Hilbert's Programme and Gödel's Theorems

Aleid Oosterwijk
Supervisor: Jaap van Oosten
Universiteit Utrecht

June 26, 2013

Contents

1 Introduction

Around 1900, Hilbert started working on the foundations of mathematics. His work on the foundations would later be known as Hilbert's programme, a programme aiming to prove consistency of mathematics by means of restricted methods. In 1930, when Hilbert and others were still working on this programme, Gödel presented his incompleteness theorems, which have been said to refute Hilbert's programme. But do they? After this 'refutation', many mathematicians have been working on the programme, adapting it to avoid the consequences of the incompleteness theorems.

To be able to answer the question whether or not Hilbert's programme has really been refuted by the incompleteness theorems, we will first look at Hilbert's programme: its historical development and an explanation of its content. This will mostly be in line with Smorynski [1986]. After that we will discuss Peano Arithmetic, some logic and primitive recursive functions, to be able to understand the next main subject: Gödel's theorems. We will discuss the theorems, their proofs, the theory Gödel developed to give these proofs, and how the theorems refute Hilbert's programme. Except for the refutation, this will mostly be in line with van Oosten [2009]. Then we will look at the adjustments made to the programme, developments in different directions, by e.g. Feferman and Gentzen, mostly in line with Zach [2006]. We will conclude by answering the question whether or not the programme has been refuted by the incompleteness theorems.

2 Hilbert's Programme

In 1921, Hilbert presented his programme. The main idea of the programme was to axiomatise all classical mathematics. The consistency of the system obtained this way should be proved, but only by means of certain restricted methods. The methods Hilbert allowed he called finitary, and during the development of the programme he specified more and more which methods were to be accepted as finitary.

2.1 Historical Development of the Programme

Originally, in mathematics the traditional Aristotelian approach to axiomatisation was accepted. This approach is roughly as follows: an axiom is a 'statement worthy of acceptance'[1], which means it is an evident truth about the structure one is axiomatising. Theorems can be deduced from these axioms via proofs, which are following certain logical rules. Using these means of proving, true axioms will lead to true theorems.

This Aristotelian approach was slowly being abandoned during the 18th and 19th centuries. Axioms no longer needed to be evident truths, one could also create systems which are different from the systems using only 'true' axioms. For example, non-Euclidean geometry was being studied by Schweikert, Gauss, Bolyai and others.[2]. George Peacock presented a new sort of algebra which was not dealing with quantities at all, a purely symbolic system. Other examples are the quaternions, an invention of Hamilton, or Cayley's numbers. All these examples show the enormous possibilities that mathematics could offer, once the axiomatic view had changed.

In 1899, Hilbert published his lectures on Euclidian geometry, the *Grundlagen der Geometrie*[3], in which Hilbert described the Euclidean geometry in a formal axiomatic way. The *Grundlagen* was the first major work written using this axiomatic freedom instead of the Aristotelian approach.

In 1902, Hilbert formulates his opinion on the axiomatic method as follows:[4]

> Every science takes its starting point when[5] a sufficiently coherent body of facts is given. It takes form, however, only by *organising* this body of facts. This organisation takes place through the *axiomatic method*, i.e., one constructs a *logical structure of concepts* so that the relationships between the concepts correspond to the relationships between the facts being organised.
>
> There is arbitrariness in the construction of such a structure of concepts; we, however, *demand* of it:
>
> 1) completeness, 2) independence, 3) consistency.

This last consideration was of special importance to Hilbert, especially the consistency of arithmetic of the reals. Before he published any of his consistency

[1] [Mendell, 2008]

[2] [Smorynski, 1986] p. 5

[3] [Hilbert, 1899]

[4] From [Zach, 2006], quoting [Majer and Hallett, 2004].

[5] In the original citation by Zach it was if instead of when. This adaptation I have made because the original text by Hilbert says: 'Jede Wissenschaft nimmt ihren Anfang dann, wenn ein genügendes zusammenhängendes Tatsachenmaterial vorliegt.' [Majer and Hallett, 2004]

proofs, Bertrand Russell discovered his famous paradox in set theory:

$$R = \{x : x \notin x\}, \text{ then } R \in R \text{ if and only if } R \notin R \tag{1}$$

This paradox could be derived in the work of Frege. Frege is famous for reconceiving the discipline of logic: he constructed a formal system in which he constituted the first predicate calculus. Also he criticized Hilbert's understanding and use of the axiomatic system.[6] After the news of this paradox, other paradoxes were found and many mathematicians and philosophers worried about the correctness of foundational work. Hilbert however, had known about these paradoxes for a few years already and remained unperturbed.[7] These paradoxes convinced Hilbert of the need for rigour in foundational mathematics, but also of the fact that logic alone wasn't enough to describe fully the foundations of mathematics.

In 1904, Hilbert showed his first consistency proofs. These showed the consistency of some very weak systems. These results were not very impressive, and one explanation for that is the fact that logic was still to be developed at that time. Only Frege had experience in working within formal systems, and his formal system was inconsistent. It was only in the 1920s that the distinction between first-order and second-order logic was made, as well as the nature of quantifiers was understood. Hilbert would do a lot of groundwork for these new understandings, but in 1904 this was all still unknown.

Brouwer, a Dutch mathematician and philosopher, looked at the formalization of mathematics from a very different perspective. As Hilbert wanted to formalize mathematics and prove its consistency merely to be able to do the 'usual mathematics' without having to worry about the foundations, Brouwer was more interested in the foundations of mathematics itself, as he was also a philosopher. Mathematics is, according to Brouwer, working out the consequences of our mathematical intuition, which is a fundamental intuition we all share. He states that the parts of mathematics that cannot be based on this intuition must be rejected. Hilbert, however, believed that every mathematical problem could be solved.

In 1908, Brouwer published a paper entitled *De onbetrouwbaarheid der logische principes (The unreliability of the logical principles)* in which he questioned the laws of logic. Brouwer states that the Law of the Excluded Middle ($\varphi \vee \neg\varphi$), does not carry over to the infinite case. He states this as follows:[8]

> [...] the question of the validity of the principium tertii exclusi is equivalent to the question *whether unsolvable mathematical problems can exist.* There is not a shred of a proof for this conviction, which has sometimes been put forward, that there exist no unsolvable mathematical problems.

