequent of π_1', then cross π_1' and π_2 making a depth 0 cut to form a proof π_1'' of this sequent.

$$\dfrac{\dfrac{\pi_1'}{\Gamma \Longrightarrow \beta(c), \Delta, \forall x\beta(x)} \quad \dfrac{\dfrac{\dfrac{\pi_2}{\forall x\beta(x), \Gamma \Longrightarrow \Delta}}{structural}}{\forall x\beta(x), \Gamma \Longrightarrow \beta(c), \Delta}}{\Gamma \Longrightarrow \beta(c), \Delta}$$

If the formula $\forall x\beta(x)$ had depth d in the endsequent of π_1, then $\beta(c)$ has depth $< d$ in the endsequent of π_1''.

(a) Now consider in π_2 the inferences that originate the direct ancestors of the cut formula: still can be weakening, or a $\Im$-inference of a left introduction rule:

$$\dfrac{\beta(t), \Pi \Longrightarrow \Lambda}{\forall x\beta(x), \Pi \Longrightarrow \Lambda}$$

Let us first consider the latter (left-introduction rule).

Replace these inferences with:

$$\dfrac{\dfrac{\dfrac{\pi_1''[t]}{\Gamma \Longrightarrow \beta(t), \Delta} \quad \beta(t), \Pi \Longrightarrow \Lambda}{\Pi, \Gamma \Longrightarrow \Delta, \Lambda}}{\forall x\beta(x), \Pi, \Gamma \Longrightarrow \Delta, \Lambda}$$

and from this (adding weak inferences as necessary) obtain a proof π_2' of $\forall x\beta(x), \Gamma \Longrightarrow \Delta$ where $\forall x\beta(x)$ has now depth ≤ 0. Note that the cut has depth $< d$. In case $\forall x\beta(x)$ has depth $-\infty$ we apply once more the first theorem; in case of depth $= 0$ make a cut (of depth 0) crossing π_1 and π_2'.

Check yourself that $h(\pi') \leq 2 \cdot h(\pi)$ and $|\pi'| < |\pi|^2$.

(d) $\alpha = atomic$, hence $d = 0$ or $d = 1$.In this case the point in which the first direct ancestor of the cut formula originates can be a weakening, a $\Im-$rule or an axiom $\alpha \Longrightarrow \alpha$. Build π_1' replacing occurrences of $\alpha \Longrightarrow \alpha$ which contain a first direct ancestor of the cut formula with:

$$\dfrac{\dfrac{\pi_2}{\alpha, \Gamma \Longrightarrow \Delta}}{\alpha, \Gamma \Longrightarrow \Delta, \alpha}$$

Note that α in the succedent has depth $-\infty$. So, at the end (with the help of structural rules) we get a proof π_1' of $\Gamma \Longrightarrow \Delta, \alpha$ with α in the endsequent of depth ≤ 0 and $\pi_1' \preccurlyeq \pi_1$. A proof π_2' of $\alpha, \Gamma \Longrightarrow \Delta$ is obtained specularly in the same way, again with α in the endsequent of depth ≤ 0 and $\pi_2' \preccurlyeq \pi_2$.

(a) If α has depth $-\infty$ in the endsequent either of π_1' or of π_2' then obtain from it a proof π' of $\Gamma \Longrightarrow \Delta$ by applying the first theorem.

(b) If α has depth 0 in both endsequents of π'_1 and π'_2, cross π'_1 and π'_2 and obtain π' making a cut on α. Note that this cut is *anchored*.

Check that $h(\pi') \leq 2 \cdot h(\pi)$. As or the size, $|\pi'| \leq (c-1) \cdot |\pi|^2$ follows from the fact that $(c-1) \cdot |\pi_i| + 1$ is a bound of the number of initial sequents of π_i. QED

Corollary 22. *Suppose π has depth $\leq d$ and $d \geq 0$. Then a proof π' exists with the same endsequent, in which all free cuts have depth $< d$ and $h(\pi') < 2^{h(\pi)+1}$ and $\pi' \preccurlyeq \pi$.*

Proof. Induction on $h(\pi)$. Let us define:

$$f(i) = \text{minimum number } z \text{ such that, if } h(\pi) \leq i, \text{ then } h(\pi') \leq z.$$

Notice that:

(a) $f(0) = 0$, because in this case there is no cut in π.

(b) If $i = h(\pi) = 1$, then:

i. if π does not contain *free cuts*, put $\pi' = \pi$.

ii. otherwise, since the height is 1, we can have only initial sequents, structural rules and *atomic* cuts (why atomic? Note that if the cut formula were introduced by weakening its *depth* would be $-\infty$ and if were a principal formula of a $\Im$-inference, the cut would be not free); hence by the previous theorem $h(\pi') < 2$, hence $f(1) = 1$.

3. If $i \geq 2$, suppose e.g. that π ends with a rule with two upper sequents whose proofs are π_1 and π_2. Apply (IH) to them and obtain π'_1 and π'_2 whose height is $\leq f(i-1)$, whose free cuts have depth $< d$ and such that $\pi'_1 \preccurlyeq \pi_1$ and $\pi'_2 \preccurlyeq \pi_2$. Form a new proof ξ replacing in π the subproofs π_1 and π_2 with π'_1 and π'_2:

i. If ξ does not end with a free cut, put $\pi' = \xi$. The height is $\leq f(i-1) + 1$

ii. otherwise, since $\pi'_j \preccurlyeq \pi_j$ $(j = 1, 2)$ the cut must have depth $\leq d$: if depth $< d$, then put $\pi' = \xi$; if depth $= d$, apply the previous theorem to ξ and obtain π'. The height is $\leq 2 \cdot f(i-1) + 2$.

By induction on i, prove that $f(i) < 2^{i+1}$.

QED

Now, iterating this result $d+1$ times, we obtain a proof of height $< 2^{h(\pi)+1}_{d+1}$, where every cut has depth < 0, namely $-\infty$. Hence apply once more the Theorem 109 to get a proof without free cuts and therefore Theorem 110 follows.

Corollary 23. *Let Φ be a class of formulas closed under substitution of terms and subformulas. Suppose that each $\Im-$inference has only formulas in the class Φ as principal formulas. Then for all proofs π, there is a proof π' of the same endsequent in which all cut formulas are in Φ.*

Proof. Note that logical inferences can be viewed as $\Im - inferences$. E.g. a right introduction inference of $\wedge$ can be seen an instance of a skeleton of this form:

$$\frac{C_1 \Longrightarrow D_1, \alpha \quad C_2 \Longrightarrow D_2, \beta}{C \Longrightarrow D, \alpha \wedge \beta}$$

Hence let us consider $\Im^+ = \Im - inferences$ plus logical inferences with the principal formulas in Φ plus all identities $\alpha \Longrightarrow \alpha$ with $\alpha \in \Phi$ atomic. The last theorem gives a π' with no free cuts with respect to $\Im^+$. If every $\Im^+$-inference has only formulas in Φ as principal formulas, then there is a proof π' of the same endsequent in which all cut formulas are *anchored*. Cuts in π' have depth 0, hence the cut formulas are occurrences of a principal formula in $\Im^+$, hence are in Φ. QED

An extremely powerful consequence of this corollary is the following:

(a) Suppose our acceptable inferences are the initial sequents corresponding to the non logical axioms of PA, the equality axioms and the induction rule restricted to Σ_k formulas (the fragment denoted $\mathsf{I\Sigma_k}$).

(b) Take $\Phi = \Sigma_k \cup \Pi_k$ and let π be a proof *free-cut-free* in $\mathsf{I\Sigma_k}$ of a sequent $\Gamma \Longrightarrow \Delta$ where Γ, Δ are made of formulas in Φ.

(c) Suppose by contradiction that there is a formula α occurring in π such that $\alpha \notin \Phi$. Hence, by the above corollary, α is not a cut formula, since all cut formulas in π are in Φ (notice that the principal formulas of the induction rules and of axioms are in $\Sigma_k \subseteq \Phi$).

(d) Therefore in π' the formula α has not been deleted and will occur in the final sequent as a side formula, or as a subformula (i.e. it occurs as an auxilary formula at some step). Contradiction: against the assumption that all formulas in $\Gamma \Longrightarrow \Delta$ were in Φ.

(e) Hence we conclude that in π occurs only formulas from Φ.

7.3. Bounded Arithmetic and Polynomial Time Computability

The philosophical motivation for the introduction of *Bounded Arithmetic* theories can perhaps be traced back to dissatisfaction with traditional finitism and intuitionist constructivism in constructive mathematics, not considered by some as a genuine alternative to realism: "finitism is the last refuge of platonism"(Nelson (1986), p. 10). In Parikh (1971), a pioneering work that sought to construct a system reflecting an "anthropomorphic point of view" in mathematics, it is proposed that numbers that are too large such as $10^{10^{10}}$ should be considered infinite and formal theories of arithmetic are proposed where exponentiation is not assumed to be defined over all numbers. This whole discussion is connected to the theme of feasibility and of the computational infeasibility of exponentiation that we have discussed in relation to the Church-Turing thesis.

There are two principal approaches to bounded arithmetic. The original approach involved theories such as $\mathsf{I\Delta_0}$ and $\mathsf{I\Delta_0} + \Omega_1$ (e.g. Cook and Nguyen (2010) is a handbook based on this approach). The first of these theories was introduced in Parikh (1971), where every Δ_0-formula defines what is called "a concrete predicate"; later, in Buss (1986), bounded theories such as S^i_2 and T^i_2, based on a broader language, have been introduced and extensively studied. One of the main features is their close connection to low-level computational complexity. The T^i_2 will be defined by restricting induction to Σ^b_i -formulas, where by induction we mean the usual one. For S^i_2, we need a different kind of induction schema. We will deal here mainly with the theories $\mathsf{S}^1_2, \mathsf{S}^2_2, \mathsf{S}^3_2$.... The idea behind this approach is to modify $\mathsf{I\Delta_0} + \Omega_1$ so that the definable functions in these theories are more directly related to the levels of the so-called *Polynomial Time Hierarchy*, instead of the *Linear Time Hierarchy*. Actually the union of

these theories $\bigcup_i \mathsf{S}^i_2$ is equivalent to the theory $\mathsf{I}\Delta_0 + \Omega_1$. All these theories are *predicative* in the sense of Nelson (1986), that is, interpretable in Robinson's Q. In particular, through an application of (partial) cut-elimination we will see that the Σ^b_1 definable functions of this extended language in the theory S^1_2 are the *polynomial time computable functions*.

Recall that $\Sigma^L_0 = \mathsf{LINTIME}$ are the languages decided in $c \cdot n$ time (for some c) and $\Sigma^L_{i+1} = \mathsf{NLINTIME}(\Sigma^{\mathsf{L}}_i)$ are the languages accepted in time $c \cdot n$ by a nondeterministic machine with oracle in Σ^L_i and finally $\mathsf{LTH} = \bigcup_i \Sigma^L_i$ is the so-called *Linear Time Hierarchy*. Note that LTH is a class of relations. We define a class of functions in terms of *function graph* i.e. the set of pairs $\langle x, y\rangle$ such that $f(x) = y$: a function f in computable in linear time iff this set belong to LTH. It is well known that $\mathsf{LTH} = \Delta^{\mathbb{N}}_0$, namely the relations $\Delta_0 - definable$ in the standard model.

Recall also that we say that a function f is Σ_1-*definable* in $\mathsf{I}\Delta_0$ iff there is a Σ_1-formula ϕ such that $\phi(\overline{n}, \overline{f(n)})$ is true for all n and the theory proves $\forall x \exists! y \phi(x, y)$. In this case it holds that this is true of f iff its graph is in LTH and there is a bounding term t of the language of the theory for existential quantifier, i.e. the theory actually proves $\forall x \exists! y \leq t\phi(x, y)$.

For many purposes, it is useful to extend $\mathsf{I}\Delta_0$ with the axiom:

$$\Omega_1 = \forall x \exists y (x^{|x|} = y)$$

where $|x|$=smallest integer bigger or equal to $log_2(x + 1)$=length of x in base two.

The theory $\mathsf{I}\Delta_0 + \Omega_1$ allow more flexible constructions, since the axiom Ω_1 just captures the polynomial increase of the lengths in such a way that definable functions of this theory satisfy the condition that $|f(x)| \leq p(|x|)$, for some polynomial p, instead of $|f(x)| \leq c \cdot |x|$ as in $\mathsf{I}\Delta_0$.

The function $x^{|x|}$ is superpolynomial and has a *polynomial growth rate*, i.e. if t is a term builded with functions $S, \cdot, +, x^{|x|}$, then a polynomial p_t exists such that $|t(x)| \leq p_t(|x|)$. Parikh's result that we are going to prove extend to $\mathsf{I}\Delta_0 + \Omega_1$ as well as to Buss's theories, and from this it follows that they does not prove the totality of exponential[2].

Parikh raised the question of whether exponentiation is necessary to carry out Gödel arithmetisation. The argument was as follows (here in Buss (1999) reconstruction). For instance, one wants the theory to be able to define the notion of substituting a term into a formula and prove the result is a formula. The following argument by Parikh, together with the difficulty in proving Löb's third derivability condition in this theory, led to believe that an 'intensional' type of arithmetization was not possible in $\mathsf{I}\Delta_0$. Actually, if the number of symbols of $\theta(x)$ is m and that of the term t is n, then the number of symbols of $\theta[t/x]$ is about $m \cdot n$. By using "efficient codings" we have that $\ulcorner \theta(x) \urcorner = 2^{O(m)}$ and $\ulcorner t \urcorner = 2^{O(n)}$, and therefore $\ulcorner \theta[t/x] \urcorner = 2^{O(m \cdot n)}$. But $\ulcorner \theta[t/x] \urcorner = 2^{O(m \cdot n)} \leq 2^{|\ulcorner \theta \urcorner| \cdot c \cdot n} \leq \ulcorner \theta(x) \urcorner^{O(n)}$ and since the number of symbols of a word whose code is x is bounded by $|x|$, lastly we have $\ulcorner \theta(x) \urcorner^{O(|\ulcorner t \urcorner|)}$. The conclusion is that the value of $\ulcorner \theta[t/x] \urcorner$ cannot be bounded by a polynomial of $\ulcorner t \urcorner$ and $\ulcorner \theta(x) \urcorner$. However, not all the power of the exponential function is required here: actually we need just the function $(x, y) \mapsto x^{|y|}$. This explain the axiom Ω_1.

Another motivation for the introduction of these extensions of $\mathsf{I}\Delta_0$ is related to the *intensional approach to arithmetization*. We have seen that every recursive function is numeralwise representable even in very weak theories such as R and Q: they can 'represent' all particular instances $f(n)$ of a recursive function f, but not prove general properties of the function. This in is in contrast to the approach to the arithmetization of syntax that Feferman (1960) called *intensional*. Here, when we define concepts such as "formula", "term", "substitution", "proof", "theorem", etc, we demand that the theory can prove general properties of these concepts,

[2] To get an idea of how much concrete number theory or combinatorics can be done in these theories, see for example Beame, Impagliazzo, Pitassi (1993) or Ajtai (1994) or D'Aquino (1992), Berarducci and D'Aquino (1995) and D'Aquino and Macintyre (2000).

which is not required in the case of arithmetisation based on the concept of representability. The intesional arithmetization can be carried out in S^1_2 as well as in $\mathsf{I}\Delta_0 + \Omega_1$ (see Buss (1986)). The above remarks about substitution suggest that an intensional arithmetization could hardly be done in $\mathsf{I}\Delta_0$, without resorting to the *shortening* technique discussed below (i.e. an intensional arithmetization of metamathematics can be given rather artificially already in Q by replacing "x is a proof"with "x is a proof in the initial segment J", as we explained when discussing the problem of Gödel's second theorem for Q).

Working in such weak theories generally entails the need to economise in the use of resources, to such an extent that Parikh (1971) stated as an open question the issue of whether the exponentiation function is required for the arithmetization of metamathematics in Gödel's incompleteness theorems. The way in which we code the syntax become relevant. We will refer to Wilkie and Paris (1987) for a careful Gödel coding such that, if n is the number of symbols of a formula θ:

(a) $n \leq |\ulcorner\theta\urcorner| \leq c \cdot n$, where $|x|$ is the base-two length of x, for some constant c. Hence also:

(b) $2^n \leq \ulcorner\theta\urcorner \leq 2^{d \cdot n}$, for some constant d.

We call *efficient* such a coding. Let us take two examples of how to economise:

(a) Speaking of Gödel's incompleteness results, coding sequences is an essential step. For coding sequences we used exponentiation, but this function is not total in $\mathsf{I}\Delta_0$. The coding of sequences e.g. in Hájek and Pudlák (1993) allows us to manipulate sequences provably in $\mathsf{I}\Delta_0$ and thereby define with a Δ_0 formula the graph of the function $x^y = z$, proving in $\mathsf{I}\Delta_0$ its main properties (except totality!).

To code a sequence of numbers in a more efficient way, they use a pair of numbers:

i. The first number will be the number determined by the concatenation of binary expansions of the numbers to be coded.

ii. The second one will be a binary code of the markers which determine beginnings and ends of the coded numbers.

In fact we code sequences of arbitrary $0-1$ words in such a way. Suppose we want to code a sequence of $0-1$ words:

$$0011, 101, 010$$

Then we take two numbers whose binary expansion is the following:

$$11101010, \quad 10001001001$$

The first one is the concatenation of the words above (where we have to omit the first two 0's) and the second one is a sequence of markers which determines the partition. If this pair is considered to be a code of *a sequence of numbers* then it will code $(3, 5, 2)$. This coding of finite sequences is used to define in $\mathsf{I}\Delta_0$ the exponential relation $x^y = z$ (however, it is not possible in this theory to show that such a z always exists!)[3].

(b) The second example concerns the efficient representation of numerals. We need to use in arithmetization the function $n \mapsto \ulcorner\overline{n}\urcorner$. Using efficient numerals and an efficient arithmetization, this transformation is p-time. Actually, to represent the number n by the numeral $S(S(...S(\overline{0})...))$ is problematic, when we don't have the exponentiation as a total function. The classical numerals $S^n(0)$ cannot be used, since their length (number

[3] The first Δ_0 definition of the graph of exponentiation, formalized in the theory $\mathsf{I}\Delta_0$ is due Gaifman and Dimitracopoulos (1982)

of symbols) is greater than n. Hence the Gödel number of this numeral will be of order $2^{c \cdot n}$, for a constant c. Hence the function $n \mapsto \ulcorner \overline{n} \urcorner$, sending a number to the code of its numeral will be exponential. At the opposite the length of the *efficient numerals* is bounded by a polynomial of $|n|$. Indeed, we think of the base two representation of n is $a_k a_{k-1} ... a_0$, where each a_i is 0 or 1, and we represent n as:

$$a_0 + 2(a_1 + 2(a_2 + ...(a_{k-1} + 2a_k)...))$$

Hence the Gödel number of the numeral of n is now of order $2^{c \cdot \log_2(n)}$.

Parikh proved that, although there exist formulas $\eta(x, y, z)$ having, provably in $\mathsf{I}\Delta_0$, the basic properties of exponentiation $x^y = z$, none of these formulas is such that $\mathsf{I}\Delta_0$ proves $\forall x \forall y \exists z\, \eta(x, y, z)$. The result follows from this theorem.

Theorem 111. *If $\theta(x, y)$ is* bounded *and $I\Delta_0 \vdash \forall x \exists y \theta(x, y)$, then exists a term t such that:*

$$I\Delta_0 \vdash \forall x \exists y < t\theta(x, y)$$

Proof. Recall that terms t in the language of PA are polynomials of the form $x^k + c$, which grow slower than the exponential function, and Parikh's theorem says that we can bound the value of Δ_0 definable functions in $\mathsf{I}\Delta_0$ only by a term in this language and therefore these functions *can increase the length of the input only linearly.* As the exponential relation $x^y = z$ actually has a Δ_0 definition, it follows that $\mathsf{I}\Delta_0$ cannot prove the totality of exponentiation.

An extremely laborious proof was given in Buss (1986) in sequent calculus; although more akin to the spirit of these readings, we prefer to report the semantic one contained in Hájek and Pudlák (1993) and Krajíček (1995), an elegant model-theoretic argument, using compactness. Suppose by contradiction that the premiss is true, but this conclusion is false. Therefore (for a new constant c) the following is consistent, for every t:

$$\mathsf{I}\Delta_0 + \forall y < t(c) \neg\theta(c, y)$$

Note that it is unprovable also any disjunction:

$$\bigvee_{i \leq k} \exists y < t_i(c)\theta(c, y)$$

(otherwise one could take $t = t_0 + t_1 + ... + t_k$ and $\exists y < t(c)\theta(c, y)$ would be provable, since each disjunct would imply this formula). Hence any finite subset of sets:

$$\Gamma_k = \mathsf{I}\Delta_0 + \forall y < t_0(c) \neg\theta(c, y) + ... + \forall y < t_k(c) \neg\theta(c, y)$$

is consistent. It follows by compactness that the set:

$$\Gamma = \mathsf{I}\Delta_0 + \{\forall y < t_i(c) \neg\theta(c, y) | t_i \textit{ is a term}\}$$

is consistent too (note that any finite subset of Γ is a subset of Γ_n, for some n). So, Γ has a model $\mathcal{M}$. Now consider an initial segment:

$$I = \{b \in \mathcal{M} | \mathcal{M} \models b < t(c),\ \textit{for some } t(x)\}$$

of this model, i.e. a subset closed downwards and by addition and product. It is well known that I remains a model of $\mathsf{I}\Delta_0$: actually $\mathsf{I}\Delta_0$ is a *bounded theory of arithmetic*, namely is axiomatized by bounded formulas, e.g. the induction principle can be formulated as follows:

$$\phi(\overline{0}) \wedge \forall x < z(\phi(x) \rightarrow \phi(x + 1)) \rightarrow \phi(z)$$

If $\psi(x)$ is a bounded formula and I an initial segment of a model $\mathcal{M}$ closed under addition and multiplication, an *absoluteness principle* holds, namely if $a \in I$, then $I \models \psi(a)$ iff $\mathcal{M} \models \psi(a)$. But:

$$I \models \exists x \forall y \neg \theta(x, y)$$

(contradiction).

Indeed, we have proved that $\mathsf{I}\Delta_0$ cannot prove the totality of such functions that eventually majorize all polynomials. QED

Corollary 24. *x^y is not provably total in $\mathsf{I}\Delta_0$.*

Remark 5. *Although Parikh's Theorem was originally established for this theory, it can be easily extended to Buss' theories $\mathsf{S}^{\mathsf{i}}_2$ (see Verbrugge (1993)).*

Here some key results. Points 2., 3. and 4. somehow support the argument of Nelson (1986) where Q is considered the reference theory from which to start, admitting only extensions that can be interpreted in it.

Theorem 112. *The following hold:*

(a) Q *is not intepretable in* R.

(b) S^1_2 *is intepretable in* Q.

(c) $\mathsf{I}\Delta_0 + \Omega_1$ *is intepretable in* Q.

(d) $I\Delta_0 + exp$ *is not intepretable in* Q .

(e) $I\Delta_0 + \neg exp$ *is intepretable in* Q.

(f) $\mathsf{I}\Delta_0 + \Omega_1$ *is intepretable in* $\mathsf{I}\Delta_0$.

(g) $I\Sigma_1$ *is not intepretable in* $I\Delta_0 + exp$.

where $exp = \forall x \exists y(2^x = y)$. We remark that the graphs of the functions 2^x and $x^{log(x)}$ are definable in $\mathsf{I}\Delta_0$ (which does *not* prove their totality) but we emphasize the difference between points 3. and 4.: although it has some induction, the theory $\mathsf{I}\Delta_0 + \Omega_1$ (as well as $\mathsf{I}\Delta_0$) is not too far from Q. The presence of exponentiation 2^x as a total function represents on the contrary, a sort of *impassable barrier* and determines a big jump in complexity. An evidence of the jump in complexity highlighted in Nelson (1986) is that by applying a technique called *shortening*, displayed in Solovay (1976) and involved in the results of interpretability or non-interpretability listed above, we can successively close an initial segment of a model of Q under each of the functions $2x, x^2, x^{|x|}, x^{|x|^{||x||}}$... However Paris and Dimitracopulos (1983) showed that it is *not always possible* to close also under exponentiation: there exist a model and an initial segment of it that cannot be restricted in such a away to be closed also under exponentiation. For this reason, we view $x^{log(x)}$ as being more akin to feasible polynomial growth rate functions than to the infeasible exponential function.

These theorems are long and complex, so here we report only one of the simplest cases, where however we see the aforementioned technique due to Solovay at work (see Ferreira G. and Ferreira F. (2013) and Hájek and Pudlák (1993), ch.V, section C for a complete account). Let us therefore consider a syntactical counterparts of initial segments we have seen in models of arithmetic.

Definition 56. *A formula $I(x)$ is* inductive *for a theory T, if this theory proves $I(\overline{0}) \wedge \forall x(I(x) \rightarrow I(Sx))$. It defines an* initial segment *of T, if moreover satisfies $I(x) \wedge y \leq x \rightarrow I(y)$. A formula $J(x)$ is a sub-initial segment of $I(x)$ in T, if the theory proves $J(x) \rightarrow I(x)$*[4].

[4] We point out that in much literature what we have called initial segment is called cut. Below, we have used another term to avoid misunderstandings.

Definition 57. *A theory* T *is locally intepretable in a theory* S *iff each finite part of* T *is intepretable in* S.

