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.