[6] [Zalta, 2013]

[7] 'The father of set theory, Cantor, had noticed similar difficulties already in 1895 (as witnessed by Bernstein and by letters to Hilbert and Dedekind).' about the Burali-Forti paradox, in [Cantini, 2012], 'He discovered the paradoxes while working on his survey papers of 1895 and 1897 and he wrote to Hilbert in 1896 explaining the paradox to him.' about the paradoxes he discovered simultanuously with Burali-Forti, in[O'Connor and Robertson, 1988].

[8] [Smorynski, 1986] p. 14, quoting Brouwer.

Or, as he puts it in his thesis[9]:

> [...] does it then follow from the non-contradictoriness of the logical system, that such a mathematical building exists? [...] So a fortiori it is not sure, that for every mathematical problem either a solution can be given, or it can be logically proved, that it is unsolvable; something that Hilbert meanwhile deems, in "Mathematische Probleme", every mathematician is utterly convinced of.

The equivalence Brouwer states here, is a consequence of how Brouwer sees mathematics. If one states $\varphi \vee \neg\varphi$, one must find a (finite) construction accomplishing the task demanded by φ, so that φ is true, or find a (finite) construction that precludes the process of performing the task demanded by φ, so that φ is false. The basis for this statement is Brouwer's placement of mathematics in the intellect[10]: a statement is true when it is known to be true, and statement is false when it is known to be false. This explains the equivalence between The Law of the Excluded Middle and the solvability of all mathematical problems.

It must be noted that Brouwer never claimed that the Law of the Excluded Middle was in fact false. Brouwer accepted the consistency of the Law of the Excluded Middle, but he distinguished between 'provable truths' and provable 'non-contradictories', so he wasn't led to believe the Law of the Excluded Middle was in fact true.

Brouwer's opinion can be illustrated by the following example, from [Kuiper, 2004], regarding the decimal expansion of π. In Brouwer's time, it was still unknown whether or not the sequence 0123456789 ever appeared in the decimal expansion of π. He used this fact in the following example. Define the denumerable infinite set $\{x_n\}$ on $[0, 1]$ as follows:

$x_n = 2^{-n}$ if for some $m < n$, at the m^{th} decimal place of the expansion of π, for the first time the sequence 0123456789 appears.

$x_n = 1 - 2^{-n}$ if this is not the case.

In classical logic, one would say that either 1 or 0 is a limit point of this set, as we know that either that sequence does appear in the decimal expansion, or it does not. In intuitionistic logic, one can not use that principle so one can not say that either 1 or 0 is a limit point. (Note that this means that also the theorem of Bolzano-Weierstrass does not hold in intuitionistic logic!) Because, if we could say that either 0 or 1 is a limit point, this would mean that we *know* that either the sequence appears in the decimal expansion, or that is does not. Now this means, according to Brouwer, that we could find a construction showing that indeed that sequence appears, or a construction showing that is does not. And at that time this was clearly not the case.

By now we know that this sequence does appear in the decimal expansion of π. The argument is still valid though, as it also holds for many other problems, that have not yet been solved.

It is clear that the questioning of the Law of the Excluded Middle and his criticism that consistency need not be enough as a condition to imply existence,

[9] [Brouwer, 1907], pp. 141-142, translated from: '[...]volgt dan uit de niet-strijdigheid van het logische systeem, dat zulk een wiskundig gebouw bestaat?[...]Het is dus a fortiori niet zeker, dat van elk wiskundig probleem f de oplossing kan worden gegeven f logisch kan worden aangetoond, dat het onoplosbaar is; iets, waarvan intusschen HILBERT in ,,Mathematische Probleme" meent, dat ieder wiskundige ten innigste is overtuigd.'

[10] [Smorynski, 1986] p. 14

are both opposed to Hilbert's view.

In 1917, Hilbert posed five epistemological questions[11]:

1. the problem of the solvability in principle of each mathematical question;
2. the problem of the additional verifiability[12] of the results of a mathematical investigation;
3. the question of a criterion of simplicity of mathematical proofs;
4. the question of the relation between content and formalism in mathematics and logic;
5. the problem of the decidability of a mathematical question through a finite number of operations.

For the fourth question, we will return to it later, when discussing Hilbert's new version of his programme.

To Brouwer, the first question could be answered negatively, as it is equivalent to the Law of the Excluded Middle, as we have seen before in the quotes of Brouwer given and the explanation of one of his counterexamples. Moreover, to him, the first and fifth question were essentially the same: for him solvability meant solvable by means of a finite number of operations.

In 1921, Hilbert started working on foundations again, something he had been considering since 1917, when he hired Bernays as his assistant. He lectured several times and the first, but inadequate, description of Hilbert's programme appeared to print. Hilbert's programme intended to prove the consistency of arithmetic, but the ground rules were different than they had been in 1904, as Hilbert made a distinction between actual mathematics and metamathematics, a distinction that will be explained shortly.

Bernays, who was also schooled as a philosopher, indirectly explained the importance of and emphasis on consistency in Hilbert's programme to certain extent, something that had not been explained so far. For example, in one of his papers he writes[13]:

> What matters for the question of pure mathematics is only whether the usual, axiomatically characterised mathematical continuum is possible in itself, that it is a consistent creation.

For Hilbert, this was probably a truism, as he commented in the following way:[14]

> If the arbitrarily given axioms do not contradict each other through their consequences, then they are true, then the objects defined through the axioms exist. That, for me, is the criterion for truth and existence.

[11] Translation from [Smorynski, 1986], p. 18

[12] In the translation in [Smorynski, 1986] it said controllability, but I adapted this translation, because in the original text,[Hilbert, 1917], it was Kontrollierbarkeit, for which I think verifiability is a better translation.

[13] [Smorynski, 1986] p. 24

[14] From [Smorynski, 1977], p. 825, citing Meschkowski, citing Hilbert.