Let us add associativity and commutativity of $+, \cdot$ to Q, plus the following distributivity $x \cdot (y + z) = x \cdot y + x \cdot z$ and let us denote Q^+ the resulting arithmetical system. It is easily proved that if $\mathsf{T} \supset \mathsf{Q}^+$ and $I(x)$ is inductive in T, the there exists a subcut $J(x)$ of it. The central result is the following.

Theorem 113. (Solovay's shortening) *If* $\mathsf{T} \supset \mathsf{Q}^+$ *and* $I(x)$ *is inductive in* T*, then there exists a subcut* $J(x)$ *of it, closed under* $+, \cdot$.

Proof. We usw the tecnique of shortening, due to Solovay. Let:

(a) $J_0(x) = \forall y(I(y) \to I(y + x))$

(b) $J(x) = \forall y(J_0(y) \to J_0(y \cdot x))$

By using associativity of $+$ prove in Q^+ that $J_0(x)$ is closed under addition, i.e. if $J_0(x)$ and $J_0(y)$, then $J_0(y + x)$ and from this follows the closure under $+$ of $J(x)$ as well. With similar argument prove that $J(x)$ is closed under multiplication too.

We now show that $J(x)$ is an initial segment. We claim that $J(x) \wedge y \leq x \to J(y)$. We have to show that if $J_0(z)$, then $J_0(z \cdot y)$, that by definition is $J(y)$. Still by definition $J_0(z \cdot y)$ is $I(v) \to I(v + z \cdot y)$, for any v. Since $J(x)$ and $J_0(z)$, by definition of J we have $J_0(z \cdot x)$ and since $y \leq x$, there must be some w such that $y + w = x$ and there fore $J_0(z \cdot (y + w))$. Now we ude the distributivity axiom to obtain $J_0(z \cdot y + z \cdot w)$. But we had $I(v)$ and therefore, by definition of J_0 we have $I(v + (z \cdot y + z \cdot w))$. Since I is a cut and:

$$v + z \cdot y \leq (v + z \cdot y) + z \cdot w = v + (z \cdot y + z \cdot w)$$

we have $I(v + z \cdot y)$ and therefore $J_0(z \cdot y)$, as claimed.

QED

Theorem 114. *The theory* $\mathsf{I}\Delta_0$ *is locally intepretable in* Q^+.

Proof. Fix a finite number of formulas with only bounded quantifiers

$$\phi_0(x, p), ..., \phi_n(x, p)$$

and for each i, let:

$$I_i(x, p) = \phi_i(\overline{0}, p) \wedge \forall y \leq x(\phi_i(y, p) \to \phi_i(Sy, p)) \to \phi_i(x, p)$$

Then define $I(x) = \forall p(\bigwedge_{i \leq n} I_i(x, p))$. This is an inductive formula and by the previous theorem it can be shortened to a cut J closed under addition and multiplication and therefore to a "model" of Q^+. Moreover, since $J(x) \to I(x)$ and $I(x) \to I_i(x, p)$, the inductiion for every ϕ_i holds in J. QED

7.4. Cut-elimination and Polynomial time definable functions

We introduce now Buss' approach to bounded arithmetic and we show some importan application to the theory of computational complexity:

(a) The language af Buss's theory of Bounded Arithmetic is the following:

$$S, 0, +, \cdot, |x|, \lfloor \frac{1}{2}x \rfloor, \#, \leq$$

where $|x| = \lceil \log_2(x+1) \rceil$ =small $y \geq \log_2(x+1)$ is the length of the binary representation of x, $\lfloor \frac{1}{2}x \rfloor$ is the greatest integer less or equal to $\frac{1}{2}x$ and $x\#y$ means $2^{|x|\cdot|y|}$. We will use $\#, \cdot, \lfloor \frac{1}{2}x \rfloor$ to write terms of the form $2^{p(|x|)}$ where $p(x)$ is a polynomial.

(b) We actually "think" in base two: the operation $\lfloor \frac{1}{2}x \rfloor$ erases the last bit from the base two representation of x. Numerals ("dyadic numerals") are defined as follows:

$$\overline{2k+1} = \overline{2k} + S(\overline{0}), \quad \overline{2(k+1)} = S(S(\overline{0})) \cdot \overline{k+1}$$

Note that the length of $\overline{k}$ is of the order of $\log(k)$.

(a) We want to discuss here of a sequent calculus LKB for theories of bounded arithmetic. For these, we must first add the equality initial sequents and some rules for bounded quantifiers.

$$\frac{\phi(t), \Gamma \Longrightarrow \Delta}{t \leq s, \forall x \leq s\phi(x), \Gamma \Longrightarrow \Delta} \qquad \frac{b \leq s, \Gamma \Longrightarrow \Delta, \phi(b)}{\Gamma \Longrightarrow \Delta, \forall x \leq s\phi(x)}$$

$$\frac{b \leq s, \phi(b), \Gamma \Longrightarrow \Delta}{\exists x \leq s\phi(x), \Gamma \Longrightarrow \Delta} \qquad \frac{\Gamma \Longrightarrow \Delta, \phi(t)}{t \leq s, \Gamma \Longrightarrow \Delta, \exists x \leq s\phi(x)}$$

(b) If α is a BASIC axiom, then we add the initial sequent $\Longrightarrow \alpha$.

BASIC will be the set of basic axioms for these operators. A richer language corresponds to a higher number of axioms:

(a) $x \leq b \to a \leq S(b)$

(b) $a \neq S(a)$

(c) $0 \leq a$

(d) $a \leq b \wedge a \neq b \leftrightarrow S(a) \leq b$

(e) $a \neq 0 \to 2 \cdot a \neq 0$

(f) $a \leq b \vee b \leq a$

(g) $a \leq b \wedge b \leq a \to a = b$

(h) $a \leq b \wedge b \leq c \to a \leq c$

(i) $|0| = 0$

(j) $|S(0)| = S(0)$

(k) $a \neq 0 \to |2 \cdot a| = S(|a|) \wedge |S(2 \cdot a)| = S(|a|)$

(l) $a \leq b \to |a| \leq |b|$

(m) $|a\#b| = S(|a| \cdot |b|)$

(n) $|S(0) = 0\#a$

(o) $a \neq 0 \to 1\#(2{\cdot}a) = 2{\cdot}(1\#a) \wedge 1\#(S(2{\cdot}a)) = 2 \cdot (1\#a)$

(p) $a\#b = b\#a$

(q) $|a| = |b| \to a\#c = b\#c$

(r) $|a| = |b| + |c| \to a\#d = (b\#d) \cdot (c\#d)$

(s) $a \leq a + b$

(t) $a \leq b \wedge a \neq b \to S(2{\cdot}a) \leq 2{\cdot}b \wedge S(2{\cdot}a) \neq 2{\cdot}b$

(u) $a + b = b + a$

(v) $a + 0 = a$

(w) $a + S(b) = S(a + b)$

(x) $a + b \leq a + c \leftrightarrow b \leq c$

(y) $a \cdot b = b \cdot a$

(z) $a \cdot (b + c) = (a \cdot b) + (a \cdot c)$

() $S(0) \leq a \to (a \cdot b \leq a \cdot c \leftrightarrow b \leq c)$

() $a \neq 0 \to |a| = S(|\lfloor \frac{1}{2}a \rfloor|)$

() $a = \lfloor \frac{1}{2}b \rfloor \leftrightarrow 2 \cdot a = b \vee S(2 \cdot a) = b$

We define now a hierarchy of bounded formulas of this language, analogous to the Arithmetical Hierarchy, but now we count the alternations of bounded quantifiers:

(a) The set $\Delta_0^b = \Sigma_0^b = \Pi_0^b$ is equal to the set of formulas in which all quantifiers are sharply bounded, i.e. of the form $\forall x \leq |t|$ or $\exists x \leq |t|$.

(b) For $i > 0$, the sets Σ_i^b and Π_i^b are inductively defined by the following conditions:

i. If α and β are Σ_i^b-formulas, then so are $\alpha \vee \beta$ and $\alpha \wedge \beta$.

ii. If α is a Π_i^b formula and β is a Σ_i^b-formula, then $\alpha \to \beta$ and $\neg\alpha$ are Σ_i^b-formulas.

iii. If α is a Π_i^b-formula, then α is a Σ_i^b-formula.

iv. If α is a Σ_i^b-formula and t is a term, then $(\forall x \leq |t|)\alpha$ is a Σ_i^b-formula.

v. If α is a Σ_i^b-formula and t is a term, then $(\forall x \leq t)\alpha$ is a Σ_i^b-formula.

vi. The four inductive conditions defining Π_i^b are dual.

The terms of this language define *functions of polynomial growth rate*. The theories T_2^i will be defined by adding to the BASIC axioms the standard induction INDrestricted to Σ_i^b-formulas. The theories S_2^i will be defined instead by adding the induction schema PIND:

$$\alpha(0) \wedge \forall x(\alpha(\lfloor \tfrac{1}{2}x \rfloor) \to \alpha(x)) \to \forall x \alpha(x)$$

restricted to Σ_i^b-formulas. We have that $\mathsf{S}_2 = \cup_i \mathsf{S}_2^i = \cup_i \mathsf{T}_2^i = \mathsf{T}_2$. Moreover

$$\mathsf{S}_2^1 \subseteq \mathsf{T}_2^1 \subseteq \mathsf{S}_2^2 \subseteq \mathsf{T}_2^2 \subseteq \mathsf{S}_2^3 \subseteq ...$$

but it is open whether they are distinct. Anyway T_2 (and therefore S_2) is equivalent to $\mathsf{I}\Delta_0 + \Omega_1$. The polynomial time computable functions can be inductively defined in a similar way to what we did for primitive recursive functions, that is, by means of axioms as follows:

(a) *Initial functions.*

i. The nullary constant function 0.
ii. The successor function $S(x)$.
iii. The doubling function $2x$.
iv. The conditional function $Cond(x, y, z)$=if $x = 0$ then y, else z.
v. The *projection* functions are polynomial time functions

(b) The *composition* of polynomial time functions is a polynomial time function.

(c) The function f defined by *limited iteration on notation* from g and h, polynomial time:

i. $f(0, x) = g(x)$
ii. $f(z, x) = h(z, x, f(\lfloor \frac{1}{2}z \rfloor, x))$ for $z > 0$ provided $|f(z, x)| \leq p(|z|, |x|)$ where $p(x)$ is a polynomial.

Following Buss we define in a rather unusual way the *Polynomial Time Hierarchy. Predicates* are here functions $0, 1$ (their characteristic functions). A predicate is polynomial time computable provided its characteristic function is polynomial time. We distinguish between *logarithmic* bounds $p(|x|)$ and *polynomial* bounds $2^{p(|x|)}$ to the quantifiers. Logarithmically bounded quantification corresponds to *sharply bounded* quantification. We denote P_0^p the smallest class containing the initial functions, closed under composition and logarithmically bounded quantifiers.

(a) $\Delta_0^p = \Sigma_0^p = \Pi_0^p$ are the predicates of P_0^p.

(b) Σ_i^p is the class of polynomially bounded predicates $R(x)$-definable by $R(x) = (\exists y \leq 2^{s(|x|)})(Q(x,y))$ for some polynomial $s(n)$ and Δ_i^p predicate Q

(c) Π_i^p is the dual class of their complements.

(d) $P_{i+1}^p$is the class of functions computable on a deterministic Turing machine in polynomial time with oracle in Σ_i^p.

(e) Δ_{i+1}^p is the class of predicates with characteristic function in $\Box_{i+1}^p$.

Hence $\mathsf{P} = \Delta_1^p$, $\mathsf{NP} = \Sigma_1^p$, $\mathsf{FP} = \Box_1^p$ and $co - \mathsf{NP} = \Pi_1^p$. The class of polynomial time functions is $\Box_1^p$, and the class of polynomial time predicates is Δ_1^p.

Definition 58. *A function f is Σ_i^b-* definable in a theory T*, if there is a formula $\phi \in \Sigma_i^b$ such that:*

(a) ϕ defines the graph of f,

(b) $\mathsf{T} \vdash \forall x \exists y \phi(x,y)$

(c) $\mathsf{T} \vdash \forall x \forall y \forall z(\phi(x,y) \wedge \phi(x,z) \rightarrow y = z)$

Definition 59. *A predicate $P(x)$ is Δ_i^b definable in T provided there are $\phi \in \Sigma_i^b$ and $\psi \in \Pi_i^b$ provably equivalent in T, that define $P(x)$.*

Definition 60. *A theory is said to be* bounded *if it is axiomatizable with a set of bounded formulas.*

In the previous discussion we used Parikh's theorem for all the theories of the hierarchy S_2^i. The syntactic proof in the sequent calculus contained in Buss (1986) is rather complex. However it can be achieved semantically in a simpler way along the same lines as in Verbrugge (1993).

Theorem 115. *If S_2^i proves $\forall x \exists y \theta(x,y)$, and θ is bounded, then $\forall x \exists y \leq t(x) \theta(x,y)$, for some term t.*

Proof. Suppose that this is not true. Then take:

$$\mathsf{S}_2^i \cup \{\forall y \leq \overbrace{c\#...\#c}^{k-times} \neg\theta(c,y) | k \in \omega\}$$

where c is a new constant and observe that this set is finitely satisfable. Hence by compactenss the whole set is satisfable, i.e. has a model, say $\mathcal{M}$ where a certain element a interprets c. Take the submodel $\mathcal{U}$ defined as:

$$\mathcal{U} = \{b \in \mathcal{M} | \exists k (b \leq \overbrace{c\#...\#c}^{k-times})\}$$

This model is closed under the operations $+, \cdot, S, \#, \lfloor \frac{1}{2} x \rfloor$ Note that the theory S_2^i can be axiomatized by Π_1 sentences. Actually we can write also the induction axiom as:

$$\forall y (\theta(\overline{0}) \wedge \forall x \leq y (\theta(\lfloor \tfrac{1}{2} x \rfloor) \rightarrow \theta(x)) \rightarrow \forall x \leq y \theta(x))$$

where $\theta \in \Sigma_i^b$. But Π_1 sentences are preserved "downward" hence this is a model of S_2^i too. On the other hand $\mathcal{U}$ does not verify $\exists y \theta(a,y)$, otherwise there would be a k and a $b \in \mathcal{M}$ such that:

(a) $b \leq \overbrace{c\#...\#c}^{k-times}$ and

(b) $\theta(a,b)$

against the hypothesis, and therefore $\mathsf{S}^{\mathsf{i}}_2 \nvdash \forall x \exists y \theta(x,y)$ (contradiction, against the assumption). QED

Some important facts.

(a) Every polynomial time function is Σ^b_1-definable in $\mathsf{S}^{\mathsf{i}}_2$.

(b) Every polynomial time predicate is Δ^b_1-definable in $\mathsf{S}^{\mathsf{i}}_2$.

(c) Every $\Box^p_i$ function is Σ^b_i-definable in $\mathsf{T}^{\mathsf{i}-1}_2$ and in $\mathsf{S}^{\mathsf{i}}_2$.

(d) Every Δ^p_i predicate is Δ^b_i-definable in $\mathsf{S}^{\mathsf{i}}_2$.

(e) A predicate is Σ^p_i if and only if there is a Σ^b_i-formula which defines it.

We are going to show the main result of Buss (1986), i.e. the inverse implications of these points, in particular:

(a) Every Σ^b_1-definable function in $\mathsf{S}^{\mathsf{i}}_2$ is polynomial time computable.

(b) Δ^b_1-definable predicate in $\mathsf{S}^{\mathsf{i}}_2$ is polynomial time computable.

Generalizations to other levels holds.

(a) To do this, first we must formulate the PIND induction as a rule:

$$\frac{\alpha(\lfloor \frac{1}{2} b \rfloor), \Gamma \Longrightarrow \Delta, \alpha(b)}{\alpha(\overline{0}), \Gamma \Longrightarrow \Delta, \alpha(t)}$$

where b occurs only as indicated.

(b) Axioms α are formalized as initial sequents $\Longrightarrow \alpha$

(c) In view of what we are about to say it must be emphasized that many functions needed for arithmetization of syntax are $\Sigma^b_1 - definable$ in S^1_2 (e.g. the coding of finite sequences, projections, concatenation of finite sequences) and many predicates are $\Delta^b_1 - definable$ in S^1_2 (e.g. $Seq(x)$, $Len(x)$).

Theorem 116. (Buss 1985) *Let $i \geq 1$ and let us suppose $\mathsf{S}^{\mathsf{i}}_2$ proves $\forall x \exists y \theta(x,y)$ where $\theta \in \Sigma^b_i$. Then there exists a term t, a formula ψ and a function $g \in \Box^p_i$ such that $\mathsf{S}^{\mathsf{i}}_2$ proves:*

(a) $\forall x \forall y (\psi(x,y) \rightarrow \theta(x,y))$

(b) $\forall x \forall y \forall z (\psi(x,y) \wedge \psi(x,z) \rightarrow y = z)$

(c) $\forall x \exists y \leq t \psi(x,y)$

(d) for all n it is true in the standard model $\psi(\overline{n}, \overline{g(n)})$

Corollary 25. *If g is Σ^b_i-definable in $\mathsf{S}^{\mathsf{i}}_2$, then $g \in \Box^p_i$.*

Buss's witnessing method. Although this theorem applies in its most general form for $i \geq 1$, in this exposition we can focus on the case where $i = 1$, in order to avoid introducing further details. Let therefore $\theta(c) \in \Sigma^b_i$-formula in prenex form. Then $Witness^i_\theta(w, c)$ (which is provably Δ^b_i in $\mathsf{S}^{\mathsf{i}}_2$) is defined inductively as follows:

(a) If $\theta \in \Pi^b_{i-1} \cup \Sigma^b_{i-1}$ then $Witness^i_\theta(w, c) \leftrightarrow \theta(c)$

(b) $Witness^i_\theta(w, c)$ distributes over $\vee$ and $\wedge$, e.g. w witnesses $\alpha \wedge \beta$ iff $(w)_1$ witnesses α *and* $(w)_2$ witnesses β.

(c) If $\theta \notin \Pi^b_{i-1} \cup \Sigma^b_{i-1}$ and has the form $\forall x \leq |s(c)|\psi(c, x)$, then $Witness^i_\theta(w, c)$ is the conjunction of:

i. $Seq(w) \wedge Len(w) = |s| + 1$
ii. $\forall x \leq |s(c)|Witness^i_{\psi(c,b)}((w)_{x+1}, c, x)$

(Hence $w = \langle w_0, ..., w_{|s|}\rangle$ witnesses the truth of θ iff each w_i witnesses $\psi(c, i)$).

(d) If $\theta \notin \Pi^b_{i-1} \cup \Sigma^b_{i-1}$ and has the form $\exists x \leq t(c)\psi(x, c)$, then $Witness^i_\theta(w, c)$ is the conjunction of:

i. $Seq(w) \wedge Len(w) = 2 \wedge (w)_1 \leq t$
ii. $Witness^i_{\psi(c,b)}((w)_2, c, (w)_1)$

(Hence $w = \langle n, v\rangle$, $n \leq t$ and v witnesses $\psi(c, n)$).

(e) In case of negated formulas not in $\Pi^b_{i-1} \cup \Sigma^b_{i-1}$ we internalize the negation in order to bring us back to the cases listed above.

Some properties. For all $\theta \in \Sigma^b_i$, the theory $\mathsf{S}^{\mathsf{i}}_2$ proves:

$$\exists w(Witness^i_\theta(w, c)) \leftrightarrow \theta(c)$$

In the following We use this notation. If $\Gamma = \langle A, B, C\rangle$ then $\bigwedge\Gamma = (A \wedge (B \wedge C))$ and $\bigvee\Gamma = (A \vee (B \vee C))$.

Theorem 117. (The main theorem) *Let us suppose* $\mathsf{S}^{\mathsf{i}}_2$ *proves* $\Gamma, \Pi \Longrightarrow \Lambda, \Delta$ *where* Γ, Δ *are composed by* Σ^b_i *formulas and* Π, Λ *are composed by* Π^b_i *formulas and* c *are all the variables of the sequent. Take* $G = \bigwedge\Gamma \wedge \bigwedge\{\neg\gamma | \gamma \in \Lambda\}$ *and* $H = \bigvee\Delta \vee \bigvee\{\neg\delta | \delta \in \Pi\}$. *Then there is a function* f *which is* $\Sigma^b_i - definable$ *in* $\mathsf{S}^{\mathsf{i}}_2$ *such that:*

(a) $f \in \Box^p_i$

(b) $\mathsf{S}^{\mathsf{i}}_2 \vdash Witness^i_G(w, c) \rightarrow Witness^i_H(f(w, c), c)$

Before proving this theorem, let us take a look of how we apply it. Suppose $\Gamma = \Pi = \Lambda = \emptyset$ and $\Delta = \exists y\theta(c, y)$. We apply Parikh's theorem to show $\Delta = \exists y \leq t\theta(c, y)$. By the main theorem there is a $\Sigma^b_i - definable$ function f that witnesses this formula. Take $\psi(x, y) \leftrightarrow y = (f(x))_1$ and notice that $g(x) = ((f(x))_1 \in \Box^p_i$.

*Proof of the Main Theorem.*By induction on the number of sequents in a free-cut free proof. Let us consider a proof of

$$\Gamma, \Pi \Longrightarrow \Lambda, \Delta$$

where for simplicity of exposition we assume $\Pi = \Lambda = \emptyset$. By the free-cut-free elimination theorem we can assume that, since Γ, Δ have all formulas in $\Sigma^b_i \cup \Pi^b_i$, the same holds for all formulas occurring in the proof. Since there are $\Sigma^b_i - \mathsf{PIND}$ inferences, all cut formulas will be in Σ^b_i. We will show here some relevant cases (the remaining cases as exercises).

(a) If $\Gamma \Longrightarrow \Delta$ is an initial sequent (BASIC axioms, logical axioms or equality axioms) these ar all composed by quantifier free formulas, hence by definition:

$$Witness^i_\theta(w,c) \leftrightarrow \theta(c)$$

and putting $f(n) = 0$ for all n this function satisfies the theorem.

(b) If the last inference has the form:

$$\frac{\alpha, \Theta \Longrightarrow \Delta}{\alpha \wedge \beta, \Theta \Longrightarrow \Delta}$$

Let D be $\alpha \wedge \bigwedge \Theta$ and E be $(\alpha \wedge \beta) \wedge \bigwedge \Theta$. By (IH) there is $g \in \Box^p_i$ which is Σ^b_i-definable in S^i_2 and this theory proves:

$$Witness^i_D(w,c) \rightarrow Witness^i_{\bigvee \Delta}(g(w,c),c)$$

Take $h(w) = \langle((w)_1)_1, (w)_2\rangle$ so that:

$$Witness^i_E(w,c) \rightarrow Witness^i_D(h(w),c)$$

and finally by putting $f(w,c) = g(h(w),c)$ we obtain:

$$Witness^i_E(w,c) \rightarrow Witness^i_{\bigvee \Delta}(f(w,c),c)$$

(c) If the last inference has the form:

$$\frac{\beta, \Theta \Longrightarrow \Delta \quad \gamma, \Theta \Longrightarrow \Delta}{\beta \vee \gamma, \Theta \Longrightarrow \Delta}$$

Let D be $\beta \wedge (\bigwedge \Theta)$ and E be $\gamma \wedge (\bigwedge \Theta)$ and F be $(\beta \vee \gamma) \wedge (\bigwedge \Theta)$. Apply (IH): hence there are $g, h \in \Box^p_i$ such that S^i_2 proves:

$$Witness^i_D(w,c) \rightarrow Witness^i_{\bigvee \Delta}(g(w,c),c)$$

$$Witness^i_E(w,c) \rightarrow Witness^i_{\bigvee \Delta}(h(w,c),c)$$

Let us define

$$f(w,c) = \begin{cases} g(\langle((w)_1)_1, (w)_2, c\rangle), & \text{if } Witness^i_\beta((w)_1)_1, c) \\ h(\langle((w)_1)_2, (w)_2\rangle, c) & \text{otherwise} \end{cases}$$

In other words, if w witnesses $(\beta \vee \gamma) \wedge (\bigwedge \Theta)$, then either $((w)_1)_1$ witnesses β or $((w)_1)_2$ witnesses γ, and $(w)_2$ witness $\bigwedge \Theta$. In the former case use g to find a witness of $\bigvee \Delta$; in the latter case use h. It follows:

$$Witness^i_F(w,c) \rightarrow Witness^i_{\bigvee \Delta}(f(w,c),c)$$

(d) If the last rule is:

$$\frac{a \leq t, \beta(a), \Theta \Longrightarrow \Delta}{\exists x \leq t \beta(x), \Theta \Longrightarrow \Delta}$$

(where a must not appear in the lower sequent) take for D the formula $a \leq t \wedge (\beta(a) \wedge \bigwedge \Theta)$ and E be $\exists x \leq t \beta(x) \wedge \bigwedge \Theta$. By (IH) there is a $g \in \Box^p_i$ such that:

$$Witness^i_D(w,c,a) \rightarrow Witness^i_{\bigvee \Delta}(g(w,c,a),c)$$

We consider two cases:

i. if $(\exists x \leq t\beta) \notin \Sigma^b_{i-1}$, then let $h(w) = \langle 0, \langle((w)_1)_2, (w)_2\rangle\rangle$, so that $Witness^i_E(w,c) \rightarrow Witness^i_D(h(w), c, ((w)_1)_1)$ so let $f(w,c) = g(h(w), c, ((w)_1)_1)$ and note that the theory proves:

$$Witness^i_E(w,c) \rightarrow Witness^i_{\bigvee \Delta}(f(w,c), c)$$

ii. if $(\exists x \leq t\beta) \in \Sigma^b_{i-1}$, then let $h(w) = \langle 0, \langle 0, (w)_2\rangle\rangle$ and $f(w,c) = g(h(w), c, (\mu x \leq t\beta(x))$

(e) If the last rule is

$$\frac{\Theta \Longrightarrow \beta(s), \Sigma}{s \leq t, \Theta \Longrightarrow \exists x \leq t\beta(x), \Sigma}$$

take D as $\beta(s) \vee (\bigvee \Sigma)$ and E be $s \leq t \wedge (\bigwedge \Theta)$ and F be $\exists x \leq t\beta(x) \vee (\bigvee \Sigma)$ and by (IH) we have:

$$Witness^i_{\bigwedge \Theta}(w,c) \rightarrow Witness^i_D(g(w,c), c)$$

But by definition we have $Witness^i_E(w,c) \rightarrow s \leq t \wedge Witness^i_{\bigwedge \Theta}((w)_2, c)$ so let $f(w,c) = \langle\langle s(c), (g((w)_2, c)_1\rangle, (g((w)_2.x)_2\rangle$ and note that:

$$Witness^i_E(w,c) \rightarrow Witness^i_F(f(w,c), c)$$

(f) Other logical rules are analized in a similar way (see Buss (1986)).