What neither he nor Bernays realised at the time, was that Brouwer seemed to return to the Aristotelian approach in some way: he was after truth, not mere consistency, something well illustrated by a remark in 1927 by Brouwer[15]:

> nothing of mathematical value will thus be gained: an incorrect theory, even if it cannot be inhibited by any contradiction that would refute it, is none the less incorrect, just as a criminal policy is none the less criminal, even if it cannot be inhibited by a court that would curb it.

So, the very goal of the programme wouldn't convince Brouwer, even when achieved.

In the 1921 version of his programme, Hilbert made the distinction between mathematics and *meta*mathematics. Mathematics is what we all know as mathematics: analysis, set theory, etc. It is abstract, infinitary and has no empirical meaning. Consistency is enough to guarantee its validity. Metamathematics is the direct study of signs and their combinations. It is intuitive and contentual. Hilbert wanted the formalisation of mathematics, with precisely defined axioms and rules of inference, and the consistency of this system proven in metamathematics.

The distinction between mathematics and metamathematics was first made by Brouwer in 1907[16], and Hilbert picked it up as it suited his purposes well: Brouwer would have to accept his consistency proof, as he accepted metamathematical reasoning. Also he hoped other objections[17] against his theory could

[15] Quoted in [Dalen, 2013], p. 442

[16] The distinction was made in his dissertation, [Brouwer, 1907], for example here (p. 98):

> Eigenlijk is het gebouw der intuitieve wiskunde zonder meer een daad, en geen wetenschap; een wetenschap, d.w.z. een samenvatting van in den tijd herhaalbare causale volgreeksen, wordt zij eerst in de wiskunde der tweede orde, die het wiskundig bekijken van de wiskunde of de taal der wiskunde is: eerst daar bestaat een causaal verband in de wijze van opvolging der wiskundige systemen eenerzijds, en der wiskundige teekens, woorden of begrippen andererzijds; maar daar, evenals bij de theoretische logica, hebben we ook weer te doen met een toepassing der wiskunde, met een ervaringswetenschap. Men vergelijke in dit verband de ontwikkelingen van het derde hoofdstuk.

and here(pp. 132-133):

> We leggen er verder den nadruk op, dat het syllogisme en de verder logische principes kunnen worden gerekend te gelden voor de taal der logische redeneeringen, die handelen over eindige elementengroepen, of aftelbaar oneindige, of gebieden binnen continua, maar in elk geval uitsluitend wiskundig opgebouwde systemen; de overtuiging van de betrouwbaarheid hunner toepassing steunt op de zekerheid, dat het wiskundig opbouwbare systemen zijn, waarover wordt gesproken. En wanneer het gelukt taalgebouwen op te trekken, reeksen van volzinnen, die volgens de wetten der logica op elkaar volgen, uitgaande van taalbeelden, die voor werkelijke wiskundige gebouwen, wiskundige grondwaarheden zouden kunnen accompagneeren, en het blijkt dat die taalgebouwen nooit het taalbeeld van een contradictie zullen kunnen vertoonen, dan zijn ze toch alleen wiskunde als taalgebouw en hebben met wiskunde buiten dat gebouw, bijv. met de gewone rekenkunde of meetkunde niets te maken.
>
> Dus in geen geval mag men denken, door middel van die taalgebouwen iets van andere wiskunde, dan die direct intuitief op te bouwen is, te kunnen te weten te komen. En nog veel minder mag men meenen, op die manier de grondslagen der wiskunde te kunnen leggen, m.a.w. de betrouwbaarheid der wiskundige eigenschappen te kunnen verzekeren.

[17] Poincaré had accused Hilbert of using induction to prove induction.

be sidestepped this way.

In 1922, Hilbert presented a formal system of metamathematical arithmetic. Also he sketched a proof of the consistency of a fragment thereof. This proof sketch shows how Hilbert thought his programme should be carried out. Indeed, his student Ackermann would almost succeed in carrying through the programme, using this proof sketch as guideline.

This proof sketch, and the explanation of later improvements of the programme can be found in section 2.3

2.2 Finitism

In his new version of the programme, Hilbert introduced a new word: 'finit', from the Latin word for finite. He used this word to differentiate between finite and finitistic: a set consisting of two infinite elements is finite (as it only has two elements) but one would not want to call that set finitistic, as its elements are infinite.

Actually Hilbert only adopted the already existing notion of finitistic mathematics. Finitistic mathematics had been developed by Kronecker, Skolem and others, after recognising and criticising the use of abstract objects in the intuitionstic mathematics of Brouwer. Their foundational viewpoint was that a mathematical definition is genuine if and only if it leads to the goal by a *finite* number of trials. Only concrete combinatorial operations on finite mathematical objects are allowed.

Hilbert considered finitistic mathematics as the truly meaningful part of mathematics. Clearly, finite mathematics is constructive and therefore, the Law of the Excluded Middle, for example, can be used without problems.

According to Hilbert, problems arise when we apply procedures, reliable in the finite case, to the infinite. For example, when using quantifiers $\exists$ and $\forall$. These are, according to Hilbert, just abbreviations of infinite conjunctions and disjunctions. For example, $\forall v \varphi(v)$ is the abbreviation of $\varphi(0) \wedge \varphi(1) \wedge \varphi(2) \wedge \ldots$. Expressions like this had no precise negations. Later on, in 1925, Hilbert would argue that only the existential quantifier was infinitistic, and that the universal quantifier could be used without restraint.

One can say that, according to Hilbert, by using quantifiers, an infinitistic element is introduced into logic. Nevertheless, Hilbert would add formulae containing quantifiers to his formal system of finitary arithmetic and apply the Law of the Excluded Middle to them. He would call these axioms and formulae *transfinite*, roughly meaning that the cardinality is comparable with that of the natural numbers. Hilbert claimed that the addition of these transfinite elements was not necessary and just done for simplification:

> To be sure one can presumably prove a finitistic statement also without application of transfinite means of proof ... but this claim is of the sort of the claim that in general every mathematical assertion must allow itself either to be verified or refuted. [18]

We already noted that finite mathematics is constructive and hence a part of intuitionistic mathematics. On the other hand, intuitionistic mathematics goes beyond finite mathematics, as became clear when Gödel showed that classical

[18] Smorynski citing Hilbert in [Smorynski, 1986], p. 28

arithmetic, which clearly goes beyond finite arithmetic, could be reduced to intuitionistic mathematics, via his 'negative translation'[19]. But this happened in 1933, which means by that time also his incompleteness theorems had been presented, and the programme, in this form, already had been refuted.