(g) If the last rule is a CUT:

$$\frac{\Gamma \Longrightarrow \Delta, \beta \quad \beta, \Gamma \Longrightarrow \Delta}{\Gamma \Longrightarrow \Delta}$$

Being the proof free-cut-free, β must be Σ^b_i. Take D be $\beta \vee (\bigvee \Delta)$ and E be $\beta \wedge (\bigwedge \Gamma)$, By (IH) there are $g, h \in \Box^p_i$ such that the theory proves:

$$Witness^i_{\bigwedge \Gamma}(w,c) \rightarrow Witness^i_D(g(w,c), c)$$

$$Witness^i_E(w,c) \rightarrow Witness^i_{\bigvee \Delta}(h(w,c), c)$$

Hence we define:

$$f(w,c) = \begin{cases} (g(w,c))_2 & \text{if } Witness^i_{\bigvee \Delta}((g(w,c))_2, c) \\ h(\langle (g(w,c))_1, w\rangle, c) & \text{otherwise} \end{cases}$$

Hence we obtain:

$$Witness^i_{\bigwedge \Gamma}(w,c) \rightarrow Witness^i_{\bigvee \Delta}(f(w,c), c)$$

(h) If the last rule is:

$$\frac{\beta(\lfloor \frac{1}{2} b \rfloor), \Theta \Longrightarrow \beta(b), \Sigma}{\beta(\overline{0}), \Theta \Longrightarrow \beta(t), \Sigma}$$

By (IH) if w witnesses $\beta(\lfloor \frac{1}{2} b \rfloor) \wedge \bigwedge \Theta$ there is $g \in \Box^p_i$ such that $g(w,b)$ witnesses $\beta(b) \vee \bigvee \Sigma$

Now let:

(a) $f(w,0) = \langle (w)_1, 0\rangle$; arguing as before, if w witnesses $\beta(\overline{0}) \wedge \bigwedge \Theta$, then $(w)_1$ witnesses $\beta(\overline{0})$ and therefore $\beta(\overline{0}) \vee \bigvee \Sigma$,

(b) Now let $f(w,b) = \langle X_0, X_1\rangle$, for $b > 0$, where:

i. $X_0 = g(\langle (f(w, \lfloor \frac{1}{2}b \rfloor)_1, (w)_2\rangle)_1$

ii. $X_1 = k((f(w, \lfloor \frac{1}{2}b \rfloor)_1, (w)_2)_2, (g(\langle f(w, \lfloor \frac{1}{2}b \rfloor, (w)_2\rangle)_2)$

and where:

$$k(v,w) = \begin{cases} v, & \text{if } Witness^i_{\bigvee \Sigma}(v,c) \\ w & \text{otherwise} \end{cases}$$

Note that if w witnesses $\beta(\overline{0}) \wedge \bigwedge \Theta$, then $f(w,b)$ witnesses $\beta(b) \vee \bigvee \Sigma$. The theory proves that:

(a) as we have seen, if w witnesses $\beta(\overline{0}) \wedge \bigwedge \Theta$, then $f(w,0)$ witnesses $\beta(0) \vee \bigvee \Sigma$.

(b) under the same condition, if $f(w, \lfloor \frac{1}{2}b \rfloor)$ witnesses $\beta(\lfloor \frac{1}{2}b \rfloor) \vee \bigvee \Sigma$, then $f(w,b)$ witnesses $\beta(b) \vee \bigvee \Sigma$.

(c) Hence by PIND we conclude that if w witnesses $\beta(\overline{0}) \wedge \bigwedge \Theta$, then $f(w,t)$ witnesses $\beta(t) \vee \bigvee \Sigma$.

To summarize:

(a) Buss's theorem says that if f is Σ^b_i-definable in $\mathsf{S}^{\mathsf{i}}_2$, then $f \in \Box^p_i$

(b) In case $i = 1$ we have that if f is Σ^b_1-definable in S^1_2, then f is polynomial time computable.

(c) Parson's theorem says that if f is Σ_1-definable in $I\Sigma_1$, then f is primitive recursive.

We remark that a proof of Parson's frequently mentioned theorem concerning the functions that can be defined in PRA can be obtained with Buss's method of the witness predicate too.

7.5. Further remarks and guide for further study

To conclude, let us make some proposals for further exploration of this topic. The literature on the weak fragments of arithmetic and their relation to computational complexity is endless and we therefore forego a priori the idea of giving a complete account of it, limiting ourselves to highlighting a few topics that we consider of particular importance, which can be approached with the tools we have introduced in the previous chapters.

A development worthy of consideration is the one who investigated what happens by narrowing the logical basis of theories. In this short presentation, we have dealt with theories based on classical logic: what can be said about intuitionist logic-based theories? *Intuitionistic Bounded Arithmetic* has been studied as well since Buss (1985): here a hierarchy of bounded formulas of the language of his theories is introduced, called $h - \Sigma^b_i$-formula (hereditarily Σ^b_i), i.e. those formulas that are Σ^b_i and whose subformulas are still Σ^b_i (for example, if $\phi \in \Pi^b_i$ and $\psi \in \Sigma^b_i$, then $\phi \to \psi \in \Sigma^b_i$, but ϕ is not). For each k, the theories $I - S^k_2$ are based on the $h - \Sigma^b_k$-induction schema (or rule) and for formulas in $h - \Sigma^b_k$ prove the excluded middle and the sability laws. Buss shows that if $\mathsf{I} - \mathsf{S}^{\mathsf{n}}_2 \vdash \forall x \exists y \phi(x,y)$, where ϕ is *arbitrarily complex* (note this difference with the classical case), then there exists a function $f \in \Box^P_n$, such that $\forall x \phi(x, f(x))$ is *true*. Actually, by a metod inspired ti Kleene's realisability, it is shown that the definable function f of $\mathsf{I} - \mathsf{S}^{\mathsf{n}}_2$ (i.e. those such that an arbitrarily complex formula ϕ exists such that $\phi(\overline{n}, \overline{f(n)})$ is true, for all n, and $\mathsf{I} - \mathsf{S}^{\mathsf{n}}_2 \vdash \forall x \exists! y \phi(x,y)$) are precisely those in $\Box^p_i$ (the class of functions computable in polynomial time with oracle in Σ^p_{i-1}, that coincides with

the class of predicates definable by a Σ^b_{i-1} formula). Further progress has been made in Cook and Urquhart (1993) and Harnik (1992).

We did not dwell on the theories T^i_2. However, a characterisation of the Σ^b_i-definable functions in T^{i-1}_2 is possible by proving that S^i_2 is $\forall\Sigma^b_i$-conservative on T^{i-1}_2 (where $\forall\Sigma^b_i$ means a universal quantifier followed by a Σ^b_i- formula). A different characterization of the Σ^b_1-definable (multivalued) functions of T^1_2 in terms of the so-called "polynomial local search problems" is given in Buss and Krajíček (1994). We only mention an important development, consisting in an application of the following well known result to these theories. Actually we consider an extension by definition of T^i_2, by adding symbols for all $\Box^p_{i+1}$ functions (which are Σ^b_{i+1}-definable in it) and the defining equations as axioms and obtain a kind of *Herbrand theorem.* Recall that a consequence of the cut elimination theorem for LK is that a cut free proof of a sequent $\Gamma \Longrightarrow \Delta$ where Γ, Δ are made of formulas in prenex normal form can be "divided" in two part. We define the *midsequent* as follows:

(a) If non quantifier rule occurs in the proof, then the midsequent is the last sequent.

(b) Otherwise, it is the topmost sequent which is a premiss of a quantifier rule (hence above there are only structural an propositional rules, and above only structural and quantifiers rules)

An application of this is the important *théorème fondamental* of Herbrand (1930). Let us consider a formula in prenex normal form, e.g. to fix the ideas:

$$\exists x \forall y \exists z \forall v \theta(x, y, z, v)$$

Let $f(x), g(x, y)$ new function symbols. Then for Herbrand's theorem that formula is provable in LK, iff there are terms $s_0, ..., s_n, t_0, ..., t_n$ such that the following:

$$\theta(s_0, f(s_0), t_0, g(s_0, t_0)) \vee ... \vee \theta(s_n, f(s_n), t_n, g(s_n, t_n))$$

is a propositional tautology (see Girard (1987) pp. 117-121 for a detailed proof). Variants of this have many applications to the problems we are discussing. A Herbrand-type theorem can be found in Krajíček, Pudlák, Takeuti (1991).

Theorem 118. *For $i \geq 1$, suppose $\phi(a, x, y)$ has the form $\exists \Pi^p_{i+1}$ and that T^i_2 proves $\exists x \forall y \phi(a, x, y)$. Then there are $\Box^p_{i+1}$-functions $f_0, ..., f_k$ such that T^i_2 proves:*

$$\phi(a, f_0(a), b_0) \vee \phi(a, f_1(a, b_0), b_1) \vee ... \vee \phi(a, f(a, b_0, .., b_{k-1}), b_k)$$

This research has important implications for an open problem: it is not known whether the hierarchy of theories of bounded arithmetic is proper, so we can hope that the connections between fragments of arithmetic and computational complexity that we have seen will also help us address the similar problem for the polynomial time hierarchy. The above result allows the following result to be deduced, due to Buss (1995), Krajíček, Pudlák, Takeuti (1991) and Zambella (1996).

Theorem 119. *If $\mathsf{T}^i_2 = \mathsf{S}^{i+1}_2$, then the polynomial time hierarchy collapses and this is provable in T^i_2.*

Last but not least, a natural continuation of the investigation around weak fragments of arithmetic leads one to consider *second-order* theories. The language of BASIC axioms is extended by adding second order variables $X^t, Y^s ...$ ranging over finite sets of numbers, where $t, s...$ are bounds to the value of the elements of the respective sets. We add also the membership relation $\in$, so that $x \in X^t$ is a new formula. The classes of formulas $\Sigma^{1,b}_i$

and $\Pi_i^{1,b}$ are introduced in analogy with the first order case counting the alternations of second-order quantifiers and not counting the alternations of first order quantifiers, where $\Sigma_0^{1,b}$ is the class of formulas with bounded first order quantifiers, but no unbounded quantifiers and no second-order quantifiers. The basic theory $\mathsf{I\Sigma_0^{1,b}}$ has the following axioms:

(a) BASIC axioms.

(b) $\forall X^t \forall Y^s \forall y \leq t + s(y \in X^t \leftrightarrow y \in Y^s) \rightarrow X^t = Y^s$ (extensionality).

(c) $\forall X^t \forall x \forall y (y \in X^t \rightarrow y \leq t(x))$.

(d) The scheme IND for $\Sigma_0^{1,b}$ formulas.

(e) $\mathsf{\Sigma_0^{1,b} - CA}$, i.e. the comprehension scheme:

$$\forall x \forall Y^x \forall y < x (y \in Y^x \leftrightarrow \theta(y))$$

(where $\theta \in \Sigma_0^{1,b}$).

The focus was mainly on these fragments, which have similarities with some fragments of the first order. We want to give an idea of what constitutes a key result linking the first- and second-order fragments, following Krajíček (1995) pp. 83-92. The second order family of fragments $\mathsf{V_1^i}$ actually had different presentations in different works. Each theory $\mathsf{V_1^i}$ is however equivalent to the above theory $\mathsf{I\Sigma_0^{1,b}}$ plus IND on all $\#-$free $\Sigma_i^{1,b}$ formulas. The family of fragments denoted $\mathsf{U_1^i}$ is obtained analogously but with PIND in place of IND. The families $\mathsf{U_2^i}$ and $\mathsf{V_2^i}$ are obtained in a similar manner, but admitting the respective induction schemes on $\Sigma_i^{1,b}$ formulas in the full language with $\#$. The strong analogy between first- order and second order fragments has the name of "RSUV isomorphism" and was highlighted by Takeuti (1993) and Razborov (1993). It is shown that there are translations $*, \circ$ between first-order and second-order languages such that:

(a) if $\phi \in \Sigma_\infty^{1,b}$, then $\phi^* \in \Sigma_\infty^b$

(b) if $\psi \in \Sigma_\infty^b$, then $\phi^\circ \in \Sigma_\infty^{1,b}$.

This translation fulfils the following conditions, linking fragments of the first and second order:

(a) if $S_2^i \vdash \psi$, then $V_1^i \vdash \psi^\circ$

(b) if $V_1^i \vdash \phi$, then $S_2^i \vdash \phi^*$

(c) $S_2^1 \vdash \psi \leftrightarrow (\psi^\circ)^*$

(d) $V_1^1 \vdash \phi \leftrightarrow (\phi^*)^\circ$.

The same relation subsist between $\mathsf{U_1^i}$ in place of $\mathsf{V_1^i}$ and R_2^i in place of S_2^i, where R_2^i is obtained from S_2^i replacing the PIND rule with the rule:

$$\frac{\Gamma, \phi(\lfloor \frac{1}{2} b \rfloor) \Longrightarrow \Delta, \phi(b)}{\Gamma, \phi(\overline{0}) \Longrightarrow \Delta, \phi(|t|)}$$

for $\phi \in \Sigma_i^b$ and adding the language the functions minus $\dot{-}$ and msf:

(a) $\mathsf{msf}(a, 0) = a$

(b) $\mathsf{msf}(a, i+1) = \lfloor \frac{1}{2}\mathsf{msf}(a, i) \rfloor$.

The fragments $\mathsf{R}^{\mathsf{i}}_3$ and $\mathsf{S}^{\mathsf{i}}_3$ include in the language the function $x\#_3 y = 2^{|x|\cdot|y|}$. The RSUV isomorphism extends to $\mathsf{S}^{\mathsf{i}}_3$ and $\mathsf{V}^{\mathsf{i}}_2$ as well as to $\mathsf{R}^{\mathsf{i}}_3$ and $\mathsf{U}^{\mathsf{i}}_2$. The main connection of these second-order fragments with computational complexity theory is condensed in this result:

(a) the $\Sigma^{1,b}_1$-definable functions of $\mathsf{U}^{\mathsf{1}}_2$ are the PSPACE-computable.

(b) the $\Sigma^{1,b}_1$-definable functions of $\mathsf{V}^{\mathsf{1}}_2$ are the EXPTIME-computable.

In other words, these theories have proof-theoretic strengths corresponding to polynomial space and exponential time computation. It is not known whether these two complexity classes coincide. Likewise, we know that $\mathsf{U}^{\mathsf{i}}_2 \subseteq \mathsf{V}^{\mathsf{i}}_2 \subseteq \mathsf{U}^{\mathsf{i+1}}_2$, but it is not known whether the theories $\mathsf{V}^{\mathsf{1}}_2$ and $\mathsf{U}^{\mathsf{1}}_2$ are different. See for instance Buss, Krajíček and Takeuti (1993) and Buss and Beckmann (2014) for further investigations and for improved witnessing theorems. Another interesting chapter (which we will not open for lack of space and to avoid excessive scattering) is that of the relationship with propositional systems and their complexity, of propositional proof systems corresponding to certain first-order and second-order bounded arithmetic theories, according to certain translations. We refer to Cook and Nguyen (2010), Buss (1997) and Krajíček (1995) for an extensive presentation.

8. Random sequences, incompleteness and information

8.1. What is a random sequence?

> When a rule is extremely complex, that which conforms to it passes for random. (Leibniz)

Another mathematically significant development in the research around the phenomenon of incompleteness on which we consider it important to focus attention, is that which has led to highlighting the link between this concept and that of randomness and information. Grigory Chaitin, starting in the 1970s (see the bibliography), reformulated the incompleteness theorems within the framework of algorithmic theory of information, presenting his results as a "dramatic extension" of the phenomenon already highlighted by Gödel. He claims that high complexity is the alleged reason of the unprovability of infinitely many true sentences: a true statement that can be expressed in the language of a theory is unprovable because its information content is greater than that of the axioms of that theory itself; or in other words, in a formal system no number can be proved random unless its complexity is less than that of the formal system itself.

At the heart of Chaitin's work, therefore, is the concept of the complexity of an object and the measure of the difficulty of describing it. *The mathematical concept of randomness* is an attempt to give an idealized model of *randomness*, as the recursive functions do in the case of computability.

When a sequence of numbers is random? Some solutions are not completely satisfactory: Borel (1909) introduced the notion of 'normality' of a sequence of decimal digits, namely the property for which the frequency of each block of digits of length $k \geq 1$ in each finite initial segment of length m of that sequence approximates to 10^{-k} as m goes to the infinity (see Calude (1994) for a thorough analysis). More precisely let us define a *basis*, namely a number $b \geq 2$ and let a *digit* in base b an element of $\{0, 1, 2..., b-1\}$.

A *block* in base b is a finite sequence w of digit and if we denote $w[i, j]$ a *subblock* from i to j of w (where $1 \leq i \leq j \leq |w|$), then let $occ(w, u)$ the cardinality of the set $\{i|w[i, i+|u|-1] = u\}$, i.e. the number of occurrences of the string u in w. For instance, if $w = 2122113211$ and $u = 211$ and $occ(w, u)$ =cardinality of the set $\{3, 8\}$, namely 2.

The expansion of $r \in [0, 1]$ in base b is given by $(r)_b = \sum_{i=1}^{\infty} a_i b^{-i}$, where the a_i are digits of the basis. The real number r is called *normal* for the basis b, if for each block u:

$$\lim_{n\to\infty} \frac{occ((r)_b[1, n], u)}{n} = \frac{1}{b^{|u|}}$$

It is *absolutely normal*, if this holds for every basis. Some important facts in this regard are the following:

Duccio Pianigiani, University of Siena, Italy, duccio.pianigiani@unisi.it, 0000-0001-9441-7226

Referee List (DOI 10.36253/fup_referee_list)
FUP Best Practice in Scholarly Publishing (DOI 10.36253/fup_best_practice)

Book References DOI 10.36253/979-12-215-0778-2.references

(a) (Borel 1909). Almost all real numbers are absolutely normal.

(b) (Turing 1938). There is a computable absolutely normal number.

(c) (Champernowne 1933) The number:

$$0,1234567891011121314151617181920212223 24....$$

is normal in base ten.

Champernowne's number, in particulart, shows that normality is not a fully satisfactory definition of 'randomness', because its development is easily predictable. Some other attempts of characterizations of random sequences link randomness to the stability of frequency. However 1010101010... has this feature, while not seeming entirely random. In particular, notice that there is a subsequence (the one composed of only places even) that does not obey this law.

Randomness is associated with the idea of the growth of disorder, i.e. the decreasing of information, but even this is sometimes misleading. Indeed the string 01101010000010011110011 although apparently random, consists of the first 23-digit of the binary expansion of $\sqrt{2}-1$.

Already in Kolmogorov (1965), among the known approaches to the problem of the quantitative definition of information, besides that in terms of algorithmic complexity due to the author himself, is mentioned the concept of entropy introduced in Shannon (1948). The general concept of entropy in thermodynamics, as is well known, was introduced by Clausius in the middle of the nineteenth century, and its success is linked (also beyond its original scope) to names such as Boltzmann, Gibbs, Shannon and Von Neumann. The properties of Shannon's concept of entropy and those of Kolmogorov, Solomonov and Chaitin's concept of complexity are in many respects similar: both constitute bit-based measures of information; in both the information conveyed by an object depends on the length of its description. Propositions formulated in the terms of one can be reformulated in the terms of the other, and Romashchenko's theorem (see Hammer, Romashchenko, Shen and Vereshchagin (2000)) establishes, in very general terms, that any linear inequality true for Kolmogorov's complexity is also true for Shannon's entropy, and vice versa.

Considering infinite binary sequences, four principal characterizations have been provided:

(a) in terms of stability of frequency (Von Mises, Wald, Church),

(b) in terms of incompressibility (Solomonov, Kolmogorov, Chaitin)

(c) in terms of typicity (Martin-Löf)

(d) in terms of unpredictability (the theory of martingale).

As for unpredictability, a sequence $\sigma \in 2^{\omega}$ is called random, if given its initial segment $\sigma \upharpoonright n$ its first n bits we cannot predict the next bit. For example, the outcome of an ideal coin is unpredictable in the sense that the knowledge of the first n-outcomes do not helps to predict the next. We will not deal here of martingale theory. Jean Ville invented martingales in the 1930s in order to improve Richard Von Mises' concept of a collective, and Claus-Peter Schnorr made martingales algorithmic in the 1970 and characterized Martin-Löf randomness in terms of martingales. Random sequences, as equivalently defined in 2,3,4, although chaotic, nevertheless may have a strong computational power: we will see a particular sequence that computes the Halting Problem. More surprising is the result that was originally obtained from Gács (1986) and Kučera (1985) who proved the following result.

Theorem 120. *Any sequence is Turing-reducible to a random sequence.*

Contrary to what one might think, not only is it not true that no useful information can be extracted from random sequences, but on the contrary it seems that many random sequences are able to " calculate everything ":

Any type of information that can be coded into an infinite binary sequence, no matter how structured that might be, can be obfuscated into an algorithmically random infinite binary sequence, from which it is effectively recoverable (Barmpalias and Lewis-Pye (2018)).

8.2. From Von Mises to Martin-Löf

We start from Von Mises' *Kollektive* to understand how we arrive at the notion of the Martin-Löf test. In addition to Borel, the first remarkable attempt to formalize the 'random sequence' notion was made by Richard Von Mises in the 20s of '900 in the context of his investigation into the foundations of probability (see Zaffora Blando (2024), Van Lambalgen I (1987) and Van Lambalgen II (1987) among the extensive bibliography of this author for a detailed discussion). The problem of giving an axiomatization of the concept of probability is the sixth of the well-known list of twenty-three problems compiled by Hilbert in 1900. In 1919, Richard Von Mises, a member of the Vienna Circle and probabilist of the frequentist school, presented its axiomatization based on the concept of *Kollektiv*, i.e. on a certain type of abstract characterization of an infinite sequence of independent trials, that meets certain global and local regularities, about the extraction of infinite subsequences. We limit ourselves in this short exposition to the case where the results of a possible experiment are 0 or 1. We denote by $2^{<\omega}$ the set of finite binary strings and with 2^{ω} the set of *infinite* binary strings (Cantor space).

A binary sequnce $\sigma \in 2^{\omega}$ is a *Kollektiv* if and only if it satisfies the following conditions:

(a) if $S_n = \sum_{i \leq n} \sigma(i)$, then $\lim_{n\to\infty} \frac{S_n}{n}$ exists and is a real p in $[0, 1]$ (considering a uniform distribution generated by the toss of a fair coin then $p = \frac{1}{2}$).

(b) If R : is a "selection rule" extracting a subsequence $R(\sigma)$ of σ, then this subsequence satisfies point 1. with the same probability:

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i \leq n} (R(\sigma)_i) = p$$

namely, extracting infinite subsequences, the stability of the frequency is preserved (impossibility of a *gambling system*).

The axiom 1. guarantees that the limiting relative frequencies of 1 and 0 exist. The meaning of point 2. is that a sequence should count as random if all infinite subsequences have the same frequency of 0's as 1's in the limit. A sequence such as 101010... clearly should not count as random in this sense. In other words, we want to guarantee that the sequence "cannot be gamed" by identifying a scheme that can be exploited to devise a successful gambling system. These 'collectives' are the formal counterpart of the notion of 'random sequences'.

However, it was necessary to better define the notion of "selection rule of a place". If we admit arbitrary selection rules, there is no collective *strictu sensu* (the so-called "Kamke's argument": take $R = \langle n | \sigma(n) = 1 \rangle$, extracting the sequence 1111...). It must nevertheless be emphasised that this objection is its strongly non-constructive character and nevertheless highlights the need to better denote the notion of a place-selection rule. Kamke (1932) therefore proposed to restrict the sequences which may represent admissible place selections to those which are lawlike; Wald (1937), accepting this criticism, demonstrated that a collective exists (thus the very notion of a collective is consistent), as long as it is restricted to a countable set of rules. Among the attempts to specify the notion of the 'collective' Church (1940) proposal to take as admissible place selections the computable functions $f : 2^{<\omega} \to \{0, 1\}$ actually constitutes the first definition of algorithmic randomness. However the most sharp criticism of Von Mises'

definition came from Ville (1939): he showed that in a sense the collective are not random enough, as there are some collectives with limit frequency of '1 'equal to 1/2, but $\forall n(\frac{S_n}{n} \geq \frac{1}{2})$. This collective violates a fundamental probabilistic law known as *Law of iterated logarithm*. This law implies that for almost all sequences frequency reaches its limit exhibiting large oscillations above and below the origin, while the Ville sequence reaches its limit "from above". Ville actually proved the following theorem.