2.3 Explanation of the Programme

An outline of Hilbert's approach to the proof of the consistency of the transfinite.

Because of the following equivalences, Hilbert's logic only included the two connectives ¬ and →:

$$\varphi \vee \psi \Leftrightarrow \neg\varphi \to \psi \qquad\qquad \varphi \wedge \psi \Leftrightarrow \neg(\varphi \to \neg\psi) \tag{2}$$

In his proof, variables are used for formulae as well as for numbers, and the elements of the 'language' are the constant 0, the successor function $(\cdot) + 1$ and the predecessor function δ. The only rule of inference is

$$\varphi \text{ and } \varphi \to \psi \text{ infer } \psi \tag{3}$$

This is called *modus ponens.*

The axioms are as follows:

1. $A \to (B \to A)$
2. $(A \to (A \to B)) \to (A \to B)$
3. $(A \to (B \to C)) \to (B \to (A \to C))$
4. $(B \to C) \to ((A \to B) \to (A \to C))$
5. $A \to (\neg A \to B)$
6. $(A \to B) \to ((\neg A \to B) \to B)$
7. $a = a$
8. $a = b \to (A(a) \to A(b))$
9. $\neg(a + 1 = 0)$
10. $\delta(a + 1) = a$

There were two things that Hilbert overlooked: the axiom $\neg a = 0 \to \delta(a) + 1 = a$, and a value for $\delta(0)$. Even after adding such additional axioms, the system is still at best a fragment of finitistic arithmetic. Hilbert said that one should add 'recursion and intuitive induction' to obtain the full metamathematical arithmetic.[20] In the system, so far no quantifiers are allowed, which means that induction would have to be added as an inference rule:

$$\text{from } A(0) \text{ and } A(a) \to A(a + 1) \text{ infer } A(b) \tag{4}$$

[19] See [Troelstra and van Dalen, 1988], p.25

[20] From [Smorynski, 1986], p. 29

At that time, Hilbert's conception of recursion was, just as that of all his contemporaries, only partially formed and it's hard to say what exactly he meant by recursion and intuitive induction. Now we know that the form of recursion allowed should be primitive recursion (this will be explained in section 5).

Hilbert has given a proof-sketch of consistency of the theory given above, which we name $\mathfrak{T}_0$, which will be presented here. First it is necessary to define exactly what a formal derivation is.

Definition 1. A formal derivation of a formula φ is a sequence of formulae $\varphi_0, \varphi_1, \ldots, \varphi_n = \varphi$ such that each φ_i is an instance of an axiom or follows from two earlier formulae φ_j and φ_k $(j, k < i)$ by modus ponens.[21]

As Smorynski states in [Smorynski, 1986]: One can show that axioms 1-6 are complete with respect to purely propositional reasoning. Whence, the consistency of $\mathfrak{T}_0$ reduces to the unprovability of $\neg(0 = 0)$: if the system would be inconsistent, anything could be proved. So if there is something which cannot be proved, the system is consistent.

Theorem 2. The system $\mathfrak{T}_0$ is consistent.

A reproduction of Hilbert's proof, given in 1921, is given by Smorynski in [Smorynski, 1986]:

> *Proof.* Suppose $D = \varphi_0, \varphi_1, \ldots, \varphi_n$ is a formal derivation of $\neg 0 = 0$. As a first step, we can modify D by throwing away any φ_i (other than φ_n) which is not used as a premise of an application of modus ponens. Moreover, by repetition of formulae, we can assume each φ_i occurs only once as such a premise. Call the resulting derivation D'.
>
> Second, we can omit number variables from D' by substituting, say, 0 for each occurrence of a variable in D'. Call the result D''. Third, we can simplify the terms to the point that each formula is a propositional combination of equations of *numerals*,
>
> $$0, 0+1, 0+1+1, \ldots \tag{5}$$
>
> Fourth, every formula can be brought into a logical normal form (of some sort - Hilbert doesn't say which kind).
>
> Each formula of the derivation is now subject to a control[22], i.e. one can check each formula for 'correctness' or 'falsity'. But it can be shown that each formula of this final derivation is correct, whence the end formula $\neg 0 = 0$ is correct - a contradiction. Thus, there was no derivation D of $\neg 0 = 0$. □

Step two guarantees the finitary nature of the derivation, as it reduces every term into a computable term capable of being reduced to a numeral. (This is only true if one adds the additional axioms given.)

One possible interpretation[23] of step three could be, as stated by Smorynski in [Smorynski, 1986]:

[21] From [Smorynski, 1986], p. 30

[22] Originally in [Smorynski, 1986] it was control, but as before I think verifiability is a better translation of the original text by Hilbert.

[23] For another possible interpretation, see [Smorynski, 1986] pp. 31-32

> Calculate the value of each term t occurring in the given derivation D'' and replace t in each occurrence by the numeral for this value. Every formula is now a propositional combination of equations of numerals. The 'correctness' or falsity of each equation is simply a matter of comparison, and the computation of truth values of propositional combinations is a simple matter.
>
> If this indeed what Hilbert had in mind, then 'bringing a formula into a logical normal form' refer to a simple calculation of the truth value of a propositional combination of sentences once the individual truth values are known.