Theorem 121. (Ville 1939) *For each countable set of rules of selection of a place, there is a collective that violates the law of the iterated logarithm.*

He suggested that a limit of the collectives was to obey a single law of chance, the law of large numbers, *and not to all these laws*, as we would expect from a random sequence. A random sequence has to be 'typical', in the sense of 'noexceptional ': we do not consider as 'typical' a sequence that differs from others for some peculiarities, such as contain a finite number of '1 '; a binary sequence is therefore considered *typical* , if it is not rare, it has no specific feature.

Martin-Löf proposed a satisfactory solution to the objections addressed to the notion of *Kollektiv* as a formal account of the concept of 'random sequence', establishing a relationship between stochastic computable properties and particular subsets of zero-measure of 2^ω: are considered *random* the sequences that do not fulfils property *effectively rare*. We distinguish between *typical* sequences and *special* sequences: *typical* means not exceptional, ordinary, as a synonym for *random*. More formally, a sequence $\sigma \in 2^\omega$ is *typical* if and only if every set of sequences each containing only a small number of sequences, does not contain σ (the statement should not be taken literally, since $\sigma \in \{\sigma\}$ and therefore each individual is not typical). It is not a typical sequence, the sequence that is distinguished by a specific property, for example, to possess a finite number of 1; the probability of obtaining a non-typical binary sequence with the launch of a correct coin is 0. Conversely, the probability of getting a typical sequence is 1. A typical sequence has all the opposite properties with respect to a non-typical sequence. It meets all the " laws of chance ", that is, the properties of which it is proved that they have the value 1.

Let therefore $\sigma \in 2^{<\omega}$, where $2^{<\omega}$ is the set of *finite* binary strings, and let $[\sigma]$ the set of *infinite* binary sequences that begin with σ, namely that extend it, also called cylinders (the basic open of the product topology on 2^ω). We can read $[\sigma]$ as: "the first $|\sigma|$ flips of a coin gave σ". Mostly the *uniform measure* is used an it is given by:

$$\mu([\sigma]) = 2^{-|\sigma|}$$

(where $|\sigma|$ denotes the length of σ) i.e. the probability that a $\tau \in 2^\omega$ obtained flipping a fair coin will be in $[\sigma]$, or the probability that a $\tau \in 2^\omega$ begins with σ. The measure μ, defined on cilynders, can be extended to all sets $A \subseteq 2^\omega$. In general $\mu(A)$ is the probability that $\omega-$flips of a fair coin give rise to a binary sequence belonging to A. (see Nies (2009) or Downey and Hirschfeldt (2010) for all measure-theoretic detalis).

A probabilistic law, theoretically, is a set of measure-one and therefore, it will have the form $\mu\{\sigma \in 2^\omega | A(\sigma)\} = 1$. So, coming back to Ville's argument, if $A_0, A_1, A_2...$ is the collection of all sets such that $\mu(A_i) = 1$, then a sequence σ is typical, if belong to all these A_i, namely if $\sigma \in \bigcap_i A_i$. However notice that for the singleton $\{\sigma\}$, we have $\mu(\{\sigma\}) = 0$, since $\{\sigma\} \subseteq [\sigma_{\restriction n}]$, for all n and therefore $\mu(\{\sigma\}) \leq \mu([\sigma_{\restriction n}]) \leq 2^{-n}$, for all n. It follows that $\mu(\overline{\{\sigma\}}) = \mu(2^\omega \setminus \{\sigma\}) = 1$. But then each sequence is not in a set of measure 1, that is, the complement of its singleton. Ergo there are no " typical sequences " at all. We must restrict the collection of sets measure 1. We will say that a sequence that satisfies rare property (i.e. of zero measure) *effectively given*, is not random. Instead of belonging to all sets of measure one, we will say that a sequence is random, if avoids all null sets effectively given. A set X is *null*, if there is an infinite sequence of open sets $\{V_i\}_{i\in\mathbb{N}}$ such that $\mu(V_n) \leq 2^{-n}$ and $X \subseteq \bigcap_n V_n$. There are actually equivalent definitions of *null sets* which, being often used, are worth mentioning:

(a) A set A is *null*, if has measure zero.

(b) A set A is *null*, if there is an infinite sequence of open sets $V_0, V_1, V_2....$ such that $A \subseteq \bigcap_n V_n$ and $\mu(V_n) \leq 2^{-n}$.

(c) A set A is *null*, if there is an infinite sequence of open sets $V_0, V_1, V_2....$ such that $A \subseteq \bigcap_n V_n$ and $\lim_n \mu(V_n) = 0$.

What is meant by "effectively given"? Martin-Löf (1966) narrowed the concept of a "typical" sequence in order to avoid the difficulties that we have mentioned: we will not ask that a *random* sequence belongs to each set of measure one, i.e. that does not belong to no null set, but only that does not belong to null sets *effectively given*: this notion is, so to say, *costructivized*, considering only the laws of probability that can be proved in an effective way (it is provable that not all the laws of probability are effective in this sense). Hence, a sequence is random, if it meets all the probabilistic laws effectively given, i.e. belongs to all sets effectively given of measure one (i.e. all those we have called " typical " properties), or, equivalently, is *not* a member of any set of *zero measure* effectively given.

This brings us to the concept of *test*. If A is computably enumerable, that is, if $A = W_e$, then we speak of sets *effectively* open. The computably enumerable opens can be represented as:

$$X = [W_e] = [Dom(\phi_e)] = \{\tau \in 2^\omega | \sigma \subset \tau, \ for\ some\ \sigma \in W_e\}$$

for some index e. Formally a *Martin-Löf test* is a uniformly computably enumerable sequence of open sets $\{V_i\}_{i\in\mathbb{N}}$ such that for each $i \in \mathbb{N}$, $\mu(V_i) \leq 2^{-i}$. Thinking to an enumeration of computably enumerable sets $W_0, W_1, W_2...$, "uniformly" computably enumerable means that a test is univocally determined by an index e of a function $\phi_e(x)$ such that:

$$V_i = \bigcup_{\sigma \in W_{\phi_e(i)}} [\sigma] = [W_{\phi_e(i)}]$$

Notice that $\mu(\bigcap_n V_n) = 0$, namely is *effectively null* (sometimes is assumed that $V_i \supseteq V_{i+1}$, however this is not essential, since considering $U_n = \bigcup_{m>n} V_m$, we have that $\bigcap_i U_i = \bigcap_i V_i$ and the previous condition is satisfied for $\{U_i\}_i$).

Definition 61. *We say that:*

(a) An infinite sequence σ pass the test of $\{V_i\}_{i\in\mathbb{N}}$, *if* $\sigma \notin \bigcap_n V_n$.

(b) An infinite sequence σ *is* ML-random, *if pass all tests* $\{V_i\}_{i\in\mathbb{N}}$, *i.e. avoids all countable intersections of such null sets.*

This is the way in which in this formalism we express the fact that a sequence has no attribute "non-typical". An important concept is that of *Universal test.* It is possible to give an enumeration of the tests, starting from a enumeration of computably enumerable sets, discarding those that are too large:

$$W_{g(e,i),s} = \begin{cases} W_{\langle e,i\rangle,s} & \text{if } \mu\{[\sigma] | \sigma \in W_{\langle e,i\rangle,s}\} \leq 2^{-i} \\ \emptyset & otherwise \end{cases}$$

(where $n \in W_{e,s}$ if and only if $\phi_e(n)$ converges in at most s steps). If we take:

$$V_i = \bigcup_e \{[\sigma] | \sigma \in W_{g(e,i+e+1)}\}$$

we observe that $\mu(V_i) \leq \sum_e 2^{-(i+e+1)} \leq 2^{-i}$. The sequence $\{V_i\}_{i\in\mathbb{N}}$ costitutes a *universal* test, since for all other test $\{Z_i\}_{i\in\mathbb{N}}$, we have $\bigcap_i Z_i \subseteq \bigcap_i V_i$.

We will go no further in our discussion of the Martin-Löf approach because the work we will refer to is based on another of the previously mentioned equivalent notions of randomness: that of incompressibility.

8.3. Kolmogorov-Solomonoff-Chaitin's theory and incomputability

To account for the definition given by Solomonoff (1964) and Kolmogorov (1965), later adopted with some variations by Chaitin, let's start highlighting a few fundamental underlying ideas. Reasoning in general terms, certain seemingly random sequences can be described by relatively simple means, while the other, as for example 1001110111010110 does not seem have other description, except that by the mere repeatition *verbatim*: we will say that such sequences are *incompressible*, because their descriptive complexity is at least equal to their length. Reasoning more abstractly, if $\tau_0\tau_1\tau_2...\tau_n$ is a binary string (i.e. a member of the set $2^{<\omega}$ of finite binary sequences), it might be generated by a "shorter program" of the string itself, coded for example with k bits, for $k < n+1$: a finite string will be considered *random*, if cannot be generated by a program shorter than itself.

From this arises the following preliminary definition, relative to a Turing machine $\mathcal{M}$ (remember that $|w|$ denotes the length of the binary string w):

$$K_{\mathcal{M}}(x) = \begin{cases} |w| & w = \text{the shortest sequence such that } \mathcal{M}(w) = x, \text{ if exists} \\ \infty & \text{otherwise} \end{cases}$$

However, in order to disengage the definition from the particular machine $\mathcal{M}$, this was formulated rather in terms of a universal Turing machine U. For example $U(0^e1\sigma) \simeq \mathcal{M}_e(\sigma)$. Let therefore U be a universal machine; the Kolmogorov complexity of a finite stringe w with respect to U is given by the following function:

$$K_U(w) = min\{|z| \mid U(z) = w\}$$

The function $K_U(w)$ has to be readen "the length of the shortest program that (with respect to U) gives a description of w". This definition is related to the previous one by the fact that for each machine $\mathcal{M}$ there is a constant $c_{\mathcal{M}}$ such that $K_U(w) \leq K_{\mathcal{M}}(w) + c_{\mathcal{M}}$. Hence for two universal machines the Kolmogorov complexity is the same up to an additive constant, in the sense that if U, U' are two universal machines, we have that for some constant c, $K_U(w) \leq K_{U'}(w) + c$. We can therefore fix once and for all U and simply write $K(w)$. It is generally true that $K(w) \leq |w| + s$, for some constant s, and then we are talking of a total function.

Incompressible strings really exist? Yes!

Theorem 122. *There are incompressible strings of any fixed length n.*

Proof. Observe that there are $\sum_{k=0}^{n-1} 2^k = 2^n - 1$ programs of length minus than n, while the programs of length n are exactly 2^n:

$$\begin{array}{cccccccc} \emptyset & 0 & 1 & 00 & 01 & 10 & 11... \\ 0 & 1 & 2 & 3 & 4 & 5 & 6... \end{array} \tag{1}$$

If the programs σ of length n were all compressible, we would have $K(\sigma) < n$: hence for all σ of length n we would have a τ of length less than n that print it. There would therefore be a τ of this kind that prints two different programs (a contradiction). QED

Lemma 36. *There is a constant e such that $K_U(\sigma) \leq |\sigma| + e$.*

Proof. Let us take $\mathcal{M}_e(\sigma) = \sigma$ the *copying-machine*; hence σ is a program of $\mathcal{M}_e$ that prints σ: at worst, therefore, the minimum program of U that prints σ will have length $0^e 1\sigma$. QED

Lemma 37. *For all partial recursive function ϕ_e, $K(\phi_e(\tau)) \leq K(\tau) + c$.*

Proof. Let σ be of minimal length such that $U(\sigma) \simeq \tau$ and let:

$$Z(0^{\log(e)}1e\sigma) \simeq \phi_e(U(\sigma)) \simeq \phi_e(\tau)$$

Hence $K(\phi_e(\tau)) \leq |0^{\log(e)}1e\sigma| \leq 2|e| + 1 + |\sigma| \leq K(\tau) + c$. QED

An element of surprise and dissatisfaction with respect to this complexity measure, however, is its *character not additive*, that is, the complexity of the concatenation $\sigma = \tau\alpha$ may be bigger than the sum of the complexity of σ plus that of α. From an algorithmic point of view, a fact even more troublesome is the following.

Theorem 123. *$K(x)$ is not a computable function.*

Proof. Suppose that on the contrary it iscomputable. Suppose that we have ordered lexicographically the finite binary sequences $\emptyset < 0 < 1 < 00 < 01 < 10 < 11...$ and let:

$$y_m = min\{\sigma | K(\sigma) > m\}$$

where the minimum is taken with respect to the order $<$; let us consider all these $y_0, y_1, y_2.....$

(a) if $K(x)$ were computable, then there would be c such that $K(y_m) < |m| + c$. Indeed, let M be a machine that on n, first generates the strings $\sigma_0, \sigma_1, \sigma_2...$ in lexicographic order and the computes $K(\sigma_0), K(\sigma_1), K(\sigma_2)...$:

   i. if $K(\sigma_i) > n$, it prints σ_i and then halts.
   ii. Otherwise analyzes σ_{i+1}

Sooner or later will come a σ_i such that $K(\sigma_i) > n$; so, on input n, the program will produce y_n

(a) Hence $K_M(y_n) \leq |n|$ and we know thata in this case $K(y_n) \leq |n| + c$.

(b) Ergo $n < K(y_n) \leq |n| + c$, from which $n < |n| + c$, contradiction, because asymptotically n grows more than $|n| + c$.

QED

Although incomputabile, the function $K(x)$ is nevertheless *approximable*, i.e. there is a computable function g such that:

(a) $g(s + 1, x) \leq g(s, x)$ (decreasing in s)

(b) $\lim_{s\to\infty} g(s, x) = K(x)$ (computable from above, or right-computably enumerable)

Recall that in general if g is of this kind and $f(x) = \lim_{s\to\infty} g(s, x)$, then the set:

$$X = \{\langle\sigma, n\rangle \in 2^{<\omega} \times \mathbb{N} | f(\sigma) < n\}$$

is computably enumerable (and viceversa). Hence the set:

$$\{\langle\sigma, n\rangle \in 2^{<\omega} \times \mathbb{N} | K(\sigma) < n\}$$

is computably enumerable non computable.

Hence the incomputability of $K(x)$ can also be seen from an angle more abstract appealing to the notion of *simple set*, due to Post (see p. 49). Recall that a simple set can not be computable.

Theorem 124. *The set* $X = \{\sigma \in 2^{<\omega} | K(\sigma) < |\sigma|\}$ *is simple.*

Proof. By contradiction let $Z \subseteq \overline{X}$ computably enumerable and infinite. Remember that each computably enumerable set and infinite contains a computable set (Post), say $A = \{z_0, z_1, z_2....\}$. Notice that $K(z_i) \leq |i| + c$, because the program that prints z_i is obtainable from the machine that generates A and from the index i, that in the binary representation has length $\log(i)$. But for a z_i big enough this contradicts the fact that it is incompressible, as an element of $\overline{X}$. QED

We would like now to extend the notion of *incompressible string* to infinite sequences (i.e. real numbers between 0 and 1), that is, we would like to say that σ is random, if and only if for every n, $K_U(\sigma \upharpoonright n) \geq n$, but we will point out in this regard that machines then it must be understood as *prefix-free*, i.e. their domain will consist of binary strings, none of which is an initial segment of the other: for a result of Martin-Löf, dropping the restriction on machines to be *prefix-free*, we would have an empty definition, since no infinite sequence would satisfy the above condition.

Theorem 125. *For all* α *and infinite* n, $K(\alpha \upharpoonright n) < n - c$ *(where* $w \upharpoonright n =$ *the prime* n *bits of* w*).*

Proof. Here we see a version due to Katseff (1978): we write $\log(n)$ for abbreviating $\lfloor \log_2(n+1) \rfloor$="the integer part of $\log_2(n+1)$"; let f be the canonical correspondence between binary sequences and numbers that we have already used (a string σ is identified with the number n such that 1σ is the base-two representation of $n+1$), observing that $|f(n)| = \log(n)$. We consider a Turing machine $\mathcal{M}$ doing this:

(a) on input σ, look if $|\sigma|$ has the shape $k - \log(k)$, namely $k - |f(k)|$ and if the answer is yes, then it returns as output $f(k) \frown \sigma$ (recall that we used $\frown$ for concatenation of binary strings).

For instance if $\sigma = 0001 = f(16)$, then $|\sigma| = k - \log(k) = 4$, where $k = 7$, $\log(k) = 3$ and therefore $k - \log(k) = 4$. In this case the machine returns $f(7) \frown \sigma = 000 \frown 0001$.

Consider now that $\sigma = f(n)$ can be seen as the initial segment $f(n) = \sigma \upharpoonright m$ of an infinite sequence, fors ome m (in the previous example, $m = 4$ and $n = 16$); therefore take $\sigma \upharpoonright n = f(n) \frown \tau$ where $|\tau| = n - |f(n)| = n - \log(n)$ (once more following the previous example, $\sigma \upharpoonright n = \sigma \upharpoonright 16 = 0001 \frown \tau$, where τ contains 12 bits). Ergo, the machine $\mathcal{M}$ on input τ will return $f(n) \frown \tau$.

In conclusion, considering that:

$$K_{\mathcal{M}}(\sigma \upharpoonright n) = min\{|\alpha| | \mathcal{M}(\alpha) \simeq f(n)\tau\} =$$

$$= min\{|\alpha| | \mathcal{M}(\alpha) \simeq \sigma \upharpoonright n\} \leq |\tau|$$

for infinite n we will have $K_U(\sigma \upharpoonright n) \leq |\tau| + e \leq n - |n| + c$. QED

The problem inherent in the definition of $K(x)$ we used hitherto, has been highlighted by Chaitin with this example, consider a machine $\mathcal{M}$ that on input σ, first scans it for determine its length $|\sigma| = n$, then go back to start the computation of σ bit by bit. If now $\mathcal{M}(\sigma) \simeq \tau$, then the information necessary to get τ, intended as $K_M(\tau)$, is not dependent only by σ, but also by $|\sigma|$ and will be then encoded encoded by a string whose length is of order $n + \log(n)$. The problem does not occur if, for example, we consider machines *self-delimiting*: think to a Turing device where there is a single input tape of only reading, whose head flows only from left to right, a certain finite number of working tapes and an output tape.

The *self-delimiting* machines are not authorized to affix a symbol $*$ of "white" at the end of the input, which therefore is not delimited by any *end marker*; the input tape contains therefore only digit 0 and 1: a machine M on input σ, gives as a result τ, if after reading all σ, but the next bit, print τ; it is *self-delimiting* in the sense that it can determine where the input terminates, without the need to read the next symbol. The machine converges on $\sigma \in 2^{<\omega}$, if after a finite number of steps halts reading the last bit of the input tape and produces output. It is interesting to note that if $\mathcal{M}$ is *self-delimiting*, the set:

$$Dom(\mathcal{M}) = \{\sigma \in 2^{<\omega} | \mathcal{M}(\sigma) \downarrow\}$$

is *prefix-free*. For domains without prefixes, an important inequality, known as "Kraft's inequality", applies. The following, more refined version is actually due to Chaitin and it is the one which we use in Levin and Schnorr's theorem.

Theorem 126. (Kraft-Chaitin inequality) *Suppose we have an effective list:*

$$\langle n_0, \sigma_0 \rangle, \langle n_1, \sigma_1 \rangle, \langle n_2, \sigma_2 \rangle ...$$

(called bounded request), such that $\sum_{k \in \mathbb{N}} 2^{-n_k} \leq 1$. *Then it is possible to define a prefix-free machine such that for all* k, *there is a* τ_k *such that* $|\tau_k| = n_k$ *and the machine on input* τ_k, *outputs* σ_k.

Theorem 127. (Levin-Schnorr (1973)) *Let* $\sigma \in 2^{\omega}$. *Then the following are equivalent:*

(a) σ *is* ML*-random.*

(b) There exists a number b *such that for all numbers* n, $K(\sigma \restriction n) > n - b$

Proof. The theorem was proven independently in Levin (1973) and Schnorr (1973). $1 \Rightarrow 2$ follows from the simple observation that the cylinders $R_b = [\{\tau | K(\tau) \leq |\tau| - b\}]$ give rise to Martin-Löf test $\{R_b\}_{b \in \mathbb{N}}$. Actually $K(\tau) \leq |\tau| - b$ if and only if $\exists \sigma \exists s (U_s(\sigma) \simeq \tau \wedge |\sigma| \leq |\tau| - b)$; it is therefore a Σ^0_1 formula and the sequence $\{R_b\}_{b \in \mathbb{N}}$ can be expressed in a uniformly computably enumerable way.

Let now X_b be the set of σ such that $K(\sigma) \leq |\sigma| - b$. Hence $\mu(R_b) = \sum_{\sigma \in X_b} 2^{-|\sigma|}$. Notice that if $\sigma \in X_b$, then $|\sigma| \geq K(\sigma) + b$. Moreover using Kraft's inequality (since $Dom(U)$ is prefix-free):

$$2^{-b} \cdot \sum_{\sigma \in X_b} 2^{-K(\sigma)} \leq 2^{-b} \cdot \sum_{\sigma \in Dom(U)} 2^{-|\sigma|} \leq 2^{-b} \cdot 1$$

Hence:

$$\sum_{\sigma \in X_b} 2^{-|\sigma|} \leq 2^{-b} \cdot \sum_{\sigma \in X_b} 2^{-K(\sigma)} \leq 2^{-b} \cdot 1$$

Namely $\mu(R_b) \leq 2^{-b}$. Now, if σ does satisfy 1. then $\sigma \notin \bigcap_b R_b$ and therefore there is a b such that $K(\sigma \restriction n) > n - b$ for all n; hence does satisty 2.

As regards instead $2 \Rightarrow 1$, suppose that $\sigma \in \bigcap_m V_m$, for some test $\{V_m\}_{m \in \mathbb{N}}$. Remember that each open computably enumerable can be represented (uniformly in m) in the form:

$$V_m = [\{\alpha_{i,m} | i < k_m\}]$$

where $\{\alpha_{i,m} | i < k_m\}$ is a *prefix-free* set and $k_m \in \mathbb{N} \cup \{\infty\}$; recall the notion of *bounded request*, i.e. a set of pairs $W \subseteq \mathbb{N} \times 2^{<\omega}$ such that $\sum \{2^{-x} | \langle x, y \rangle \in W\} \leq 1$. This sum is called the "weight" of W.

Let us consider therefore the *bounded request*

$$W = \{\langle |\alpha_{i,m}| - m + 1, \alpha_{i,m}\rangle | m \in \mathbb{N}, i < k_m\}$$

We can assume w.l.o.g. that $\mu(V_m) \leq 2^{-2m}$ and check that the contribution of V_m to the weight of W is at most 2^{-m-1}. Actually $\sum_{i \leq k_m} 2^{-|\alpha_{i,m}|+m-1} = \mu(V_m) \cdot 2^{m-1} \leq 2^{-m-1}$. The sum of these values, that is, $2^{-1} + 2^{-2} + 2^{-3} + ...$ for each m is equal to 1, i.e. W is a bounded request. Kraft and Chaitin's theorem states that we can effectively build a machine M_d such that $\langle x, \alpha\rangle \in W$ if and only if there is a string τ such that $|\tau| = x$ and $M_d(\tau) = \alpha$, where $x = |\alpha_{i,m}| - m + 1$ and $\alpha = \alpha_{i,m}$ (and therefore $K_{M_d}(\alpha_{i,m}) \leq |\alpha_{i,m}| - m + 1$) for some $m \in \mathbb{N}, i < k_m$. QED

8.4. Incompleteness and randomness

Chaitin's basic idea was to measure the information content of a theory using the notion of algorithmic complexity. This idea turned out to have strong implications in the analysis of the phenomenon of incompleteness. Here we shall analyse in particular the relationship with Gödel's first incompleteness theorem (for the second, see Kritchman and Raz (2010)). Chaitin proved a version of the first incompleteness theorem which says that, among true, but *unprovable* formulas there are all true statements $K(u) > c$ for a certain constant c (for all finite binary strings u). According to Chaitin this constant is somehow a measure of the information content of the theory. This very elegant version of Chaitin's result is attributed in Van Lambalgen III (1987) to Albert Visser and Dick de Jongh. Recall that to define $K(x)$ we used the universal machine U defined on input of the form $0^e 1\sigma$, that simulates ϕ_e on σ. Hence if $\phi_e(e) = n$, then $K(n) \leq K_{\phi_e}(n) + e + 1 \leq 2e + 1$.

Theorem 128. *Let $w \in 2^{<\omega}$. There exist a constant c such that* PA *does not prove any statement of the kind of '$K(w) > c$'.*

Proof. Fix an enumeration of the derivations of PA; let ϕ_e the following partial recursive function:

$\phi_e(m) = n$ if and only if n is that n occurring in the first proof in PA of a sentence of the form "$\phi_m(m) \neq n$".

It is provable that $\phi_e(e)$ is not definite: indeed, if it were defined, for instance $\phi_e(e) = n$, then PA would prove $\phi_e(e) \neq n$ (contradiction, under the hypothesis of soundness). Hence, notice that PA + $\{\phi_e(e) = n\}$ is consistent. But this implies that PA + $\{K(n) \leq 2e + 1\}$ is consistent and therefore PA cannot prove statements of the form $K(n) > 2e + 1$. QED

The minimal c such that PA $\nvdash K(\sigma) > c$ is called by Chaitin "characteristic constant"; according to Chaitin it depends only by the complexity of the axioms and is related to the information content. Hence true statements like the above, expressed in the language of a theory as PA, are unprovable because their information content is *higher than that of the axioms* of the theory itself. This interpretation has been strongly questioned in Van Lambalgen III (1987), Franzen (2005) and in particular Raatikainen (1998) and Raatikainen (2000). For example, we can collapse the characteristics constant to zero, or we can make it arbitrarily large. These constants are dependent on factors quite different from the information content. Raatikainen's criticism, in particular, is that the above results due to Chaitin depend essentially on the *acceptable system of indexes* adopted. Let the standard indexing be $\phi_0, \phi_1, \phi_2....$ An indexing $\psi_0, \psi_1, \psi_2...$ is called acceptable, if there are computable functions f, g such that $\phi_e \simeq \psi_{f(e)}$ and $\psi_e \simeq \phi_{g(e)}$.

The acceptable indexing meet in particular the fixed point theorem, i.e. for each total recursive function f, there is a number e such that $\phi_e \simeq \phi_{f(e)}$. It is precisely through a simple

application of fixed point that we can collapse to zero the characteristic constant: in fact we take a theory T sufficiently strong a let π an acceptable coding of Turing machines; let us define *ad hoc* the following indexing:

$$\pi^n(x) = \begin{cases} 0 & \text{if } x = n \\ x+1 & \text{if } x < n \\ x & x > n \end{cases}$$

Raatikainen uses this equivalent definition of the Kolmogorov complexity:

$$K(\sigma) = min\{e|\phi_e(0) \simeq \sigma\}$$

namely the smallest description of σ on a fixed input 0. On the basis of this new indexing we can define the algorithmic complexity relative to it as:

$$K^n(\sigma) = min\{k|\exists y(\pi^n(y) = k \wedge \phi_y(0) \downarrow = \sigma)\}$$

Now we define a program P_m in this way: on input 0 look for the minimum x such that $\mathsf{Prf}_{\mathsf{T}}(x, \overline{\ulcorner K^n(\sigma) > \overline{0}\urcorner})$, for some σ; if you find it, print σ.

With the help of the fixed point theorem, we can make sure that the code of this program coincides with the parameter n in $K^n(x)$: if $f(n) = m$ is the program code, there will be an e such that $P_{f(e)} \simeq P_e$.

However P_e never halts: indeed it looks for the minimum x such that:

$$\mathsf{Prf}_{\mathsf{T}}(x, \overline{\ulcorner K^e(\sigma) > \overline{0}\urcorner})$$

and if it finds such a minimum, it prints σ. If there was a similar x with a minimal length of proof, then P_e would print σ: ergo $K(\sigma) \leq e$. But $\pi^e(e) = 0$ and therefore $K^e(\sigma) = e$; hence if hypothetically we had proved $K^e(\sigma) > 0$, thanks to the *soundness* we would have a contradiction. Since P_e never halts, there is no proof of $K^e(\sigma) > \overline{0}$, i.e. $c_T = 0$.

With similar arguments Raatikainen shows that c_T can be made arbitrarily large. What is therefore the true source of "characteristic constants"? According to him the value of c_T is actually determined simply by the smallest (by its code) Turing machine which does not halt, but for which this cannot be proved in T.

In defence of Chaitin's argument Ferbus-Zanda and Grigorieff (2014) point out, however, that modelization rarely rules out all pathological cases and it is intended to be used in "reasonable" cases (perfect modelization is illusory). Mathematically more significant is the defence in Calude and Jürgensen (2005): Chaitin's "heuristic principle", i.e. the thesis that the theorems of a theory cannot be significantly more complex than the theory itself, is defensible as long as the measure of complexity is further specified. By changing the measure of complexity, from program-size $K(x)$ to the measure $\delta(x) = K(x) - |x|$, it is proved that for any sound, consistent theory strong enough to formalize arithmetic and for any Gödel numbering of its well-formed formulas, we can compute a bound N such that no sentence x with complexity $\delta(x) > N$ can be proved in the theory and this phenomenon is independent on the choice of the Gödel numbering.

In Chaitin (2007) it is introduced a particular infinite binary sequence (i.e. a real), called Ω, with singular characteristics. With emphasis the autor says that he is able to construct a much more uncomputable real than Turing does. In this section we will attempt to provide an introduction to what is considered one of the most important concepts in Algorithmic Information Theory. It is necessary to recall a few notions first.

Definition 62. *The Turing computable reals are define as follows:*

(a) *A real α is* computable*, iff is the limit of a computable sequence of rationals*[1] *$r_0, r_1, r_2...$ for which there exists a computable function f such that for all n, $|\alpha - r_{f(n)}| < 2^{-n}$ (i.e. has a computable rate of convergence). E.g. the rationals, $\sqrt{2}$, π, e.*

(b) *A real r is called* recursively enumerable*, if can be approximated in an effective way: $r = \lim_n r_n$, for a computable non decreasing sequence of rationals $r_0, r_1, r_2....$ Equivalently, if it is* left-computable *(or* lower-semicomputable*), i.e. if $L(r) = \{q \in \mathbb{Q} | q < r\}$ is computably enumerable*

Recall that any real in $[0, 1]$ con be associated with a binary sequence: if $\varepsilon = \varepsilon_0\varepsilon_1\varepsilon_2...$ is a binary sequence we can associate to it the real:

$$r_\varepsilon = \varepsilon_0 2^{-1} + \varepsilon_1 2^{-2} + \varepsilon_2 2^{-3} + ...$$

So, a real r in the interval $[0, 1]$ will be written as $\sum_i \epsilon_i 2^{-(i+1)}$.

There are other characterizations of computable reals. E.g.

(a) $r \in [0, 1]$ is computable iff $L(r) = \{q \in \mathbb{Q} | q < r\}$ is computable;

(b) Recall that r can be written as $\sum_i \epsilon_i 2^{-(i+1)}$. We can identify a real $r \in [0, 1]$ with the set A such that the $n - th$ bit of the binary expansion $\epsilon_0, \epsilon_1\epsilon_2....$ of r is 1, iff $n \in A$ (we write $r = 0.A$ to mean that $r = \sum_{i \in A} 2^{-(i+1)}$); then we can say that r is computable iff A is computable. Therefore a real is computable if and only if its binary expansion is computable, i.e. if there is an algorithm to calculate its bits.

Theorem 129. *A real is computable iff it is the characteristic function of some computable set A.*

This notion can be relativised. Hence $r = 0.A$ is computable in $\emptyset'$ (hence is in Δ_2, by Post's results) iff A is computable in $\emptyset'$, iff the left-cut $L(r)$ is computable in $\emptyset'$. Lastly, real is computable if and only if its binary expansion is computable, but a real may be computably enumerable even if its binary expansion *is not* computably enumerable Actually we must distinguish the weaker property of being an computably enumerable real from the stronger one of being a real whose binary expansion is computably enumerable:

(a) A real r is left-recursively enumerable, iff $L(r) = \{q \in \mathbb{Q} | q < r\}$ is computably enumerable

(b) However a real left-computably enumerable *is not* one of the form $0.A$ where A is r.e, namely where $A(0), A(1), A(2), ...$ is the "characteristic function" of a set computably enumerable

(c) We call *strongly computably enumerable* the reals *left computably enumerable* in which this happens, namely whose binary expansion is the" characteristic function " of a set computably enumerable (in the sense that for some e, $\phi_e(n) \downarrow$ iff $A(n) = 1$).

Only the converse is true: if r is strongly computably enumerable then is also left-computably enumerable.

We therefore arrive at a central result.

Theorem 130. *There are reals left computably enumerable that are not strongly computably enumerable.*

[1] A sequence $\{r_i\}_i$ of rationals is computable iff there are computable functions f, g such that $r_i = \frac{f(i)}{g(i)}$.

Typical example of this kind of reals are the constants Ω_U introduced by Chaitin. There are real left-computably enumerable random; they are all and only those of the form Ω_U, for some universal machine U. The number Ω_U depends on U as the halting-set K depends on the enumeration of partial recursive functions; for a theorem due to Myhill however, all K are equivalent, up to a computable permutation. Similarly, all versions of Ω have similar properties that in this phase allow us sometimes to leaving aside from the particular machine. If for example we define:

$$\Omega_s = \sum\{2^{-|\sigma|} | U(\sigma) \; converges \; in \; at \; most \; s - steps\}$$

then Chaitin's Ω is the limit of this non-decreasing computable sequence of rational $\Omega = \lim_s \Omega_s$ (and is therefore computably enumerable in the first sense), however, its binary representation is *random* (and therefore not computably enumerable). An incompressible binary sequence σ, *cannot be* computably enumerable, if for computably enumerable we mean in the second sense, namely that there is a number e such that $\phi_e(i) \downarrow$ if and only if $\sigma(i) = 1$ and where $\sigma(i)$ is the $i - th$ bit of σ. This is the content of the following result.

Lemma 38. (Barzdins 1968) *If σ is computably enumerable in the strong sense, then $K(\sigma \upharpoonright n) \leq 2|n| + e$.*

Proof. For printing $\sigma \upharpoonright n$ we must know: 1) the length n (coded by $\log(n)$-bits), 2) the number k of elements of W_e less than n (coded also with at most $\log(n)$-bits) and 3) the index e. If we provide these informations we consider these steps of computation: generate W_e, until k elements have been enumerated. Then observe that the input $\langle n, k \rangle$ is coded by at most $2 \cdot \log(n) + c$-bits. QED

As we have said, this is the case of numbers Ω_U, whose binary expansion is incompressible. These numbers may be seen as expressing the probability that a universal Chaitin machine U halts when it receives as input a binary string, determined for example through the launch of a coin. Consider that the probability of getting the program u by launching a coin is $2^{-|u|}$, and therefore the probability of fish a program u such that $U(u) = n$ is:

$$\sum\{2^{-|u|} | U(u) = n\}$$

From this comes the definition of the probability that a universal machine U sooner or later halts, on programs selected by lot:

$$\Omega_U = \sum\{2^{-|p|} | U(p) \downarrow\}$$

The Ω_U are real numbers with very peculiar properties. Recall that they depend on the choice of U and then there are, not just one, but a class. In Kučera and Slaman (2001) is proved that each *random left computably enumerable* real coincides with some Ω_U. To summarise, the Ω_U also fulfils these properties:

(a) in terms of binary expansion, they are "chaotic", and therefore incompressible and a fortiori not computably enumerable

(b) From Kraft's inequality (Theorem 126) we have $\Omega_U \leq 1$. Since U does not converge on all strings, actually we have $\Omega_U < 1$. Since on the other hand converges on *some* string, also we have $0 < \Omega_U$; hence, the Ω_U These are reals irrationals strictly between 0 and 1.

(c) Each Ω_U is computable from $\emptyset'$; hence it is Δ_2^0 and by the "Limit lemma" on p.60 is computable in the limit: $\Omega_U(x) = \lim_s f(x, s)$ for some computable functions f with values 0-1. In other words, it is generated by a procedure of type "trial and error" (Putnam).

Let us fix now a universal machine U, namely the U defined as $U(0^e1\sigma) \simeq \phi_e(\sigma)$, for which we have $K_U(\sigma) \leq K_M(\sigma) + c_M$, for all other machines M and let $\Omega = \Omega_U$. Before proving some important properties of Ω, recall that for the binary expansions of real numbers $\alpha = \sum_i \varepsilon_i 2^{-(i+1)}$ where $0 < \alpha \leq 1$ and in $\varepsilon_0\varepsilon_1\varepsilon_2...$, for all i, $\varepsilon_i \in \{0, 1\}$ the following holds:

$$\underbrace{\sum_{i<n} \varepsilon_i 2^{-(i+1)}}_{\alpha\restriction n} < \alpha \leq \underbrace{\sum_{i<n} \varepsilon_i 2^{-(i+1)}}_{\alpha\restriction n} + \underbrace{\sum_{i\geq n} \varepsilon_i 2^{-(i+1)}}_{2^{-n}}$$

where $\alpha \restriction n = \alpha(0), \alpha(1), ..., \alpha(n-1)$.

We remark that the real numbers in $[0, 1]$ that have in common with α the first n elements of their dyadic expansion (the "cylinder" of $\alpha \restriction n$) are in this interval. Hence these additional inequalities hold:

(a) $\Omega \restriction n < \Omega$

(b) $\Omega \leq \Omega \restriction n + 2^{-n}$

If $\Omega_s = \{\sum 2^{-|\sigma|} | U(\sigma) \downarrow \ \textit{in at most } s - \textit{steps}\}$, we have that if $\Omega_s > \Omega \restriction n$, then for all σ of length shorter than or equal to n, if it did not appear in programs that contribute to the determination of Ω_s, then $U(\sigma) \uparrow$. Suppose on the contrary that this is false; hence σ would add a contribution of the amount $2^{-|\sigma|} \geq 2^{-n}$, from which:

$$\Omega \geq \Omega_s + 2^{-|\sigma|} > \Omega \restriction n + 2^{-n}$$

(against as determined at 2.). These observations allow us to highlight the link between *Chaitin's Ω and Turing's Halting Problem.* Knowledge of Ω permits indeed to *solve* the "Halting Problem": later we will prove that in fact $\Omega =_T K$. Actually the knowledge of the first $n - bits$ of Ω_U would solve the halting problems coded with at most n bits, or decide the sentences that can be codes with at most n-bits. It should first be noted that:

(a) if the length of a program is n, its contribution to determine Ω_U is 2^{-n};

(b) also applies the general condition for which $\Omega_U \leq \Omega \restriction n + 2^{-n}$.

So we begin a systematic check of all programs of any length, until we have found enough programs that halts and that allow us to go beyond $\Omega \restriction n$ (suppose to have reached a approximation $\Omega' > \Omega \restriction n$). Hence, if a string w of length less than or equal to n is not among these, $U(w)$ will never halts: if the machine stopped, accordingly we would that $\Omega_U \geq \Omega' + 2^{-|w|} > \Omega \restriction n + 2^{-n}$, against the above conditions. We are therefore able to decide the halting problem for w and Chaitin spoke of "enormous wisdom concentrated in a small space".

Theorem 131. *Ω_U is a random real.*

Proof. Let $\sigma_0, \sigma_1\, \sigma_2....$ an enumeration of the domain of U. Add $2^{-|\sigma_i|}$ to Ω as σ_i is enumerated and let:

$$\Omega_s = \sum_{U(\sigma)\downarrow \ in \leq s \ steps} 2^{-|\sigma|}$$

Observe that there are 2^n strings of length n and therefore at most a finite number of possible changes in passing from the initial segment $\Omega_s \restriction n$ to the segment $\Omega_{s+1} \restriction n$ (if there were more than 2^n possible changes, would mean that the number decreases or exceeds 1, but both things are impossible, since $\Omega_s \leq \Omega_{s+1}$).

Hence there will be a stage k at which it will become stable $\Omega_k \upharpoonright n = \Omega \upharpoonright n$. If you knew what $\Omega \upharpoonright n$ is, we would be able to determine $k = \psi(\Omega \upharpoonright n)$; namely, we could express $\Omega \upharpoonright n$ in this way:

$$\Omega \upharpoonright n = (\sum_{i \leq \psi(\Omega \upharpoonright n)} 2^{-|\sigma_i|}) \upharpoonright n$$

for some partial recursive function ψ, where $\sigma_0, \sigma_1, \sigma_2 ...$ is an enumeration of $Dom(U)$. Observe that at stage $\psi(\Omega \upharpoonright n)$ will be, however, already appeared all strings σ of length n such that $U(\sigma)\downarrow$. Suppose that this is not true: then if another appeared after, we would have $\Omega \upharpoonright n + 2^{-n} < \Omega$, being an initial segment of Ω (against the inequality at point 2. above mentioned). But by definition for all $\tau \in 2^{<\omega}$ such that $K(\tau) \leq n$, there is a σ of length less or equal to n such that $U(\sigma) \simeq \tau$. Hence $\sigma = \sigma_i$, for some $i \leq \psi(\Omega \upharpoonright n)$. In other words, if $\tau \notin \{U(\sigma_i) | i \leq \psi(\Omega \upharpoonright n)\}$, then $K(\tau) > n$. If now F is a function that on input $(\Omega \upharpoonright n)$ select a $\tau \notin \{U(\sigma_i) | i \leq \psi(\Omega \upharpoonright n)\}$, note that $K(F(\Omega \upharpoonright n)) > n$. However we have mentioned that for all $\tau \in 2^{<\omega}$, $K(F(\tau)) \leq K(\tau) + c$, from which $K(\Omega \upharpoonright n) > n + c$ for all n, namely, Ω is incompressible. QED

Theorem 132. $\emptyset' \equiv_T \Omega$.

Proof. First show that $\emptyset' \leq_T \Omega$. Let us take a machine $\mathcal{M}(0^n) = s$ if and only if $\phi_{n,s}(n)\downarrow$ and let e the code of this machine. We decide whether $n \in K$ by means of an oracle in Ω. We remark preliminarily that $U(0^e 10^n) \simeq \phi_e(0^n)$ and that $\phi_n(n)\downarrow$ if and only if for some s, $U(0^e 10^n) \simeq s$. If we have an oracle in Ω to establish its first n bits, then we are able to find s^* big enough to have $\Omega_{\upharpoonright n+e+1} = \Omega_{s^* \upharpoonright n+e+1}$. If $\phi_n(n)\downarrow$, this must have happened before the stadium s^*, otherwise we would have that for some $t \geq s^*$, $U(0^e 10^n)\downarrow = t$ and $0^e 10^n$ would contribute to the determination of Ω with $2^{-|0^e 10^n|}$ and therefore $\Omega_{\upharpoonright n+e+1} + 2^{-|0^e 10^n|} < \Omega$ (contradiction, against what we have seen above). Hence $n \in \emptyset'$ if and only if it was added before the stage s^*.

For the other way round, recall that α is left computably enumerable iff α is the limit of a non decreasing sequence of rationals. Since Ω fulfils the second property, it is therefore left-computably enumerable, and therefore is computable from the halting set K. QED

The author's *interpretation* of this undeniably interesting result has also been the object of criticism in Raatikainen (2000). Chaitin says:

> This is an impenetrable stone wall, it's a worst case. From Gödel we knew that we couldn't get a formal axiomatic system to be complete. We knew we were in trouble, and Turing showed us how basic it was, but Ω is an extreme case where reasoning fails completely (Chaitin (2007), p. 93).

But in what sense, then Ω would be an extreme example of unsolvability, if it is at the level of $\mathsf{0}'$ namely at the level of the halting set K in the upper semilattice of Turing degrees? Actually there are sets rather simple as $X = \{x | W_x \; infinite\}$ and $Y = \{x | W_x \; finite\}$ that are respectively $\Pi^0_2 - complete$ and $\Sigma^0_2 - complete$ (recall that Z is $\Sigma^0_n - complete$, if it is Σ^0_n and for any other A that is Σ^0_n, $A \leq_m Z$). Sets X and Y are therefore more difficult to calculate than Ω. Moreover, it is well known that $\mathsf{PA} + Th_{\Pi_1}(\mathbb{N})$ decides the halting problem, and then also Ω. But it is not able to decide the Paris-Harrington sentence of the previous chapter. So in what sense - Raatikainen asks - is the undecidability of Chaitin's constant *extreme*?

In Chaitin (2007) the Argentine mathematician obtains a further result of incompleteness establishing essentially that if T is a theory capable of interpreting PA, then we can explicitly compute a bound on the number of bits of Ω that can be determined by T. Of the variations on this result obtained subsequently, we believe the clearest and most elegant is the following one, due to Solovay (2000).

Theorem 133. *Let* T *be a theory sufficiently strong and* Σ_1^0 *– sound. It is possible to effectively build a universal machine* U *(provably in* PA*) such that* T *cannot even prove a single statement of the form:"the* n*-th bit of* Ω_U *is* k*".*

Proof. Preliminarly it should be noted that if the theory is Σ_1^0 sound, it is true also each Π_2^0-statement provable in it. Indeed, suppose by contradiction that $\forall x \exists y \phi(x, y)$ was provable but false. Hence for some n it will be true $\neg \exists y \phi(n, y)$. At the same time, however, will be provable also $\exists y \phi(n, y)$, against the Σ_1^0-soundness. A careful formalization shows that the statement: "the $n - th$ bit in Ω_U is k" has complexity Π_2^0; so if it can be proved in a theory Σ_1^0-sound, then is also true. We agree to start counting from zero. Therefore, the first element of a sequence will be the 0^{th} bits etc. Hence the proof consists of three steps:

(Step 1) fix a universal Turing machine V (provably in PA) and define $U(\sigma)$ by cases:

(a) if $\sigma = \emptyset$, then $U(\sigma) \uparrow$;
(b) if $\sigma = 0\tau$, let $U(\sigma) = V(\tau)$;
(c) if $\sigma = 1\tau$, then go to the second step:

(Step 2) Define first an algorithm $\psi(e, \sigma)$ as follows:

(i) preliminarly determine a pair $\langle n, k \rangle$ by means of the following computation: list the theorems of T until a theorem of the form "the n-th bit of Ω_{ϕ_e} is k" appears. If it appears, fix n, k (if the computation does not converge, let $U(\sigma) \uparrow$, for all σ that falls under the case (c)).
(ii) Fixed n, k and recalling that $\sigma = 1\tau$, if $|\tau| \neq n$, let $U(1\tau) \uparrow$; if instead $|\tau| = n$, then $U(\sigma)$ will be defined as follows: let r the diadic rational[2] whose binary expansion is $\tau k 000000....$ and let $r' = r + 2^{-(n+1)}$; look for the minimum s such that $\Omega_{\phi_{e,s}} \in (r, r')$ (clearly this search may fail). Then verify if $\phi_{e,s}(\sigma) \downarrow$. If yes, let $U(\sigma) \uparrow$, otherwise $U(\sigma) = \emptyset$.

The following applies:

i. The universality (provable) of U, follows from the definition of U on strings that begin with 0, from that of V.
ii. The domain of U is prefix-free: Suppose fact that $\sigma_1, \sigma_2 \in Dom(U)$ and $\sigma_1 \subseteq \sigma_2$; as by definition U is not defined on the empty string, then σ_1, σ_2 will be of length bigger than 0. Hence they will be of the form $\sigma_i = r * \tau_i$, where $\tau_1 \subseteq \tau_2$:
   A. if $r = 0$, then $\tau_1, \tau_2 \in Dom(V)$, that is prefix-free and therefore $\sigma_1 = \sigma_2$.
   B. if $r = 1$, than by definition $U(1 * \tau_i)$ is defined if and only if $|\tau_i| = n$. But therefore $|\sigma_1| = |\sigma_2|$ and then $\sigma_1 = \sigma_2$.

(Step 3) We can think of this algorithm through which we define U, starting from ϕ_e and σ, as a function $\psi(e, \sigma)$ and then apply to it the fixed point theorem. This, together with the parameterization theorem allow to deduce that there is an e such that $\psi(e, \sigma) \simeq \phi_e(\sigma)$. Let therefore this $\phi_e = U$ just that involved in the previous step.

[2] Recall that a rational of the form $r \cdot 2^{-n}$ (where r is an integer and n a natural number) is called "dyadic"; each dyadic rational has two binary expansions, and one of two has a tail of the form 0000.... Recall also that a dyadic interval is of the form:

$$[\frac{m}{2^s}, \frac{m+1}{2^s}]$$

There is a correspondence between cylinders $[\sigma]$ and dyadic intervals: in the above, if $|\sigma| = s$, then $m = \sum_{i=1}^{s} 2^{s-1} \cdot \sigma(i)$. Notice that if σ is a finite binary string then the leftmost element (thinking to the binary tree) in the cylinder generated by it is $\sigma 0000000...$, and the rightmost is $\sigma 111111.....$ Indeed, $m + 1 \cdot 2^{-s} = m \cdot 2^{-s} + 2^{-s} = m \cdot 2^{-s} + \sum_{i>s} 2^{-i}$, namely σ followed by an infinite tail of only digits 1 (see e.g. Nies (2009) 12-3.

We check that the hypothesis that the theory can determine some bits of

$$\Omega_U = \Omega_{\phi_e}$$

leads to a contradiction: suppose in fact that we have obtained n, k according to the described procedure; consider (r, r') where $r = h/2^{n+1}$ and $r' = r + 2^{-(n+1)} = h + 1/2^{n+1}$, such that:

$$\Omega_{\phi_e} \in (\frac{h}{2^{n+1}}, \frac{h+1}{2^{n+1}}) = (r, r')$$

Since the theory is *sound*, if it says that the $n - th$ bit of Ω_{ϕ_e} is k, then this must be true. So the binary expansion of r must be of the form $\tau k0000.....$, where $|\tau| = n$. For m big enough we will have that $\Omega_{\phi_e,m} \in (r, r')$; hence consider that (r, r') is the pair required by the computation at point 2.ii and the search for an m such that $\Omega_{\phi_e,m} \in (r, r')$ is successful. Let $\sigma = 1\tau$. We wonder therefore, if $\phi_{e,m}(\sigma)$ converges and the answer is no, because otherwise by definition $U(\sigma) \uparrow$; but since $W_{e,m} \subseteq W_e$, if $\phi_{e,m}(\sigma)$ converged, we would have $\sigma \in W_e$. On the other hand $W_e = Dom(U)$ and therefore $U(\sigma) \downarrow$, contradiction. Hence $\phi_{e,m}(\sigma)$ does not converge and again by the definition in 2.ii $U(\sigma) \downarrow = \emptyset$; hence there is a σ in the domain of U such that $\sigma \notin Dom(\phi_{e,m})$ and $|\sigma| = n + 1$, from which $\Omega_U \geq r + 2^{-(n+1)}$, against the definition of r and the fact that $\Omega_U \in (r, r')$ (contradiction).