Hilbert extended his proof with the aim of obtaining the consistency of *transfinite arithmetic*, i.e. arithmetic with quantifiers. [24]

2.3.1 The Introduction of the τ-function

First, Hilbert introduced *choice functions.* Given a formula $\varphi(a)$ with free variable a, Hilbert added a new term $\tau_a(\varphi)$, which was an attempt to find a counterexample, and together these satisfied the transfinite axiom

$$\varphi(\tau_a(\varphi)) \to \varphi(a) \tag{6}$$

This axiom can intuitively be explained as follows. If $\varphi(a)$ holds, then the implication is always true. If $\varphi(a)$ does not hold, then $\tau_a(\varphi)$ is a 'counterexample' of φ: for this 'value' φ does not hold. Then $\varphi(\tau_a(\varphi))$ does not hold, and the implication is true.

Formally, the axiom was

11. $A(\tau(A)) \to A(a)$

As we see A is a variable for formulae, so it has no numerical variable a occurring in it. Thus, $\tau(A)$ has no subscript indicating which variable is being bound. Nonetheless, τ_a does bind variables, and problems may arise with substitution: we will return to binding variables and substitution in section 4.

Now, quantifiers could be introduced as abbreviations, using τ:

$$\forall a \varphi(a) : \varphi(\tau_a(\varphi)) \qquad\qquad \exists a \varphi(a) : \varphi(\tau_a(\neg\varphi)) \tag{7}$$

The first one can be explained using the axiom that was introduced: this just implies $\varphi(a)$, for all a. The second one is φ of a counterexample of $\neg\varphi$, which is an example of φ.

Once he had these, it is possible to derive the usual laws for quantifiers:

$$\begin{aligned} \forall a \varphi(a) \to \varphi(a),\ \neg\forall a \varphi(a) \Leftrightarrow \exists a \neg\varphi(a) \\ \varphi(a) \to \exists a \varphi(a),\ \neg\exists a \varphi(a) \Leftrightarrow \forall a \neg\varphi(a) \end{aligned} \tag{8}$$

Hilbert noted that there were difficulties with the nestings of τ's (for every formula containing a τ-function, a new τ-function should be introduced to complete the system, but then again new functions should be introduced, etc.), but he indicated how the proof of theorem 2 could be extended to cover a special instance of this new axiom.

[24] (not in quotation) This is what we would call ordinary set theory (ZF) now.

Define, with f a function variable:

$$\tau(f) = \tau_a(f(a) = 0). \tag{9}$$

This means that the corresponding instance of the newly added axiom becomes[25]

12. $f(\tau(f)) = 0 \to f(a) = 0$

To simplify matters, we define a new axiom in which we only allow one fixed function F which is obtained by recursions, and which is assumed unary. The new axiom becomes:

13. $F(\tau(F)) = 0 \to F(a) = 0$

The resulting system with axioms 1-10 and 13 and the recursion equations needed to compute F, will be called $\mathfrak{T}_1$.

Theorem 3. The system $\mathfrak{T}_1$ is consistent.

A translation of the proof is given by Smorynski in [Smorynski, 1986]:

> *Proof sketch.* Following the steps of the proof of theorem 2 we can transform any derivation D into a derivation D' which consists solely of propositional combinations of equations involving numerals and the symbol $\tau(F)$. (This is in fact not true, but this is what Hilbert claimed.) We now attempt a control[26] of the proof by assigning $\tau(F)$ the value 0. That is, we replace all occurrences of $\tau(F)$ in D' by the numeral 0 to obtain a new 'derivation' D''. We would like to show that all sentences in D'' are correct. Those that come from axioms 1-10 or the recursion equations defining F are correct, and modus ponens preserves correctness. So any false sentences must have been introduced by 13:
>
> $$F(0) = 0 \to F(z) = 0 \tag{10}$$
>
> for some numerals z. If all of these are correct, then every sentence in D'' is correct and the original D could not have derived $\neg 0 = 0$ as this formula is false and is left unchanged by the proof transformations.
>
> If, one the other hand, the instance,
>
> $$F(0) = 0 \to F(z) = 0 \tag{11}$$
>
> is incorrect, we simply go back to D' and replace all occurrences of $\tau(F)$ by the numeral z. The instances of 13 now become instances like
>
> $$F(z) = 0 \to F(x) = 0 \tag{12}$$
>
> for various numerals x. But these implications are correct formulae, because $F(z) = 0$ is false! Again, all sentences in D'' are correct, whence D could not have been a derivation of $\neg 0 = 0$
>
> □

[25] The function $\tau(f)$ wouldn't be accepted by Brouwer, as Hilbert noted. He gave a sketch of the consistency proof for theorem 2 and following this he gave a few examples of what this consistency proof would yield that Brouwer wouldn't allow. For these examples see [Smorynski, 1986] p. 33

[26] I think this should be verification, as explained before.

In the lecture where Hilbert presented this proof, he said as well that 'one had but to carry through the details of his proof sketch in order to complete the laying foundations for analysis, and therewith begin the corresponding work on set theory.'[27] Hilbert was aware of the fact that this wouldn't be without difficulties. What the extent of these difficulties would be, we will see in later sections.

2.4 Adjustments to the First Version

In 1925, Hilbert lectured again and now made a distinction between *real propositions*, *finitary general propositions*, and *ideal propositions*, instead of the old dichotomy between finitary propositions and transfinite formulae. [28]

Smorynski describes the difference between these kinds of propositions, as explained by Hilbert, as follows[29]:

> The real propositions are, Hilbert says, of 'no essential interest' in themselves. They are simple propositional combinations of equations involving primitive recursive functions and fixed numerals:
>
> $$2+3=3+2, 1+1=2, \text{etc.} \qquad (13)$$
>
> They are directly contentual assertions verifiable by direct computation. Their importance lies primarily in affording a control on the results of formal mathematical proofs. [...]
>
> Real propositions do not exhaust the class of finitistic propositions. There are also what I shall call here finitary general propositions - assertions of the form 'for every numeral n, $n+1=1+n$' [...]
>
> Finally ther are the ideal propositions are not really propositions at all, but formal symbols manipulated according to pre-determined rules [...]
>
> In a lecture in December 1930,[...] Hilbert would point out the difference between a finitary general proposition, e.g.
>
> $$1+x=x+1 \qquad (14)$$
>
> and the similar ideal proposition,
>
> $$\forall x(1+x=x+1). \qquad (15)$$
>
> The former is the assertion that
>
> $$n+1=1+n \qquad (16)$$
>
> for all numerals n; the latter also asserts the equation for meaningless infinitary constructs involving the τ-function.[30] [...] Hilbert said

[27] [Smorynski, 1986] p. 34

[28] This new distincion as stated by Smorynski ([Smorynski, 1986]) is being criticised as a misunderstanding of Hilbert's writings, by e.g. Detlefsen. For a detailed comparison of the German texts and translations given by Smorynski, and Detlefsen's conclusions on this subject, see [Detlefsen, 1990], pp. 347-357.

[29] [Smorynski, 1986] pp. 39-40

[30] Actually it was the ε-function instead of the τ-function. The ε-function will be introduced in section 2.4.1

> the following: Let $\mathfrak{D}$ be a formal system of finitary arithmetic and let $\mathfrak{T}$ be some system of transfinite mathematics. Suppose $\mathfrak{D}$ proves the consistency of $\mathfrak{T}$. Then: For any universal assertion φ, if $\mathfrak{T} \vdash \varphi$ then $\mathfrak{D} \vdash \varphi$.

Hilbert is talking only about the finitary general propositions here. He had shown already before that for the real propositions, the conservation resulted from verifiability of the results of mathematical derivations. The verifiability of these mathematical derivations followed from the method Hilbert used in his consistency proof.

What he was doing now, is settling the question of the finitistic derivability of any finitistic statement obtained via transfinite means of proof.

He proved his method by means of an example, Fermat's theorem:

$$\mathrm{FT}: \forall xyzw(x > 1 \land y > 1 \land z > 1 \land w > 2 \to x^w + y^w \neq z^w) \tag{17}$$

The idea of the proof is this. We assume FT holds in a transfinite system, which we assume to be consistent and conservative, regarding elementary arithmetic sentences, over the finitistic system. Now we want to see it is finitistically derivable. So we reason finitistically and use the fact that transfinitely we know FT holds to derive a contradiction, which will lead to the finitistic statement of FT.

The proof of FT is as follows:

Proof. Suppose we are given numerals k, m, n, p such that

$$k > 1 \land m > 1 \land n > 1 \land p > 2 \land k^p + m^p = n^p. \tag{18}$$

This is a finitistic statement: it can be verified by a simple computation. This translates directly into a proof in the transfinite system. On the other hand, we know that FT is provable in the transfinite system (because we know FT is provable if we don't have the restriction of finitistic derivability). This means, by the provability of FT within the transfinite system, that FT also holds when applied to k, m, n, p:

$$k > 1 \land m > 1 \land n > 1 \land p > 2 \to k^p + m^p \neq n^p. \tag{19}$$

So now we have

$$k^p + m^p \neq n^p \qquad\qquad k^p + m^p = n^p \tag{20}$$

which are both provable transfinitely. (The first one is our assumption, the second one follows from the transfinite provability of FT.) But the transfinite system is consistent, so this means 18 is false and we have proven that for any k, m, n, p 19 holds, so we have proven FT in our finitistic system, using transfinite means of proof. □

Still no proof of the consistency of transfinite mathematics had been given. Hilbert had sketched his proof of consistency for the limited case of his transfinite axiom with the τ-function in 1922. The τ-function was later replaced by a choice function ε.

2.4.1 The ε-function

The ε-function satisfied the following additional axiom:[31]

$$A(t) \to A(\varepsilon_a(A(a))). \tag{21}$$

where t is an arbitrary term. With the ε-function it is possible to define the quantifiers, as follows:

$$\exists x A(x) \equiv A(\varepsilon_x A(x)) \qquad \forall x A(x) \equiv A(\varepsilon_x \neg A(x)) \tag{22}$$

The idea of the ε-function is as follows: The ε-function stands for a witness of the formula $A(x)$, intuitively seen. The ε-terms $\varepsilon_x A(x)$ that occur in a formal proof can be replaced by numerals, which turns the proof into a quantifier-free proof.

An example: Suppose a derivation of $\neg(0 = 0)$ exists, and that it contains only one ε-term $\varepsilon_x A(x)$. Now all occurrences of $\varepsilon_x A(x)$ should be replaced by 0. The axiom added, 21, now only has instances of the form $A(t) \to A(0)$. We assumed that no other ε-terms occur in the proof, this means that $A(t)$ and $A(0)$ are basic numerical formulae without quantifiers. Also we may assume, to simplify matters, they are without free variables. This means they can be evaluated by finitary calculation. Now all instances $A(t) \to A(0)$ could be true, then we are done. If one of them is false, then we know $A(0)$ is false while $A(t)$ is true for some t. Now we replace $\varepsilon_x A(x)$ by n instead of 0, where n is the numerical value of the term t. Then all statements are true: $A(t)$ was false for all other terms, so then the implication holds, and for t we now have $A(t) \to A(t)$ which is of course true. The proof obtained is a derivation of $\neg(0 = 0)$ from true, purely finitistic, numerical formulae using only modus ponens. This is impossible, from which we can derive that there doesn't exist a derivation of $\neg(0 = 0)$.

We can describe the function of the ε-function by the following two conservativity results, proved by Bernays in the second volume of the *Grundlagen* in 1939.[32] Let T be a finitely axiomatised theory containing no quantifiers of ε-terms, and let PC_ε be a usual formulation of the predicate calculus, extended by the ε-operator and its characteristic axiom as stated above. The first ε-theorem states that for any formula without quantifiers or ε-terms provable in PC_ε from T is already provable from T in the quantifier- and ε-free fragment EC of PC. (EC is the so-called elementary calculus of free variables.) The second ε-theorem states that any formula without ε-terms provable in PC_ε from T, is also provable from T in PC, the pure predicate calculus (without ε-operator).

The difference between the ε-function and the τ-function could intuitively be explained as follows: the τ-function yields a counterexample to a certain formula, while the ε-function yields a 'witness', an example, of a certain formula.