QED

The author summarises the relevance he attaches to his findings as follows:

> My work is a fundamental extension of the work of Gödel and Turing on undecidability in pure mathematics. I show that not only does undecidability occur, but in fact sometimes there is complete randomness, and mathematical truth becomes a perfect coin toss (Chaitin (2003), pp. 109-110).

However also in this case the received view is not free from criticism. Recall that Ω is Δ_2 definable, hence computable by a *trial-and-error* machine. For this reason Raatikainen (2000) pp. 221-222 points out that Ω can still be generated by a completely deterministic procedure. Actually the Finnish logician seems to question the plausibility of the theory of algorithmic randomness itself, although, as we have seen, these types of deterministic machines are not equivalent to Turing machines. That is, we are in the domain of hypercomputation, i.e. of hypothetical and unrealistic models of computation that transcend the Church-Turing thesis.

8.5. Farewell: randomness, incompleteness and physical theories

Although the definition of "randomness" as *avoidance of null sets* proposed by Martin-Löf and its equivalents to some extent constitute the standard notion, nevertheless other definitions of randomness have also been proposed. For instance in Schnorr (1971), it is criticised that Martin-Löf randomness was too strong, and two weaker randomness notions are introduced. Martin-Löf-randomness is conversely considered by some to be too weak a notion of randomness. The notion proposed in Demuth (1982), an expression of the Russian school of constructive mathematic, is for instance stronger than Martin-Löf-randomness. But sticking to the standard definition, are all random sequences, so to speak, "equally random" and equally powerful? It may be perplexing that a random sequence is strong enough to compute the Halting Set $\emptyset'$, as is the case with Ω (it is said that Ω can be seen as a highly compressed version of $\emptyset'$ and this seems to be unintuitive). The Gács and Kǔcera theorem 120 also returns computability-theoretically powerful random sequence. However, in Stephan (2006) it is proved that there are (only) two types of random sequences: those that compute $\emptyset'$ and those that are computationally much weaker, which fail to compute a complete extension of PA, that are less powerful than $\emptyset'$ and for somebody these are more intuitively "random". It has

been proved that almost all random sequences are indeed in this second class (see Downey and Hirschfeldt (2019) for an *excursus*).

The theory of randomness we have seen is founded on computability theory. We also pointed out that doubts have been raised as to whether algorithmic randomness is the best account of the notion of randomness. It may indeed be surprising that a deterministic algorithm is involved in the definition: are there alternative proposals? Interestingly, the concept of "random sequence" or "random real" has also been studied without resorting to the notion of an algorithm. In particular, Solovay (1970), by using sophisticated methods and results from *Descriptive Set Theory*, introduced a notion of *real random number* based on set theory, with a peculiar extension of the technique of forcing (see Jech (1978) pp. 493-563 for a detailed account and Durand, et al. (2003) and Ferbus-Zanda and Grigorieff (2014) for a discussion and for some development of non-algorithmically-based randomness approaches and Lafitte (2002) for links with strong axioms of infinity). Remaining within the sphere of *Set Theory*, Kreisel (1969) proposed the programme of expanding the theory with new primitive concepts and new predicates in the language, in addition to set membership, in order to seek solutions to classical problems that were undecidable in standard axiomatisations: among these new predicates to be added, the primitive predicate of being a random sequence. The idea therefore is to abandon the project of giving an explicit characterisation of the concept of 'random sequence', and instead move towards an axiomatisation of randomness. A first attempt at axiomatisation goes back to Myhill in 1963. The axiomatic approach has been systematically pursued in Van Lambalgen (1990) and Van Lambalgen (1992) by adopting Kreisel and Myhill's observations about the substantially intensional character of the notion of 'random sequence', or at least not fully extensional, starting from the consideration that none of the current formalizations captures its meaning exhaustively.

The ZFR theory, obtained from the set theory ZF (i.e. Zermelo-Fraenkel set theory without the *Axiom of Choice*) by adding the Van Lambalgen axioms for the primitive predicate $R(x, y)$= "x is independent of y", i.e. "y has no information for x" , or "x is random with respect to y" (where $R(x) = R(x, \emptyset)$), turned out to be for example incompatible with the *Axiom of Choice*. The Dutch logician also proposed another axiomatization, in terms of generalized quantifiers of the type $Qx\phi(x)$= "if x is generated randomly, then it is practically certain that $\phi(x)$". The theory ZFQ , obtained from ZFC (i.e. ZF plus the *Axiom of Choice*) by adding these axioms is conservative on ZFC. Actually the theories ZFQ and ZFR turn out to be relatively interpretable one into the other. A particular fragment of the axiomatization of the theory ZFQ significantly allows to decide the continuum hypothesis $\aleph_1 = 2^{\aleph_0}$, proving the so-called "Freeling's Axiom", equivalent in ZFC to its denial. It must be noted, however, that the program of introducing new axioms for new primitive concepts, such as that of randomness, rather than axioms (such as those on large cardinals or on forcing) formulated in the pure language of set theory with the concepts of set and membership, has not found full approval and has even found fierce opponents.

There is furthermore the unavoidable issue of the relationship between algorithmic randomness and randomness in the sense of the physical sciences, either as *unpredictability* (deterministic chaos) or as ontic randomness, according to the standard interpretation of Quantum Mechanics. Gödel's aversion to generalisations of the significate of his results beyond the specifically logico-mathematical terrain is well known. There is in this respect the famous episode of John Wheeler who, having gone to Gödel's office to ask whether there was a connection between Heisenberg's uncertainty principle and the incompleteness theorem, was rather rudely kicked out (Wheeler comments with irony that this was mainly due to the bad influx exerted by Einstein on Gödel). However, the suggestive analogy between mathematical incompleteness and certain natural phenomena has been an irresistible attraction for many physicists. Incompleteness in classical, as well as quantum, physics became a popular topic for some years (see e.g. Da Costa and Doria (1991) or Moore (1990)). On the side of deterministic theories, dynamical systems are called *chaotic* if they are sensitive to initial conditions. In scientific thought between the 18th and 19th centuries, the deterministic

character of theories was revealed in their predictive capacity. Since Poincaré, however, a clearer distinction has begun to be made between the concepts of *determination* and of *prediction*: indeed, in a deterministic context, Poincaré proved that the system of equations describing the the interaction of three celestial bodies is provably incapable of predicting their evolution. In Bailly and Longo (2008), the authors consider Poincaré's three-body problem, which is at the origin of the theories of deterministic chaos, as a forerunner from an epistemological point of view of Gödel's limiting theorems: if the Laplacian conception can be compared in the logical-mathematical sphere, to the Hilbertian idea of a complete formal system, Poincaré's results can then be equated by analogy with Gödel's. A parallel is therefore instinctively established between the paradigmatic positions of Hilbert and Laplace, on the one hand, and Gödel and Poincaré on the other, that is, on the one hand the demand for decidability of every statement concerning the future (Laplacean predictability), while on the other, Poincaré, with his theorem on three-body problem, showed the "unpredictability of interesting non-linear systems, which we may understand as undecidability of future states".

In Bailly and Longo (2007) the authors wonder whether there is a randomness specific to the various fields of Physics. Typically in non-linear systems, randomness is of an "epistemic"nature, unpredictability is the joint result of sensitive dependency of the boundary conditions and the theoretical properties of classical measurement. Is it possible to describe the deterministic chaos by means of Chaitin complexity? This problem was raised in the 1990s of the last century and was part of a list of open problems. For a survey of progress in the applications of the theory of algorithmic randomness to the *Ergodic Theory* see for instance V'yugin (2022) and Towsner (2020). On the other hand, another branch, *Quantum Physics* proposes an intrinsic notion of randomness, as it is associated with *any* measurement operation. Challenging this "objective", "ontic", not epistemic concept of randomness, as early as 1935 Einstein, Podolsky and Rosen spoke of the "incompleteness" of *Quantum Mechanics*, formulating the famous EPR paradox. "Completeness" means in this framework that every element of the physical reality must have a counterpart in the physical theory. The paradox forces to conclude that the quantum-mechanical description of physical reality given by wave functions is not complete in this sense. The wave function does not provide in other words an exhaustive description of the objective properties of the system; this as is known led to the formulation of various hypotheses around hidden variables in order to produce a deterministic completion of the theory which would classically lead back to our ignorance its apparent probabilistic character. However, the remarkable results obtained by John Bell, as is well known, establish the impossibility of a local, deterministic theory predictively equivalent to *Quantum Mechanics* QM, i.e. no local theory admitting hidden variables is capable of reproducing its statistical predictions.

Starting from the Einstein, Podolsky and Rosen definition of "reality", the paradox stated in Peres (1985) highlights a case in which we know the (negative) answer to the question of whether or not the spin of a given electron is $3\hbar$, although there is no direct way of posing the question in the QM formalism and this too is reminiscent of the phenomenon of incompleteness. In Peres, Zurek (1982) it is emphasized that the theory of *Quantum Mechanics* cannot be considered "closed", in the sense that it can describe in principle anything, although in every situation something remains unanalyzed:

> This may not be just a flaw of quantum theory: it is likely to emerge as a logical necessity in any theory which is self- referential, as it attempts to describe its own means of verification. In this sense it is analogous to Gödel's undecidability theorem of formal number theory.

However, the reformulation of the incompleteness theorem *as a result about algorithmic randomness* has made it possible to investigate out of metaphor its meaning in fields other than Logic and has allowed a connection with disciplines such as Physics or Information Theory. There has been much speculation also on the relationship between indeterminacy in Heisenberg's sense and incompleteness in Gödel's sense (despite the latter's hostility). Beyond the strong suggestiveness of the topic, this issue has been thoroughly investigated on a

rigorously mathematical level in Calude and Stay (2007), that used for this purpose Chaitin's reformulation of the incompleteness theorem as a *trait-d'union* with Physics, presenting the Gödel-Chaitin theorem as a true uncertainty principle. Algorithmic randomness, as these two mathematicians show, is equivalent to what they call a "formal uncertainty principle", which in turn implies incompleteness.

The research programme that investigates the relationship between algorithmic randomness and quantum randomness is actually, in our opinion, one of the most exciting developments. Yurtsever (2000) showed that a sequence of bits produced by tossing a quantum coin is, almost certainly, algorithmically random (although the proof has been considered by some to be lacking in some points). More recently Nies and Scholz (2017) develop an algorithmic theory of randomness for infinite sequences of quantum bits and introduce a quantum analog of Martin-Löf tests, showing the existence of a universal such test in the framework of the C^*-algebra formulation of Quantum Mechanics. Actually at present there are different and non-equivalent approaches to this problem and we are still in an experimental phase of what appears to be a promising research program. The bibliography we have mentioned around algorithmic randomness and in particular its ralation with the notions of randomness that emerge from the various branches of Physics, is far from being exhaustive. The breadth and depth of the technical literature is enormous and we cannot account for it, nor is it the purpose of this book.

Bibliography

Abramsky, Samson and Tzevelekos, Nikos. 2011. *Introduction to Categories and Categorical Logic*. Oxford University Computing Laboratory. arXiv : 1102.1313.

Ajtai, Miklós. 1994. "The complexity of the pigeonhole principle". *Combinatorica* 14: 417-433.

Amadio, Roberto and Curien, Pierre-Louis. 1996. *Domains and Lambda-Calculi*. Cambridge: Cambridge University Press.

Amidei, Jacopo and Pianigiani, Duccio and San Mauro Luca and Simi, Giulia and Sorbi, Andrea. 2016. "Trial and error Mathematics I" . *The Review of Symbolic Logic* 9(2): 299-324.

Amidei Jacopo and Pianigiani, Duccio and San Mauro Luca and Sorbi, Andrea. 2016. "Trial and error mathematics II: dialectical sets and quasi-dialectical sets, their degrees, and their distribution within the class of limit sets". *The Review of Symbolic Logic* 9(4): 810-835.

Amidei Jacopo and Andrews, Uri and Pianigiani, Duccio and San Mauro Luca and Sorbi, Andrea. 2019. "Trial and error mathematics III: Dialectical systems and completions of theories". *Journal of Logic and Computation* 29 (1): 157-184

Arai, Toshiyasu. 2020. *Ordinal Analysis with an Introduction to Proof Theory*. Berlin: Springer.

Ardeshir, Mohammad and S. Mojtaba Mojtahedi, Mojtaba S. 2014. "The Σ^0_1 Provability Logic of HA". *Annals of Pure and Applied Logic* 169 (10): 997-1043.

Artemov, Sergei and Beklemishev, Lev D. 2005. "Provability Logic". In Gabbay, D., Guenthner, F. (editors) *Handbook of Philosophical Logic*, 2nd Edition. Handbook of Philosophical Logic, vol 13. Dordrecht: Springer.

Artemov, Sergei and G. Japaridze Giorgie. 1990."Finite Kripke models and predicate logics of provability". *The Journal of Symbolic Logic* 55 (3): 1090-1098.

Asperti, Andrea. 1998. "Light affine logic" . In *Proc. 13th IEEE Annual Symp. on Logic in Computer Science* 300-308. Washington: IEEE Computer Society.

Asperti, Andrea and Roversi, Luca. 2002. "Intuitionistic Light Affine Logic" . *ACM Trans. Comput. Logic* 3 (1): 137-175.

Ausiello, Giorgio and Gambosi, Giorgio and d'Amore, Fabrizio. 2002. *Linguaggi, Modelli, Complessità*. Milano: Franco Angeli.

Avigad, Jeremy and Feferman, Solomon. 1998. "Gödel's Functional ("Dialectica") Interpretation". S. R. Buss (editor) *Handbook of Proof Theory* 337-405.
Amsterdam: Elsevier.

Duccio Pianigiani, University of Siena, Italy, duccio.pianigiani@unisi.it, 0000-0001-9441-7226

Referee List (DOI 10.36253/fup_referee_list)
FUP Best Practice in Scholarly Publishing (DOI 10.36253/fup_best_practice)

Book References DOI 10.36253/979-12-215-0778-2.references

Awodey, Steve. 1996. "Structure in mathematics and logic: A categorical perspective" . *Philosophia Mathematica* 4 (3): 209-237.

Awodey, Steve. 2008. *Category Theory* (II edition). Oxford: Oxford University Press.

Baillot, Patrick. 2004. "Type inference for light affine logic via constraints on words" . *Theoretical Computer Science* 328 (3): 289-323.

Baillot, Patrick and Mogbil, Virgile. 2004. "Soft lambda-Calculus: A Language for Polynomial Time Computation" . In I. Walukiewicz (editors) *Foundations of Software Science and Computation Structures 27-41. FoSSaCS 2004. Lecture Notes in Computer Science 2987.* Berlin: Springer.

Bailly, Francis and Longo, Giuseppe. 2007. "Randomness and determinism in the interplay between the continuum and the discrete" . *Mathematical Structures in Computer Science* 17: 289-305.

Bailly, Francis and Longo, Giuseppe. 2008. "Phenomenology of Incompleteness: from Formal Deductions to Mathematics and Physics" . In R. Lupacchini (editor) *Deduction, Computation, Experiment* Berlin: Springer.

Barendregt, Henk P. 1984. *The Lambda Calculus: Its Syntax and Semantics*, second, revised edition. Amsterdam: North-Holland.

Barendregt, Henk P. 1992. "Representing undefined in Lambda calculus" . *Journal of Functional Programming* 2 (3): 367-374.

Barmpalias, George and Lewis-Pye, Andrew. 2018. "Optimal redundancy in computations from random oracles" . *Journal of Computer and System Sciences* 92: 1-8.

Barwise, John and Jerome Keisler ed. 1989. *Handbook of Mathematical Logic.* Amsterdam: North Holland.

Barzdins, Janis. 1968. "Complexity of programs to determine whether natural numbers not greater than n belong to a recursively enumerable set". *Soviet Math. Dokl.* 9: 1251-54.

Beame, Paul and Impagliazzo, Russell and and Pitassi, Toniann. 1993. "Exponential lower bounds for the pigeonhole principle" . *Computational Complexity* 3: 97-140.

Beckmann, Arnold and Buss, Samuel R.2011. "Corrected upper bounds for free-cut elimination". *Theoretical Computer Science* 412 (39): 5433-5445.

Beckmann, Arnold and Pollett, Chris and Buss, Samuel R. 2003. "Ordinal notations and well-orderings in bounded arithmetic". *Annals of Pure and Applied Logic* 120, (1–3): 197-223,

Beklemishev, Lev. 1992. "Independent enumerations of theories and recursive progressions". *Siberian Mathematical Journal* 33: 760-783.

Beklemishev, Lev. 1995. "Iterated local reflection versus iterated consistency". *Annals of Pure and Applied Logic* 75 (1-2): 25-48.

Beklemishev, Lev. 2006. "The Worm principle". In Z. Chatzidakis, P. Koepke, W. Pohlers (editors) *Logic Colloquium '02* 75-95. Cambridge: Cambridge University Press.

Beklemishev, Lev. 2014. "Positive provability logic for uniform reflection principles". *Annals of Pure and Applied Logic*, 165 (1) : 82-105

Beklemishev, Lev and Flaminio, Tommaso. 2016. "Franco Montagna's Work on Provability Logic and Many-valued Logic". *Studia Logica* 104: 1-46.

Bellantoni, Stephen and Cook, Stephen. 1992. "A new recursion-theoretic characterization of the polytime functions". *Computational Complexity* 2: 97-110.

Berarducci, Alessandro. 1990. "The interpretability logic of Peano Arithmetic" . *The Journal of Symbolic Logic* 55: 1059-1089.

Berarducci, Alessandro and D'Aquino, Paola. 1995."Δ_0-complexity of the relation $y = \Pi_{i \leq nF(i)}$" . *Annals of Pure and Applied Logic* 75 (1-2): 49-56.

Berarducci, Alessandro and Mamino, Marcello. 2023. "Provability logic: models within models in Peano Arithmetic" . *Bollettino dell'Unione Mat. Ital.* 16: 25-41.

Berarducci, Alessandro and Otero, Margarita. 1996. "A Recursive Nonstandard Model of Normal Open Induction" . *The Journal of Symbolic Logic* 61 (4): 1228-41.

Bernardi, Claudio. 1975. "The fixed-point theorem for diagonalizable algebras" . *Studia Logica* 34 (3): 239-251.

Bernardi, Claudio. 1981. "On the relation provable equivalence and on partitions in effectively inseparable sets" . *Studia Logica* 40: 29-37.

Bernardi, Claudio and Sorbi, Andrea. 1983. "Classifying positive equivalence relations" . *The Journal of Symbolic Logic* 48 (3): 529-538.

Bernays, Paul. Letter to Gödel, 7 January 1970. Bernays Papers, ETH Zürich Library/WHS, Hs. 975:1745.

Bezboruah, Amala and Shepherdson, John C. 1976. "Gödel's Second incompleteness theorem for Q" . *The Journal of Symbolic Logic* 41(2): 503-512.

Bishop, Erret. 1970. "Mathematics as a numerical language" In A. Kino, J. Myhill and R. E. Vesley (editors) *Intuitionism and Proof Theory* 53-71. Amsterdam: North-Holland.

Boolos, George S. 2008. *The Logic of Provability*. Cambridge: Cambridge University Press.

Boolos, George and V. McGee, Van. 1987. "The degree of the set of sentences of predicate provability logic that are true under every interpretation". *The Journal of Symbolic Logic* 52: 165-171.

Bourbaki. 1950. "The Architecture of Mathematics". *The American Mathematical Monthly* 57 (4): 221-232.

Borel, Émile. 1909. "Les probabilités dénombrables et leurs applications arithmétiques" . *Rendiconti del Circolo Matematico di Palermo* 27: 247-271.

Bovykin, Andrey and Kaye, Richard. 2002. "Order-types of models of Peano arithmetic" . In Y. Zhang (edior) *Logic and Algebra. Contemporary Mathematics 302* 275-285. Providence: American Mathematical Society.

Boykan Pour-El, Marian. 1968. "Effectively extensible theories". *The Journal of Symbolic Logic* 33 (1): 56-68.

Brouwer, Luitzen Egbertus Jan. 1912. "Intuitionism and formalism."*Bull. Amer. Math. Soc.* 20: 81-96, (reprinted in *Philosophy of Mathematics: Selected Readings* 1984, edited by Paul Benacerraf and Hilary Putnam, 77-89. Cambridge: Cambridge University Press.

Buchholz, Wilfried. 1987. "An independence result for $(\Pi^1_1 - CA) + BI$" . *Annals of Pure and Appied Logic* 33: 131-155.

Buchholz, Wilfried and Wainer, Stan. 1987. "Provably computable functions and the fast growing hierarchy". In S. Simpson (editor) *Logic and Combinatorics. Contemp. Math* 65 179-98. Providence: American Mathematical Society.

Buss, Samuel. 1985. "The polynomial hierarchy and intuitionistic Bounded Arithmetic" . *Structure in Complexity Theory*: 77-103.

Buss, Samuel R. 1986. *Bounded Arithmetic*. Napoli: Bibliopolis.

Buss, Samuel R. 1991. "The Undecidability of k-Provability" . *Annals of Pure and Applied Logic* 53: 75-102.

Buss, Samuel R. 1995. "Relating the bounded arithmetic and polynomial-time hierarchies" . *Annals of Pure and Applied Logic* 75: 67-77.

Buss, Samuel R. 1995. "On Gödel's Theorems on Lengths of Proofs II: Lower Bounds for Recognizing k Symbol Provability Bounded Arithmetic and Propositional Proof Complexity" . In P. Clote and J. Remmel (editors) *Feasible Mathematics II* 57-90. Basel: Birkhauser.

Buss, Samuel R. 1997. "Bounded Arithmetic and Propositional Proof Complexity". In *Logic of Computation, NATO ASI Series* 157, edited by H. Schwichtenberg, 67-121. Berlin: Springer.

Buss, Samuel R. 1998. "An introduction to proof theory". In *Handbook of proof theory*, 1-78. Amsterdam: Elsevier.

Buss, Samuel R. 1998. "First-order proof theory of arithmetic". In S. Buss (editor) *Handbook of Proof Theory* 79-147. Amsterdam: Elsevier.

Buss, Samuel R. 1999. "Bounded arithmetic, proof complexity and two papers of Parikh" . *Annals of Pure and Applied Logic* 96 (1-3): 43-55.

Buss, Samuel R. and Krajícek, Jan. 1994. "An Application of Boolean Complexity to Separation Problems in Bounded Arithmetic" . *Proceedings of The London Mathematical Society*: 1-21.

Buss, Samuel R. and Krajíček, Jan and Takeuti, Gaisi. 1993. "Provably total functions in the bounded arithmetic theories R^i_3, U^i_2, and $\mathsf{V}^{\mathsf{i}}_2$" . In *Proof Theory, Arithmetic, and Complexity*, edited by P. Clote and J. Krajíček, 116-161. Oxford: Oxford University Press.

Buss, Samuel R. and Beckmann, Arnold. 2014. "Improved witnessing and local improvement principles for second-order bounded arithmetic". *ACM Trans. Comput. Logic* 15 (1): 1-35.

Calude, Claude. 1994. "Borel normality and algorithmic randomness" . In G. Rozenberg, A. Salomaa (editors) *Developments in Language Theory* 113-129. Singapore: World Scientific.

Calude, Cristian S. and Jürgensen, Helmut. 2005. "Is complexity a source of incompleteness?". *Advances in Applied Mathematics* 35 (1): 1-15.

Calude, Cristian ed. 2007. *Randomness and Complexity, from Leibniz to Chaitin*. Singapore: World Scientific.

Calude, Christian S. and Stay, Michael. 2007. "From Heisenberg to Gödel via Chaitin". *International Journal of Theoretical Physics* 46 (8): 1053-1065.

Cantini, Andrea. 1996. "Asymmetric Interpretations for Bounded Theories" . *Mathematical Logic Quarterly* 42: 270-288.

Carlucci, Lorenzo. 2003. "A new proof-theoretic proof of the independence of Kirby-Paris' Hydra Theorem." *Theoretical Computer Science* 300: 365-378.

Carlucci, Lorenzo. 2005. "Worms, gaps, and hydras". *Mathematical Logic Quarterly* 51 (4): 342-350.

Chaitin, Gregory J. 1974. "Information-theoretic limitations of formal systems" *Journal of the ACM* 21 (3): 403-424.

Chaitin, Gregory J. 1975. "A Theory of Program Size Formally Identical to Information Theory" . *Journal of the ACM* 22 (3): 329-340.

Chaitin, Gregory. 1977. "Algorithmic Information Theory". *IBM Journal of Research and Development* 21 (4): 50-359.

Chaitin, Gregory. 1982. "Gödel's theorem and information" . *International Journal of Theoret. Physics* 22: 941-954.

Chaitin, Gregory. 1990. *Information, Randomness and Incompleteness, Papers on Algorithmic Information Theory* (second edition). Singapore: World Scientific.

Chaitin, Gregory. 2003. *Conversations with a Mathematician. Math, Art, Science and the Limits of Reason*. Berlin: Springer.

Chaitin, Gregory. 2007. *Thinking of Gödel and Turing*. Singapore: World Scientific.

Chang, Chen Chung and Keisler, H. Jerome. 1978. *Model Theory* (third edition). Amsterdam: North Holland.

Cheng, Yong. 2023. "There Are No Minimal Effectively Inseparable Theories ". *Notre Dame Journal of Formal Logic* 64 (4): 425-439.

Church, Alonso. 1936. "An unsolvable Problem of Elementary Number Theory". *American Journal of Mathematics* 58 (2): 345-363.

Church, Alonzo. 1940. "On the concept of a random sequence" . *Bulletin of the American Mathematical Society* 46: 130-135.

Cichon, Adam. 1983. "A short proof of two recently discovered independence results using recursion theoretic methods". *Proceedings of the American Mathematical Society* 87 (4): 704-706.

Cook, Stephen. 1971. "The complexity of theorem-proving procedures" . In *STOC '71 Proceedings of the third annual ACM symposium on Theory of computing*, 151-158. New York: ACM Association for Computing Machinery.

Cook, Stephen and Nguyen, Phuong. 2010. *Logical Foundations of Proof Complexity* . Cambridge: Cambridge University Press.

Cook, Stephen and Urquhart, Alisdair. 1993. "Functional interpretations of feasibly constructive arithmetic" . *Annals of Pure and Applied Logic* 63 (2): 103-200.

Cooper, S. Barry. 2003. *Computability Theory*. Boca Raton: Chapman and Hall/ CRC Press.

Cooper, Barry S. 2004. "The Incomputable Alan Turing" . In J. Delve and J. Paris, (editors) *Proceedings of the 2004 international conference on Alan Mathison Turing: a celebration of his life and achievements* (Turing'04) 1-17. Swindon: BCS Learning and Development Ltd.

Copeland, B. Jack and Shagrir, Oron. 2013. "Turing versus Gödel on Computability and the Mind" . In J. B. Copeland, C. Posy, O. Shagrir (editors) *Computability: Gödel, Turing, Church, and beyond* 1-35. Boston: MIT Press.

Copeland, Jack. *The Church-Turing Thesis*. alanturing.net

Crosilla, Laura. 2017. "Predicativity and Feferman" . In G. Jäger and W. Sieg (editors) *Feferman on Foundations: Logic, Mathematics and Philosophy* 423-447. Cham: Springer.

Crole, Roy. 1994. *Categories for Types*. Cambridge: Cambridge University Press.

Da Costa, Newton and Doria, Francisco A. 1991. "Undecidability and Incompleteness in Classical Mechanics" . *International Journal of Theoretical Physics* 30: 1041-1073.

D'Aquino, Paola. 1992. "Local behaviour of the Chebyshev theorem in models of $\mathsf{I}\Delta_0$" . *The Journal of Symbolic Logic* 57 (1): 12-27.

D'Aquino, Paola. 1997. "Toward the Limits of the Tennenbaum Phenomenon" . *Notre Dame Journal of Formal Logic* 38 (1):81-92.

D'Aquino, Paola and Macintyre, Angus. 2000. "Non-standard finite fields over $\mathsf{I}\Delta_0 + \Omega_1$" . *Israel Journal of Mathematics* 117 (1): 311-333.

Davis, Martin and Sigal, Ron and Elaine, Jean. 1994. *Computability, complexity and language*. Burlington: Morgan Kaufmann.

Dawson, John. 1997. *Logical Dilemmas. The Life and Work of Kurt Gödel*. Wellesley: A K Peters.

Dedekind, Richard. 1888. *Was sind und was sollen die Zahlen?* 1. Auflage. Braunschweig: Vieweg.

De Jongh, Dick. 1969. "The maximality of the intuitionistic predicate calculus with respect to Heyting's arithmetic."(typed manuscript) Amsterdam University.

De Jongh, Dick, and Japaridze, Giorgi. 1998. "The Logic of Provability" . In S. R. Buss (editor) *Handbook of Proof Theory* 475-546. Amsterdam: Elsevier.

De Jongh, Dick, Marc Jumelet, and Franco Montagna. 1991. "On the Proof of Solovay's Theorem" . *Studia Logica* 50 (1): 51-69.

De Jongh, Dick, and Veltman, Frank. 1990. "Provability logics for relative interpretability" . In *Mathematical Logic*, edited by P. P. Petkov, 175-208. New York: Plenum Press.

De Jongh, Dick, and Verbrugge, Rineke and Visser, Albert. 2011. "Intermediate logics and the de Jongh property". *Archive for Mathematical Logic* 50: 197-213.

Detlefsen, Michael. 1986. *Hilbert's Program*. Dordrecht: Reidel.

Detlefsen, Michael. 1979. "On interpreting Gödel Second Theorem" . *Journal of Philosophical Logic* 8: 297-313.

Detlefsen, Michael. 2001. "What does Gödel's second theorem say?". *Philosophia Mathematica* 9 (1): 37-71.

Demuth, Osvald. 1982. "Some classes of arithmetical real numbers " (in Russian). *Commentationes Mathematicae Universitatis Carolinae*. 23 (3): 453-465.

de Almeida Borges, Ana and Joosten, Joost. 2013. "An escape from Vardanyan's theorem". *The Journal of Symbolic Logic* 88 (4): 1613-1638.

Dieudonné, Jean. 1987. *Pour l'honneur de l'esprit humain*. Paris: Hachette Littérature.

Du, Dingzhu and Ko, Ker-I. 2014. *Theory of Computational Complexity*. 2nd Edition. Hoboken: Wiley.

Downey, Rod and Hirschfeldt, Denis. 2010. *Algorithmic Randomness and Complexity*. Berlin: Springer.

Downey, Rod and Hirschfeldt, Denis. 2019. "Computability and Randomness" *Notices of the American Mathematical Society.* 66: 1001-1012.

Durand, Bruno and Kanovei, Vladimir and Uspensky, Vladimir A. and Vereshchagin, Nikolai. 2003. "Do stronger definitions of randomness exist?" *Theoretical Computer Science* 290 (3): 1987-1996.

Enderton, Herbert B. *A mathematical introduction to logic* (Second edition). Harcourt/Academic Press: San Diego.

Endrullis, Jörg and Klop, Jan and Overbeek, Roy. 2021. "Star Games and Hydras." *Logical Methods in Computer Science* 17 (2):1-32.

Feferman, Solomon. 1960. "Arithmetization of metamathematics in a general setting" . *Fundamenta Mathematicae* 49 (1): 35-92.

Feferman, Solomon. 1962. "Transfinite Recursive Progressions of Axiomatic Theories". The Journal of Symbolic Logic 27 (3): 259-316.

Feferman, Solomon. 1964. "Systems of predicative analysis". *The Journal of Symbolic Logic* 29: 1-30.

Feferman, Solomon. 2006. "Are There Absolutely Unsolvable Problems? Gödel's Dichotomy". *Philosophia Mathematica* (III) 14: 134-152.

Feferman, Solomon. 2006. "Turing's Thesis" . *Notices of the AMS* 53 (10): 2-8.

Ferbus-Zanda, Marie and Grigorieff, Serge. 2008. "Is Randomness Native to Computer Science?" *Bulletin of The European Association for Theoretical Computer Science* 74: 78-118.

Ferbus-Zanda, Marie and Grigorieff, Serge. 2014."Kolmogorov Complexity in Perspective Part I: Information Theory and Randomness". In *Constructivity and Computability in Historical and Philosophical Perspective. Logic, Epistemology, and the Unity of Science* 34, edited by J. Dubucs and M. Bourdeau 57-94. Dordrecht: Springer Netherland.

Ferreira, Fernando and Ferreira, Gilda. 2013. "Interpretability in Robinson's Q". *The Bulletin of Symbolic Logic* 19: 289-317.

Fitting, Melvin and Mendelsohn, Richard. 1998. *First Order Modal Logic*, Dordrecht: Kluwer.

Franzen, Torkel. 2003. *Inexhaustibility: A Non-Exhaustive Treatment*. Lecture Notes in Logic 16. Cambridge: Cambridge University Press .

Franzen, Torkel. 2004. "Transfinite progressions: a second look at completeness". *Bulletin of Symbolic Logic* 10 (2004): 367-389.

Franzen, Torkel. 2005. *Gödel's theorem: an incomplete guide to its use and abuse* (1st ed.). Wellesley: A. K. Peters.

Friedman, Harvey. 1973. "Some applications of Kleene's methods for intuitionistic systems."In Cambridge Summer School in Mathematical Logic, edited by A.R.D. Mathias and A.R.D. and H. Rogers: 113–170. Berlin: Springer.

Gács, Peter. 1986. "Every sequence is reducible to a random one" . *Information and Control* 70 (2–3): 186-192.

Gaifman, Haim and Dimitracopoulos, Costas. 1982. "Fragments of Peano's Arithmetic and the MRDP theorem" . *Monographie de L'Enseignement Mathematique* 30: 187-206.

Gallier, Jean H. 1991. "What's so special about Kruskal's theorem and the ordinal Γ_0? A survey of some results in proof theory" . *Annals of Pure and Applied Logic* 53 (3): 199-260.

Gandy, Robin. 1980. "Church's Thesis and Principles for Mechanisms" . In J. Barwise, H. J. Keisler and K. Kunen (editors) *The Kleene Symposium* 123-148. Amsterdam: North-Holland.

Gentzen, Gerhard. 1935. "Untersuchungen über das logische Schließen" . *Mathematische Zeitschrift* 39: 176-210 and 405-431.

Gentzen, Gerhard. 1936. "Die Widerspruchfreiheit der reinen Zahlentheorie" *Mathematische Annalen* 112: 493-565 (reprinted in M. E. Szabo (editor), *Collected papers of Gerhard Gentzen* 1969. Amsterdam: North Holland 132-213).

Gentzen, Gerhard. 1938. "Neue Fassung des Widerspruchsfreiheitsbeweises für die reine Zahlentheorie" . Forschungen zur Logik und zur Grundlegung der exacten Wissenschaften, Neue Folge 4: 19-44 (reprinted in M. E. Szabo (editor), *Collected papers of Gerhard Gentzen* 1969. Amsterdam: North Holland 252–286).

Gentzen, Gerhard. 1943. "Beweisbarkeit und Unbeweisbarkeit von Anfangsfällen der transfiniten Induktion in der reinen Zahlentheorie" . Mathematische Annalen, 119 (1): 140–161 (reprinted in M. E. Szabo (editor), *Collected papers of Gerhard Gentzen* 1969. Amsterdam: North Holland 287-308).

Ghilardi, Silvio. 1999."Unification in Intuitionistic Logic" . *The Journal of Symbolic Logic* 64: 859-880.

Girard, Jean Yves. 1987. *Proof-theory and logical complexity I.* Napoli: Bibliopolis.

Girard, Jean-Yves and Lafont, Yves and Taylor, Paul. 1989. *Proofs and types.* Cambridge: Cambridge University Press.

Girard, Jean Yves. 1998. "Light linear logic" . *Information and Computation* 143: 175-204.

Gödel, Kurt. 1930. "Die Vollständigkeit der Axiome des logischen Funktionenkalküls". Translated by Stefan Bauer-Mengelberg. In Gödel's Collected Works (1986-2005) 102–23.

Gödel, Kurt. 1931. "Über Formal Unentscheidbare Sätze der Principia Mathematica und Verwandter Systeme I" . *Monatshefte für Mathematik und Physik* 38: 173-198 (reprinted in Gödel's Collected Works (1986-2005), Vol. I: 144-195 and Davis, Sigal and Weyuker (1994): 4-39).

Gödel, Kurt. 1951. "Some basic theorems on the foundations of mathematics and their implications" (Gibbs Lecture). In Gödel's Collected Works (1986-2005) III, 304-323.

Gödel, Kurt. 1986-2005. *Collected Works* I-V, edited by S. Feferman, J. W. Dawson, Stephen C. Kleene, W. Goldfarb, G. Moore, C. Parsons, R. Solovay and Jean Van Heijenoort. Oxford: Oxford University Press.

Gödel, Kurt. 1958. "Über eine bisher noch nicht benützte Erweiterung des finiten Standpunktes" . *Dialectica* 12 (3): 280.

Gödel, Kurt. 1972. "On an Extension of Finitary Mathematics Which Has Not Yet Been Used" . Revised and expanded English version of 1958, first published in Gödel's Collected Works (1986-2005) II: 271-280.

Gold, E. Mark. 1967. "Language identification in the limit". *Information and Control* 10 (5): 447-474.

Goré, Rajeev and Ramanayake, Revantha. 2012. "Valentini's cut-elimination for provability logic resolved" . *The Review of Symbolic* 5 (2): 212-238.

Grattan-Guinness, Ivor. 1979. "In Memoriam Kurt Gödel: his 1931 correspondence with Zermelo on his incompletability theorem" . *Historia Math.* 6: 294-304.

Grattan-Guinness, Ivor. 2011. "The Reception of Gödel's 1931 Incompletability Theorems by Mathematicians, and Some Logicians, to the Early 1960s ". In M. Baaz (editor) *Kurt Gödel and the Foundations of Mathematics* 57-74. Cambridge: Cambridge University Press.

Goodstein, Reuben. 1944. "On the restricted ordinal theorem" . *The Journal of Symbolic Logic* 9: 33-41.

Hájek, Peter and Pudlák, Pavel. 1993. *Metamathematics of first order arithmetic.* Berlin: Springer.

Hamano, Masahiro and Okada, Mitsuhiro. 1997. "A relationship among Gentzen's Proof-Reduction, Kirby–Paris' Hydra Game and Buchholz's Hydra Game" . *Mathematical Logic Quarterly* 43: 103-120.

Hamano, Masahiro and Okada, Mitsuhiro. 1998. "A direct independence proof of Buchholz's Hydra Game on finite labeled trees" . *Archive for Mathematical Logic* 37: 67–89. (1998).

Hammer, Daniel and Romashchenko, Andrei and Shen, Alexander and Vereshchagin, Nikolay. 2000. "Inequalities for Shannon entropies and Kolmogorov complexities". *Journal of Computer and System Sciences* 60 (2): 442-464.

Hao, Yunge and Tourlakis, George. 2021."An arithmetically complete predicate modal logic". Bulletin of the Section of Logic. Łódź University. https://czasopisma.uni.lodz.pl/bulletin/article/view/8441

Harnik, Victor. 1992. "Provably Total Functions of Intuitionistic Bounded Arithmetic". *The Journal of Symbolic Logic* 57 (2): 466-477.

Harper, Robert. 2011. "Existential Type" . https://existentialtype.wordpress.com/2011/03/27/the-holy-trinity/

Harrison, John. 2009. *Handbook of practical logic and authomated reasoning.* Cambridge: Cambridge University Press.

Hartmanis, Juris and Stearns, Richard E. 1965. "On the computational complexity of algorithms" . *Transactions of American Mathematical Society* 117 (5): 285-306.

Henkin, Leon. (1950). "Completeness in the Theory of Types". *The Journal of Symbolic Logic.*" 15 (2): 81-91.

Herbrand, Jacques. 1930. "Recherches sur la théorie de la démonstration". *Travaux de la société des Sciences et des Lettres de Varsovie, Class III, Sciences Mathématiques et Physiques* 33: 1-131.

Hilbert, David and Ackermann, Wilhelm. 1928. *Grundzüge der theoretischen Logik.* Berlin: Springer.

Hilbert, David and Bernays, Paul. 1934. *Grundlagen der Mathematik* I. Berlin: Springer.

Hilbert, David and Bernays, Paul. 1939. *Grundlagen der Mathematik* II. Berlin: Springer.

Hindley, Roger J. and Seldin, Jonathan. 2008. *Lambda-Calculus and Combinators, an Introduction.* Cambridge: Cambridge University Press.

Hofstadter, Douglas R. 1979. *Gödel, Escher, Bach: An Eternal Golden Braid.* New York: Basic Books.

Hopcroft, John E. and Motwani, Rajeev and Ullman, Jeffrey D. 2000. *Introduction to Automata Theory, Languages, and Computation* (2nd ed.). Boston: Addison-Wesley.

Iemhoff, Rosalie. 2003."Preservativity logic: An analogue of interpretability logic for constructive theories". *Mathematical Logic Quarterly* 49 (3): 219-324.

Jech, Thomas. 1978. *Set theory.* Cambridge M.: Academic Press.

Jeřábek, Emil. 2016. "Division by zero". *Archive for Mathematical Logic* 55: 997-1013.

Jeroslow, Robert G. 1975. "Experimental logics and Δ_2 theories" . *Journal of Philosophical Logic* 4 (3): 253-267.

Jones, James. 1983. "Variants of Robinson's essentially undecidable theory R". *Archiv für mathematische Logik und Grundlagenforschung* 23: 61-64.

Lafitte, Grégory. 2002. "On Randomness and Infinity". In *Foundations of Information Technology in the Era of Network and Mobile Computing. IFIP-The International Federation for Information Processing*, edited by R. Baeza-Yates, U. Montanari, M. Santoro, 267-279. Boston: Springer US.

Lafont, Yves. 2004. "Soft linear logic and polynomial time" . *Theoretical Computer Science*, 318 (1-2): 163-180.

Lakatos, Imre. 1976. *Proofs and Refutations.* Cambridge: Cambridge University Press.

Lambek, Joachim and Scott, Philip J. 1986. *Introduction to Higher-Order Categorical Logic.* Cambridge: Cambridge University Press.

Lawvere, William F. 1969. "Diagonal arguments and cartesian closed categories". In *Category theory, homology theory and their applications II. Lecture Notes in Mathematics* 92, edited by P. J. Hilton, 134-145. Berlin: Springer.

Lawvere, William F. 1970. "Quantifiers and Sheaves". In *Actes Du Congres International Des Mathematiciens, publiés sous la direction du Comité d'organisation du congrès* 1, 329-334. Nice: Gauthier Villars.

Leivant, Daniel. 1990. "Subrecursion and lambda representation over free algebras". *In Feasible Mathematics. Progress in Computer Science and Applied Logic* 9, edited by S. R. Buss and P. J. Scott, 281-291. Basel: Birkhäuser.

Leivant, Daniel and Marion, Jean-Yves. 1993. "Lambda-calculus characterisations of polytime". *Fundamenta Informaticae* 19: 167-184.

Levin, Leonid. 1973. "Universal search problems" . *Problems of Information Transmission* 9 (3): 115-116.

Lindström, Per. 1969. "On Extensions of Elementary Logic" . *Theoria* 35: 1-11.

Löb, Martin. 1955. "Solution of a Problem of Leon Henkin". *The Journal of Symbolic Logic* 20 (2): 115-118.

Lucas, John R. 1961. "Minds, machines and Godel". *Philosophy* 36 (137): 112-127.

Kamke, Erich. 1932. "Über neuere Begriindungen der Wahrscheinlichkeitsrechnung" . *Jahresbericht der Deutschen Mathematiker- Vereinigung* 42: 121-149.

Kaye, Richard. 1991. *Models of Peano arithmetic*. Oxford: Oxford University Press.

Kaye, Richard. 2011. "Tennenbaum's theorem for models of arithmetic" . In *Set Theory, Arithmetic, and Foundations of Mathematics*, edited by J. Kennedy and R. Kossak 66-79. Cambridge: Cambridge UPress.

Katseff, Howard. 1978. "Complexity dips in random infinite binary sequences". *Information and Control* 38: 258-263.

Kelly, Kevin T. 2023. *The Logic of Reliable Inquiry*. Oxford: Oxford University Press.

Kennedy, Juliette. 2022. *Gödel's Incompleteness Theorems*. Cambridge: Cambridge University Press.

Ketonen, Jussi and Solovay, Robert. 1981. "Rapidly Growing Ramsey Functions". *Annals of Mathematics* 113 (2): 267-314.

Kirby, Laurie and Paris, Jeff. 1982. "Accessible Independence Results for Peano Arithmetic". *Bulletin of the London Mathematical Society* 14 (4): 285-293.

Kleene, Stephen C. 1952. *Introduction to Metamathematics*. Amsterdam: North-Holland.

Kleene, Stephen C. 1967. *Mathematical Logic*. New York: Wiley.

Kolmogorov, Andrej N. 1965. "Three Approaches to the quantitive definition of Information". *Problems of Information Transmission* 1: 1-17.

Kossak, Roman and Schmerl, James. 2006. *The Structure of Models of Peano Arithmetic*. Oxford: Oxford University Press.

Krajíček, Jan and Pudlák, Pavel and Takeuti, Gaisi. 1991. "Bounded arithmetic and the polynomial hierarchy." *Annals of Pure and Applied Logic* 52 (1-2): 143-153.

Krajíček, Jan. 1995. *Bounded arithmetic, propositional logic, and complexity theory*. Cambridge: Cambridge University Press.

Krajewski, Stanislaw. 2004. "Gödel on Tarski" . *Annals of Pure and Applied Logic* 127: 303-323.

Kreisel, Georg. 1952. "On the Interpretation of Non-Finitist Proofs: Part II. Interpretation of Number Theory. Applications" . *The Journal of Symbolic Logic* 17 (1): 43-58.

Kreisel, Georg. 1960. "Ordinal Logics and the Characterization of Informal Concepts of Proof". In Todd J. A. (editor), *Proceedings of the International Congress of Mathematicians* 289-299. Edinburgh, Cambridge: Cambridge University Press.

Kreisel, Georg. 1969. "Two notes on the foundations of set theory. ", *Dialectica* 23: pp. 93-114.

Kreisel, Georg. 1970. "A Survey of Proof Theory II" . In *Proceedings 2nd Scandinavian Logic Symposium*, edited by J. E. Fenstad 109-170. Amsterdam: North Holland.

Kreisel, Georg. 1971. "Some Reasons for Generalizing Recursion Theory". In *Studies in Logic and the Foundations of Mathematics* 61, edited by R.O. Gandy and C.M.E. Yates 139-198. Amsterdam: Elsevier.

Kripke, Saul A. 2021. "Mathematical Incompleteness Results in First-Order Peano Arithmetic: A Revisionist View of the Early History" . *History and Philosophy of Logic* 43 (2): 175-182.

Krivine, Jean Louis. 1993. *Lambda-calculus, Types and Models*. London: Ellis Horwood.

Kritchman, Shira and Ran Raz. 2010. "The surprise examination paradox and the second incompleteness theorem". *Notices Amer. Math. Soc* 57: 1454-1458.

Kučera, Antonin. 1985. "Measure, Π^0_1 -classes and complete extensions of PA." In *Recursion Theory Week. Lecture Notes in Mathematics* 1141, edited by H.D. Ebbinghaus, G.H. Müller and G.E. Sacks 245-259. Berlin: Springer.

Kučera, Antonin and Slaman, Theodore. 2001. "Randomness and Recursive Enumerability." *SIAM Journal of Computing* 31: 199-211.

Kugel, Peter. 1986. "Thinking may be more than computing." *Cognition* 18: 128-149.

Macintyre, Angus. 2011. "The Impact of Gödel's Incompleteness Theorems on Mathematics" . In *Kurt Gödel and the Foundations of Mathematics*, edited by M. Baaz 3-75. Cambridge: Cambridge University Press.

Mac Lane, Saunders. 1971. *Categories for the Working Mathematician*. Graduate Texts in Mathematics 5. Berlin: Springer.

Magari, Roberto. 1974. "Su certe teorie non enumerabili" . *Annali di Matematica Pura ed Applicata* 4, XCVIII: 119-152.

Magari, Roberto. 1975. "Representation and duality theory for diagonalizable algebras". *Studia Logica* 34 (4): 305-313.

Mal'cev, Anatoli Ivanovic. 1971. *The Metamathematics of Algebraic Systems. Collected Papers: 1936–1967*. Studies in Logic and the Foundations of Mathematics, Volume 66. Amsterdam: Elsevier.

Mancosu, Paolo and Galvan, Sergio and Zach, Richard. 2021. *An Introduction to Proof Theory: Normalization, Cut-Elimination, and Consistency Proofs*. Oxford: Oxford University Press.

Martin-Löf, Per. 1966. "The definition of random sequences" . *Information and Control* 9: 602-619.

Marker, David. 2002. *Model Theory: An Introduction*. Berlin: Springer.

Matiyasevich, Yuri. 1970. "Enumerable sets are Diophantine". *Doklady Akademii Nauk SSSR* 191: 279-282.

McCulloch, Warren S. and Pitts, Walter. 1943. "A Logical Calculus of the Ideas Immanent in Nervous Activity". *Bulletin of Mathematical Biophysics* 5: 115-133.

Montagna, Franco. 1978. "On the algebraization of a Feferman's predicate" . *Studia Logica* 37: 221–236.

Montagna, Franco. 1979. "On the diagonalizable algebra of Peano arithmetic." Bollettino della Unione Matematica Italiana B (5), 16: 795-812.

Montagna, Franco. 1987. "The predicate modal logic of provability". Notre Dame Journal of Formal Logic 25: 179-189.

Montagna, F. 1987. "Provability in finite subtheories of PA" . *The Journal of Symbolic Logic* 52(2): 494–511.

Montagna, Franco and Mancini, Antonella. 1994. "A minimal predicative set theory" . *Notre Dame Journal of Formal Logic* 35 (2): 186-203.

Moore, Cristopher. 1990. “Unpredictability and undecidability in dynamical systems.” *Physical review letters* 64 (2): 2354-2357.

Mostowski, Andrzej W.. 1957. “On recursive models of formalised arithmetic”. *Bulletin de l’Académie Polonaise des Sciences*, Classe III, (5): 705-710.

Mojtahedi, Mojtaba. 2022. “On provability logic of HA” . arXiv:2206.00445 [math.LO].