Ackermann published a consistency proof in 1924, using this ε-function, but discovered an error later. Johann von Neumann criticised Ackermann's improved version of his proof, and also published a consistency proof, in 1927. However, Hilbert discussed Ackermann's improved proof in a lecture, and the published version was accompanied by a note by Bernays explaining the proof

[31] From Moser and Zach [2005] p. 1

[32] From Moser and Zach [2005] p. 2, p. 11 and p. 19

in more detail. This all happened in 1928, and the Hilbert school believed the proof was almost complete.

2.5 The Four Problems Posed by Hilbert

In 1928, there were four problems posed by Hilbert. The first one consisted of extending the consistency proof given by Ackermann and von Neumann of the arithmetical of the integers, to cover not only arithmetic formulae, but also function variables. (In fact, Ackermann and von Neumann had not proved the consistency for arithmetical formulae yet.) The second problem consisted of giving a consistency proof for a system in which more advanced parts of analysis could be carried out, as well as some set theory. The third problem was stated as follows: 'If one can proof the consistency of φ with the axioms of number theory, then one cannot prove such consistency for $\neg\varphi$.'[33]

This statement is based on the following reasoning: if φ and $\neg\varphi$ were both consistent, then either one could be added as an axiom to the axioms of number theory to obtain two non-isomorphic systems of arithmetic. However, Dedekind showed[34] that any two models of the Peano axioms[35] are isomorphic. A year later, Gödel would prove this in his Completeness theorem (23), according to this theorem if both φ and $\neg\varphi$ would be consistent, there would be a proof for both of them, which results in a contradiction.

The fourth problem a statement of completeness that is more familiar to us nowadays: 'If φ is not provable from the axioms of arithmetic, then adding φ as an axiom yields a contradiction.'[36]

Only two years later Gödel would present his famous theorems, refuting Hilbert's programme as formulated at that time, but we will return to this later, in section 8. The question whether or not this meant the end of Hilbert's programme, will be discussed in section 9.

[33] Citing Smorynski in [Smorynski, 1986], p. 49

[34] See [Dedekind, 1901]

[35] These are *not* the Peano axioms that will be introduced in the next section, these are second-order axioms, and the system introduced later will be of first-order. This distinction is important because the first-order Peano system *does* have non-isomorphic models.

[36] Citing Smorynski in [Smorynski, 1986], p. 49

3 Peano Arithmetic

PA, or *Peano Arithmetic*, is a system of first-order arithmetic. It is a theory using the language $\mathcal{L}_{PA} = \{0, 1, +, \cdot\}$. In this language $0, 1$ are the constants, and $+, \cdot$ are binary function symbols. At the time PA was developed, the formal notions of e.g. language, function symbols and theories were not in use yet. We will see in section 4 what the formal definition of these notions is.

PA has the following axioms:

1. $\forall x \neg(x + 1 = 0)$
2. $\forall xy(x + 1 = y + 1 \to x = y)$
3. $\forall x(x + 0 = x)$
4. $\forall xy(x + (y + 1) = (x + y) + 1)$
5. $\forall x(x \cdot 0 = 0)$
6. $\forall xy(x \cdot (y + 1) = (x \cdot y) + x)$
7. $\forall \vec{x}[(\varphi(0, \vec{x}) \wedge \forall y(\varphi(y, \vec{x}) \to \varphi(y + 1, \vec{x}))) \to \forall y \varphi(y, \vec{x})]$

These axioms speak for themselves, only the last one needs some explanation. The last axiom is actually an axiom for every possible formula $\varphi(y, \vec{x})$. The axioms are called *induction axioms.* This means there are infinitely many axioms, as there are infinitely many formulae. It can be shown that there are no finite $\mathcal{L}_{PA}$-theories which have the same models as PA, so it is necessary to have infinitely many axioms for PA.

Clearly, PA can be applied to the natural numbers, $\mathbb{N}$. In fact $\mathbb{N}$ is a *model* of PA, a notion that will be explained in section 4. In PA one can carry out most elementary number theory, as we will see in section 3.1. Also PA can represent all *recursive functions*, as was proved by Gödel and will be used to prove his incompleteness theorems. The concept of recursive functions will be explained in section 5.

3.1 Basic Properties of PA

In PA, a lot of elementary number theory can be carried out. We will give some examples and some definitions we will need later on. Also some theorems and lemmas will be stated, some of which will be proved. For the proof of the rest of the statements I refer to [van Oosten, 2009].

With the last axiom, the induction axiom, the basic rules for addition and multiplication can be proved:

1. PA $\vdash \forall x(x = 0 \vee \exists y(x = y + 1))$
2. PA $\vdash \forall xyz(x + (y + z) = (x + y) + z)$
3. PA $\vdash \forall xy(x + y = y + x)$
4. PA $\vdash \forall xyz(x + z = y + z \to x = y)$
5. PA $\vdash \forall xyz(x \cdot (y \cdot z) = (x \cdot y) \cdot z)$

6. $\mathrm{PA} \vdash \forall xy(x \cdot y = y \cdot x)$
7. $\mathrm{PA} \vdash \forall xyz(x \cdot (y + z) = (x \cdot y) + (x \cdot z)$
8. $\mathrm{PA} \vdash \forall xyz(\neg(z = 0) \wedge x \cdot z = y \cdot z \rightarrow x = y)$

Also the symbol $<$, with its usual meaning, can be defined formally in PA. After we have defined it, we will use it as a primitive symbol of PA, but remember that in fact it is just an abbrevation.

Lemma 4 (Least Number Principle)**.** Let $\varphi(x, y)$ be the formula $\exists z(x+(z+1) = y)$. In PA, φ defines a discrete linear order with least element, and which satisfies the least number principle:

$$\mathrm{PA} \vdash \exists w\psi(w) \rightarrow \exists y(\psi(y) \wedge \forall x(\varphi(x, y) \rightarrow \neg\psi(x))) \tag{23}$$

We write $x < y$ for $\varphi(x, y)$.

The least number is actually the formal statement of the following: for any formula ψ, if there is a w such that $\psi(w)$, then there is a smallest number y for which $\psi(y)$: for all $x < y$, we have $\neg\psi(x)$.