Murawski, Roman. 1999. *Recursive Functions and Metamathematics*. Dordrecht: Kluwer.

Nelson, Edward. 1986. *Predicative Arithmetic*. Princeton: Princeton Univ. Press.

Nerode, Anil and Odifreddi, Piergiorgio. 1990. *Lambda Calculi and Constructive Logics*. Mathematical Sciences Institute, Cornell University.

Nielsen, Michael A. and Chuang, Isaac L. 2010. *Quantum Computation and Quantum Information: 10th Anniversary Edition*. Cambridge: Cambridge University Press.

Nies, André. 2009. *Computability and Randomness*. Oxford: Oxford University Press.

Nies, André and Scholz, Volkher B. 2017. “Martin-Löf random quantum states” . *Journal of Mathematical Physics* 60 (9): 1-11.

Odifreddi, Piergiorgio. 1989. *Classical Recursion Theory* I. Amsterdam: North Holland.

Odifreddi, Piergiorgio. 1999. *Classical Recursion Theory* II. Amsterdam: North Holland.

Odifreddi, Piergiorgio. 1996. “Kreisel’s Church”. In P. Odifreddi (editor) In *Kreiseliana* 389-415. Boca Raton: A. K. Peters/CRC Press.

AA.VV. *Open Logic Project*. https://openlogicproject.org/

Ord, Toby and Kieu, Tien D. 2002. “On the existence of a new family of Diophantine equations for Ω ”. *Fundamenta Informaticae* 56 (3): 273-284.

Osherson, Daniel. N. and Stob, Michael and Weinstein, Scott. 1986. *Systems that learn: An introduction to learning theory for cognitive and computer scientists*. Boston: MIT Press.

Parikh, Rohit. 1971. “Existence and Feasibility.” *The Journal of Symbolic Logic* 36 (3): 494-508.

Paris, Jeff B. 1987. “Some independence results for Peano arithmetic.” *The Journal of Symbolic Logic* 43 (4): 725-731.

Paris, Jeff B. 1980. “A hierarchy of cuts in models of arithmetic.”In *Model theory of algebra and arithmetic, Proceedings of the Conference on Applications of Logic to Algebra and Arithmetic held at Karpacz, Poland, September 1–7, 1979. Lecture notes in mathematics 834*, edited by L. Pacholski, J. Wierzejewski, and A. J. Wilkie 312-337. Berlin: Springer.

Paris, Jeff and Dimitracopoulos, Costas. 1983, “A Note on the Undefinability of Cuts”. *The Journal of Symbolic Logic* 48: 564-569.

Paris, Jeff and Harrington, Leo. 1977. “A mathematical incompleteness in Peano Arithmetic”. In *Handbook of Mathematical Logic*, edited by J. Barwise and : Keisler 1133-1142. Amsterdam: North-Holland.

Parsons, Charles. 1970. “On a Number Theoretic Choice Schema and its Relation to Induction” . in A. Kino, J. Myhill, R.E. Vesley (editors) *Studies in Logic and the Foundations of Mathematics* vol. 60. Amsterdam: Elsevier.

Penrose, Roger. 1994. *Shadows of the Mind*. Oxford University Press: Oxford

Peres, Asher. 1985. "Einstein, Gödel, Bohr" *Foundations of Physics* 15: 201-205.

Peres, Asher and Zurek, Wojciech H. 1982. "Is quantum theory universally valid?". *American Journal of Physics* 50 (9): 807-810.

Peter, Rozsa. 1957. *Rekursive Funktionen*. Berlin, Boston: de Gruyter.

Péter, Rózsa. 1932. "Rekursive Funktionen." In *Verhandlungen Des Internationalen Mathematiker Kongresses Zürich 2*, edited by W. Saxer 336-337. Zurich: Orell Füssli Verlag.

Poggiolesi, Francesca. 2009."A purely syntactic and cut-free sequent calculus for the modal logic of provability". *Review of Symbolic Logic* 2 (4): 593-611.

Pohlers, Wolfram. 1993. "A short course in ordinal analysis". In *Proof Theory*, edited by P. Aczel, H. Simmons and S. S. Wainer 27-78. Cambridge: Cambridge University Press.

Post, E.L. 1936. "Finite Combinatory Processes - Formulation 1". *The Journal of Symbolic Logic* 1: 103-105.

Post, Emil. 1941. "Absolutely Unsolvable Problems and Relatively Undecidable Propositions: Account of an Anticipation." In *The Undecidable* 1965, edited by M. Davis 340-433. New York: Raven Press.

Post, Emil. 1944. "Recursively enumerable sets of positive integers and their decision problems". *Bulletin of the American Mathematical Society* 50: 284-316.

Post, Emil L. 1948. "Degrees of recursive unsolvability: preliminary report". *Bulletin of the American Math. Soc.* 54: 641-642.

Potthoff, Klaus. 1969. "Über Nichtstandardmodelle der Arithmetik und der rationalen Zahlen". *Zeitschrift für mathematische Logik und Grundlagen der Mathematik* 15 (1969): 223-236.

Pudlák, Pavel. 1985. "Cuts, consistency statements and interpretation" . *The Journal of Symbolic Logic* 50: 423-441.

Pudlák, Pavel. 1996. "On the lengths of proofs of consistency." *Collegium Logicum, annals of the Kurt Gödel Society 2*: 65-86.

Putnam, Hilary. 1965. "Trial and error predicates and the solution to a problem of Mostowski." *The Journal of Symbolic Logic* 30 (1): 49-57.

Raatikainen, Panu. 1998. "On Interpreting Chaitin's Incompleteness Theorem". *Journal of Philosophical Logic* 27 (6): 569-586.

Raatikainen, Panu. 2000. "Algorithmic information theory and undecidability" . *Synthese* 123 (2): 217-225.

Razborov, Alexander A. 1993. "An equivalence between second order bounded domain bounded arithmetic and first order bounded arithmetic." In P. Clote and J. Krajicek (editors) *Arithmetic, Proof Theory and Computational Complexity* 247-277. Oxford: Oxford University Press.

Rathjen, Michael. 2006. "The art of ordinal analysis." In M. Sanz Solé, J. Soria de Diego, J. L. Varona Malumbres, J. Verdera (editors) *Proceedings of the International Congress of Mathematicians 2* 45-70. Berlin: EMS Press.

Rathjen, Michael. 1999. "The Realm of Ordinal Analysis" . In Cooper S. B. and Truss J.K. (editors) *Sets and Proofs* 219-280. London Mathematical Society Lecture Note Series. Cambridge: Cambridge University Press.

Rathjen, Michael. 2015. "Goodstein's theorem revisited". In R. Kahle and M. Rathjen (editors) *Gentzen's Centenary: The Quest for Consistency*: 229-242. Berlin: Springer.

Robinson, Raphael. 1950. "An Essentially Undecidable Axiom System". In *Proceedings of the International Congress of Mathematics, 1950*: 729-730. Providence: AMS.

Rogers, Hartley. 1987. *The Theory of Recursive Functions and Effective Computability.* Boston: MIT Press.

Ryll-Nardzewski, Czesław. 1952. "The role of the axiom of induction in elementary arithmetic". *Fundamenta Mathematicae* 39: 239-63.

Sacks, Gerald. 2017. *Higher Recursion Theory.* Cambridge: Cambridge University Press.

Sambin, Giovanni. 1976. "An effective fixed-point theorem in intuitionistic diagonalizable algebras". *Studia Logica* 35: 345-361.

Sasaki, Katsumi. 2001. *Löb's axiom and cut-elimination theorem.* ILLC Dissertation Series DS-2001-07, Amsterdam.

Schnorr, Claus-Peter. 1971. "A unified approach to the definition of random sequences.". *Mathematical Systems Theory* 5: 246-258.

Schnorr, Claus Peter. 1973. "Process complexity and effective random tests". *Journal of Computer and System Sciences* 7 (4,) 376-388.

Schütte, Kurt. 1964. "Eine Grenze für die Beweisbarkeit der transfiniten Induktion in der verzweigten Typenlogik". *Archiv für Mathematische Logik und Grundlagenforschung* 7(1–2): 45–60.

Schütte, Kurt. 1965. "Predicative Well-Orderings". In J.N. Crossley and M.A.E. Dummett (editors), *Formal Systems and Recursive Functions* 280–303. Amsterdam: North Holland.

Schütte, Kurt. 1960. *Beweistheorie*, (Grundlehren der mathematischen Wissenschaften, 103), Berlin: Springer.

Schwichtenberg, Helmut. 1977. "Proof Theory: Some Applications of Cut-Elimination ". In *Handbook of Mathematical Logic. Studies in Logic and the Foundations of Mathematics 90*, edited by J. Barwise 867-895. Amsterdam: Elsevier.

Schwichtenberg, Helmut. 1976. "Definierbare Funktionen im Lambda-Kalkul mit Typen". *Archiv Logik Grundlagenforsch.* 17: 113-114.

Schwichtenberg, Helmut. 2012. *Mathematical Logic* (lecture notes). München: Ludwig-Maximilians-Universität.

Schwichtenberg, Helmut and Wainer, Stanley S. 2011. *Proofs and Computations. Perspectives in Logic.* Cambridge: Cambridge University Press.

Scott, Dana. 1980."Relating theories of the lambda-calculus ". In *To H.B. Curry: Essays on Combinatory Logic, Lambda-Calculus and Formalism*, edited by J. R. Hindley and J. P. Seldin 403-450. Cambridge M.: Academic Press.

Segerberg, Krister. 1971. *An essay in classical modal logic.* Filosofiska Föreningen och Filosofiska Institutionen vid Uppsala Universitet, Uppsala.

Shannon, Claude. 1984. "A mathematical theory of communication". *Bell System Technical Journal* 27: 379-423 and 623-656.

Shavrukov, Volodya Yu. 1988. "The logic of relative interpretability over Peano arithmetic". Preprint, Steklov Mathematical Institute, Moscow, 1988. In Russian.

Shepherdson, John C. 1964. "A nonstandard model for a free variable fragment of number theory". *Bulletin of the Polish Academy of Sciences* 12: 7.

Sieg, Wilfrid. 1999. "Hilbert's programs: 1917-1922". *Bulletin of Symbolic Logic* 5 (1): 1-41.

Simpson, Stephen G. 1990. "Nonprovability of Certain Combinatorial Properties of Finite Trees". *The Journal of Symbolic Logic 55 (2)*: 868-869.

Simpson, Stephen ed. 1987. *Logic and combinatorics, Proceedings of the AMS-IMS-SIAM joint summer research conference held August 4-10, 1985. Contemporary Mathematics, volume 65.* Providence: American Mathematical Society.

Simpson, Stephen G. 1999. *Subsystems of Second Order Arithmetic.* Berlin: Springer.

Sipser, Michael. 2006. *Introduction to the Theory of Computation*, second edition. Boston: Course Technology.

Skolem, Thoralf. 1923. "The foundation of elementary arithmetic established by means of the recursive mode of Thougt, without the use of apparent variable ranging over infinite domains". In *From Frege to Gödel, a source book in mathematical logic* 302-334, edited by J. Van Heijenhoort. Cambridge M.: Harvard Univ. Press.

Skolem,Thoralf. 1955. "Peano's axioms and models of arithmetic". In *Mathematical Interpretations of Formal Systems*, edited by T. Skolem, G. Hasenjaeger, G. Kreisel, A. Robinson, H. Wang. L. Hen and J. Loś 1-14. Amsterdam: North-Holland.

Smorynski, Craig. 1973. "Applications of Kripke models" . In Metamathematical Investigation of Intuitionistic Arithmetic and Analysis, edited by A. S. Troelstra 324-391. Berlin: Springer.

Smorynski, Craig. 1984. "Lectures on Nonstandard Models of Arithmetic: Commemorating Guiseppe Peano". In *Logic Colloquium '82. Studies in Logic and the Foundations of Mathematics* 112, edited by G. Lolli, G. Longo, A. Marcja 1-70. Amsterdam: Elsevier.

Smorynski, Craig. 1985. *Self-Reference and Modal Logic.* Berlin: Springer.

Smullyan, Raymond Merrill. 1961. *Theory of Formal Systems.* Princeton: Princeton University Press.

Smorynski, Craig. 1987. "Quantified modal logic and selfreference". *Notre Dame Journal of Formal Logic* 28: 356-370.

Soare, Robert I. 1996. "Computability and Recursion". *The Bulletin of Symbolic Logic* 2 (3): 284-321.

Soare, Robert I. 2009. "Turing oracle machines, online computing, and three displacements in computability theory". *Annals of Pure and Applied Logic* 160 (3): 368-399.

Soare, Robert I. 2014. "Turing and the discovery of computability". In *Turing's Legacy: Developments from Turing's Ideas in Logic. Lecture Notes in Logic*, edited by R. Downey 467-492. Cambridge: Cambridge University Press.

Solomonoff, Ray. 1964. "A formal theory of inductive inference. I and II". *Information and Control* 7: 1-22 and 224-254.

Solovay, Robert. 1970. "A Model of Set Theory in Which Every Set of Reals is Lebesgue Measurable". *Annals of Mathematics. Second Series*: 1-56.

Solovay, Robert. 1976. “On Interpretability in Peano Arithmetic.”Unpublished letter to P. Hájek. https://www.cs.cas.cz/hajek/RSolovayZFGB.pdf

Solovay, Robert M. 1976. “Provability Interpretations of Modal Logic” . *Israel Journal of Mathematics* 25: 287-304.

Solovay, Robert. 2000. “A Version of Ω for which ZFC can not Predict a Single Bit”. In *Finite Versus Infinite. Contributions to an Eternal Dilemma*, edited by C. Calude and G. Paŭn 323-334. Berlin: Springer.

Sommer, Richard. 1990. *Transfnite induction and hierarchies generated by transfinite recursion within Peano arithmetic*, Ph.D. Thesis, U.C. Berkeley.

Sommer, Richard. 1995. “Transfinite induction within Peano arithmetic ” . *Annals of Pure and Applied Logic* 76 (3): 231-289.

Sørensen, Morten Heine and Urzyczyin, Pavel. 2006. *Lectures on the Curry-Howard Isomorphism*. Amsterdam: Elsevier.

Statman, Richard. 1979. “The typed λ-calculus is not elementary recursive.” *Theoretical Computer Science* 9(1): 73-81.

Statman, Richard. 1979“ . Intuitionistic propositional logic is polynomial-space complete”. *Theoretical Computer Science* 9 : 67-72.

Stephan, Frank. 2006. “Martin-Löf random and PA-complete sets” . In Logic Colloquium ’02. Lecture Notes in Logic, edited by Z. Chatzidakis, P. Koepke and W. Pohlers W. 342-348. Cambridge: Cambridge University Press.

Švejdar, Vítězslav. 1983. “Modal analysis of generalized Rosser sentences.” *The Journal of Symbolic Logic* 48: 986-999.

Švejdar, Vítězslav. 2007. “An interpretation of Robinson arithmetic in its Grzegorczyk’s weaker variant”. *Fundamenta Informaticae* 81 (1-3): 347-354.

Tait, William W. 1968. “Normal derivability in classical logic”. In *The Syntax and Semantics of Infinitatry Languages. Lecture Notes in Mathematics 72*, edited by J. Barwise, 204-236. Berlin: Springer.

Tait, William. 1981. “Finitism” . *Journal of Philosophy* 78 (9): 524-546.

Takahashi, Masako. 1995. “Parallel Reductions in λ-Calculus ”. *Information and Computation* 118 (1): 120-127.

Takeuti, Gaisi. 1987. *Proof Theory*. Amsterdam: North Holland (Reprint Courier Corporation, 2013).

Takeuti, Gaisi. 1993. “RSUV isomorphism.” In *Arithmetic, Proof Theory and Computational Complexity*, edited by P. Clote and J. Krajicek, 364-386. Oxford: Oxford University.

Tarski, Alfred and Mostowski, Andrzej and Robinson, Raphael M. 1953. *Undecidable Theories*. Amsterdam: Elsevier.

Tennenbaum, Stanley. 1959. “Non-Archimedean models for arithmetic.” *Notices of the American Mathematical Society* 6: 270.

Terui, Kazushige. 2007.“Light Affine Lambda-calculus and polytime strong normalization”. *Archive for Mathematical Logic* 46: 253-280.

Towsner Henry. 2020. “Algorithmic randomness in ergodic theory”. In *Algorithmic Randomness: Progress and Prospects. Lecture Notes in Logic*, edited by J.N.Y. Franklin and C.P. Porter 40-57. Cambridge: Cambridge University Press.

Troelstra, Anne S. and Van Dalen, Dirk. 1988. Constructivism In Mathematics. An Introduction. I and II. Amsterdam: North Holland.

Troelstra, Anne S. and Schwichtenberg, Helmut. 2000. *Basic proof theory.* Cambridge: Cambridge University Press.

Turing, Alan M. 1936. “On computable numbers, with application to the Entscheidungsproblem”, in *Proceedings of London Math. Soc.* 42: 230-265, 544-546.

Turing, Alan M. 1937. “Computability and lambda definability”. *The Journal of Symbolic Logic* 2 (4): 153-163.

Turing, Alan. 1939. “Systems of logic based on ordinals.”Proc. London Mathematical Society 2: 161-228.

Turing, Alan. 1940b. Letter to Newman, dated 21 April, added by R.O. Gandy, AMT D/2, Contemporary Scientific Archives Centre, King’s College Library, Cambridge.

Turing, Alan M. 1947. “Lecture to the London Mathematical Society on 20 February 1947.” In *The Essential Turing*, edited by B. J. Copeland, 378-394. Oxford: Oxford U.P.

Urban, Christian and Bierman, Gavin M. 2001. “Strong Normalisation of Cut-Elimination in Classical Logic.”*Fundamenta Informaticae* 45 (1-2): 123-155.

Ursini, Aldo. 1979.“Intuitionistic diagonalizable algebras” . Algebra Universalis 9: 229-237.

Valentini, Silvio. 1983.“The modal logic of provability: Cut-elimination” . *Journal of Philosophical Logic* 12: 471-476.

Van Dalen, Dirk. 2001. “Intuitionistic Logic”. In L. Goble (editor). *The Blackwell Guide to Philosophical Logic* 224–257. Malden, Mass.: Wiley-Blackwell.

Van Dalen, Dirk. 2013. *Logic and Structure* (5th ed.). Berlin: Springer.

Van Emde Boas, Peter. 1990. “Machine Models and Simulations” . In *Handbook of Theoretical Computer Science, Algorithms and Complexity*, edited by L. Van Leeuwen, 1-66. Amsterdam: Elsevier.

Van Lambalgen, Michiel. 1987. “Von Mises’ Definition of Random Sequences Reconsidered.” *The Journal of Symbolic Logic* 52 (3): 725-755.

Van Lambalgen, Michiel. 1987. *Random Sequences.* Doctoral thesis, University of Amsterdam.

Van Lambalgen, Michiel. 1987. “Algorithmic information theory” *The Journal of Symbolic Logic* 54 (4): 1389-1400.

Van Lambalgen, Michiel. 1990. “The axiomatization of randomness” *The Journal of Symbolic Logic* 55 (3): 1143-1167.

Van Lambalgen, Michiel. 1992. “Independence, randomness and the axiom of Choice” . *The Journal of Symbolic Logic* 57 (4): 1274-1304.

Vardanyan, Valeri Aramovich. 1986. “Arithmetic complexity of predicate logics of provability and their fragments.”Doklady Akad. Nauk SSSR, 288 (1): 11-14.

Verbrugge, Rineke. 1993. *Efficient Metamathematics.* Dissertation, Universiteit Van Amsterdam.

Verbrugge, Rineke and Berarducci, Alessandro. "On the provability logic of bounded arithmetic". *Annals of Pure and Applied Logic* 61 (1-2): 75-93.

Ville, Jean. 1939. *Etude critique de la notion de collectif.* Gauthier-Villars: Paris.

Visser, Albert. 1989. "Peano's smart children. A provability logical study of systems with built-in consistency" . *Notre Dame Journal of Formal Logic* 30: 161-196.

Visser, Albert. 1990. "Interpretability logic" . In *Mathematical Logic*, edited by P. P. Petkov 175-208. New York: Plenum Press.

Visser, Albert. 2009. "Why the theory R is special". Logic Group preprint series, 279. Utrecht Univ. Now in In Friedman, H. and Tennants, N. (eds.) 2014. *Foundational Adventures. Essay in honour of Harvey Friedman* 7-23. London: College Publications.

Visser, Albert and Beklemishev, Lev. 2006. "Problems in the Logic of Provability". 2006. In D.M. Gabbay, S. S. Goncharov and N. Zakharyaschev (editors), *Mathematical Problems from Applied Logic I.* International Mathematical Series, vol 4. New York: Springer.

Visser, Albert and De Jonge, Maartje. 2006. "No Escape from Vardanyan's theorem". *Archive for Mathematical Logic* 45: 539–554.

Visser, Albert and De Jongh, Dick and Van Benthem, Johannes and Renardel de Lavalette, Jean Gerard. 1995. "NNIL, A study in intuitionistic propositional logic" In Modal Logic and Process Algebra, edited by A. Ponse M. de Rijke Y. Venema. Stanford: CSLI.

Von Neumann, John. 1951. "The general and logical theory of automata". In *Cerebral mechanisms in behavior; the Hixon Symposium*, edited by L. A. Jeffress, 1-41. Hoboken: Wiley.

Von Plato, Jan. 2014. "The Development of Proof Theory" . Stanford Encyclopedia of Philosophy https://plato.stanford.edu/entries/proof-theory-development/

V'yugin, Vladimir. 2022. "Ergodic theorems for algorithmically random points". 10.48550/arXiv.2202.13465.

Wadsworth, Christopher P. 1971. *Semantics and Pragmatics of the Lambda-Calculus.* Dissertation. Oxford University.

Wald, Abraham. 1937. "Die Widerspruchsfreiheit des Kollektivbegriffes der Wahrscheinlichkeitsrechnung". *Ergebnisse eines Math. Kolloquiums* 8: 38-72.

Wang, Hao. 1974. *From Mathematics to Philosophy.* New York: Routledge and Kegan.

Wang, Hao. 1987. *Reflections on Kurt Gödel.* Cambridge M.: MIT Press.

Wang, Hao. 1996. *A logical journey: from Gödel to philosophy.* Cambridge M.: MIT Press.

Weaver, Nik.2005."Predicativity beyond Γ_0". arXiv:math/0509244.

Wilkie, Alex J. and Paris, Jeff B. "On the scheme of induction for bounded arithmetic formulas". *Annals of Pure and Applied Logic* 35 : 261-302.

Yavorsky, Rostislav. 2002."On arithmetical completeness of first-order logics of provability". In F. Wolter, H. Wansing, M. de Rijke and M. Zakharyaschev (editors) *Advances in Modal Logic 3* 1–16. Singapore: World Scientific Publishing,

Yurtsever, Ulvi. 2000. "Quantum mechanics and algorithmic randomness". *Complexity* 6, n. 1: 27-34.

Zach, Richard. 1998. "Numbers and functions in Hilbert's finitism". *Taiwanese Journal for Philosophy and History of Science* 10: 33-60.

Zaffora Blando, Francesca. 2024. “From Wald to Schnorr: Von Mises’ definition of randomness in the aftermath of Ville’s Theorem.”. *Studies in History and Philosophy of Science* 106: 196-207.

Zambella, Domenico. 1996. “Notes on Polynomially Bounded Arithmetic”. *The Journal of Symbolic Logic* 61 (3): 942-966.

Zoethout, Jetze and Visser, Albert. 2019. “Provability Logic and the Completeness Principle”. *Annals of Pure and Applied Logic* 6: 718-753.

Key concepts

FUP Best Practice in Scholarly Publishing (DOI 10.36253/fup_best_practice)

UNIverSI
Ricerca e Didattica all'Università di Siena

TITOLI PUBBLICATI

DIDATTICA

Francesco P. Vetere, *Elementi di petrologia sperimentale*

Stefania Lamponi, *Chimica generale: esercizi svolti. Raccolta di esercizi con soluzioni dettagliate per la preparazione alla prova scritta dell'esame di Chimica generale*

Duccio Pianigiani, *Lectures in Proof Theory and Complexity*

RICERCA

Lara Lazzeroni, *Responsabilità sociale d'impresa 2.0 e sostenibilità digitale. Una lettura giuslavoristica*

UNIverSI

Ricerca e **Didattica**
all'Università di Siena

The book is based on lecture notes from the course 'Formal Systems', taught by the author for the Master's Degree in 'Applied Mathematics' at the University of Siena. It was created for educational purposes, specifically for second-level (graduate) courses.
The work is mainly oriented towards applications of Proof Theory — one of the macro-areas into which Mathematical Logic is divided — to Computability Theory and Computational Complexity Theory, albeit with entanglements with Model Theory and with Category Theory.
The book begins with some classical results concerning formal arithmetic, dating back to the 1930s, and then compares them with more recent developments, emphasising the acceleration imparted to logical study by the development of computer science.

Duccio Pianigiani graduated in Philosophy of Science at the University of Siena. He obtained a diploma from the Postgraduate School of Logic at the local Department of Mathematics, and later a PhD in the same field from the Department of Philosophy at the University of Florence. He is currently a researcher at the Department of Information Engineering and Mathematics, University of Siena.

ISSN 3035-5915 (print)
ISSN 3035-5931 (online)
ISBN 979-12-215-0778-2 (PDF)
ISBN 979-12-215-0779-9 (XML)
DOI 10.36253/979-12-215-0778-2
www.fupress.com