Theorem 5 (Division with remainder)**.** PA proves the following, in which a, b are unique:

$$\mathrm{PA} \vdash \forall xy(y \neq 0 \rightarrow \exists ab(x = a \cdot y + b \wedge 0 \leq b < y)) \tag{24}$$

Proof. We prove this by induction on x. For $x = 0$ we have $0 = 0 \cdot y + 0 \wedge 0 \leq b < y))$, for any $y > 0$.

Suppose we have $x = a \cdot y + b \wedge 0 \leq b < y)$. First we prove that such a, b exist for $x + 1$. Because $<$ is a discrete linear order we have $b + 1 < y \vee b + 1 = y$.

If $b + 1 < y$, then we have $x + 1 = a \cdot y + (b + 1)$ and we are done.

If $b + 1 = y$, then $x + 1 = a \cdot y + (b + 1) = a \cdot y + y = (a + 1) \cdot y + 0$ and we are done.

Now we still have to prove that a, b are unique. Suppose we have $x = a \cdot y + b$ and $x = a' \cdot y + b'$, with $0 \leq b, b' < y$. Suppose $a < a'$, then $a + 1 \leq a'$ so we have $a' \cdot y \geq a \cdot y + y > a \cdot y + b = x$, which is a contradiction. So $a' \leq a$. By symmetry we see $a \leq a'$ so $a = a'$. So we have $x = a \cdot y + b$ and $x = a \cdot y + b'$, by one of the properties of addition stated above we see that also $b = b'$. □

Once we have this important result, it is possible to define formulae expressing properties of numbers, below we have some examples:

- $x|y \equiv \exists z(x \cdot z = y)$
- $\mathrm{irred}(x) \equiv \forall v \leq x(v|x \rightarrow v = 1 \vee v = x)$
- $\mathrm{prime}(x) \equiv x > 1 \wedge \forall yz(x|(y \cdot z) \rightarrow x|y \vee x|z)$
- $\mathrm{pow}(x, v) \equiv x \geq 1 \wedge \mathrm{prime}(v) \wedge \forall w \leq x(w > 1 \wedge w|x \rightarrow v|w)$
- $\mathrm{pp}(x) \equiv \exists v \leq x \ \mathrm{pow}(x, v)$

For these, one can show the expected results, for example PA $\vdash \text{prime}(x) \leftrightarrow \text{irred}(x)$ and unique prime factorisation in PA. Also we can denote $\text{rm}(x|y) = b$, if $x = a \cdot y + b \wedge 0 \leq b < y$.

We can also define lcm(least common multiple) and gcd(greatest common divisor). From the least number principle we know that, as $x|x \cdot y$ and $y|x \cdot y$, there is a smallest w such that $x|w \wedge y|w$. We denote this w by $\text{lcm}(x, y)$. Now we can write, by theorem 5, $x \cdot y = a \cdot \text{lcm}(x, y) + b$, where $0 \leq b \leq \text{lcm}(x, y)$. We know $x|x \cdot y \wedge y|x \cdot y$, $x|\text{lcm}(x, y) \wedge y|\text{lcm}(x, y)$, so we have $x|b \wedge y|b$. If $b > 0$ we would have found a smaller common multiple of x and y, which is a contradiction, so $b = 0$ and we have $x \cdot y = a \cdot \text{lcm}(x, y)$ for a unique a. We denote this a by $\text{gcd}(x, y)$. Now in PA we can prove their basic properties, for example:

- PA $\vdash \forall xyu(x, y \geq 1 \wedge u|x \wedge u|y \rightarrow u|\text{gcd}(x, y))$
- PA $\vdash \forall xyab(y = a \cdot x + b \wedge 0 \leq b < x \rightarrow \text{gcd}(x, y) = \text{gcd}(x, b))$

I will not state or prove all the properties of gcd, for a more complete overview see [van Oosten, 2009].

3.2 More elementary number theory

Later on, we will need some more advanced results, that will be proved in this section. To be able to prove those results, we will need *Bézout's theorem for PA* which we will present here. Then we will prove the results that we have used, that are also very important for Gödel's *coding of sequences*, as we will see when we return to this in section 7.2. The sequences defined in this section will be used to prove that primitive recursive functions can be represented in PA, in particular the primitive recursive coding of sequences.

Theorem 6 (Bézout's Theorem for PA)**.**

$$\text{PA} \vdash \forall xy \geq 1 \exists a \leq y, b \leq x(a \cdot x = b \cdot y + \text{gcd}(x, y)) \tag{25}$$

Proof. We proof this theorem by induction on x, using properties we have shown in section 3.1. For $x = 1$ take $a = 1$ and $b = 0$. Then for any y, $1 \cdot 1 = 0 \cdot y + \text{gcd}(1, y)$, as $\text{gcd}(1, y) = 1$ for any y. Now assume that for any $x' < x$ we know PA $\vdash \forall x'y \geq 1 \exists a \leq y, b \leq x'(a \cdot x' = b \cdot y + \text{gcd}(x', y))$.

We know that for y we can write $y = c \cdot x + d$, where $0 \leq d < x$. We divide this equation by $\text{gcd}(x, y)$ to obtain $y' = c \cdot x' + d'$, where $d' < x' \leq x$. $d' < x'$ because $d < x$ and $x' < x$ because $\text{gcd}(x, y) \geq 1$. Also we know $\text{gcd}(x', d') = 1$. For if $\text{gcd}(x', d') = a$ with $a > 1$, then we can write $x = a \cdot x'' \cdot \text{gcd}(x, y)$ and $y = a(c \cdot x'' \cdot \text{gcd}(x, y) + d'')$, where $x'' = a \cdot x'$ and $d'' = a \cdot d'$. Which means a divides both x and y, which is a contradiction because x' and d' already have been divided by $\text{gcd}(x, y)$ so they don't contain any common divisors of x and y.

Because $d' < x' < x$ we can apply the induction hypothesis, and using $\text{gcd}(x', d')$ we obtain

$$u \cdot d' = v \cdot x' + 1 \tag{26}$$

which means we have $v \cdot x' = u \cdot d' - 1$. Squaring both sides gives $(vvx') \cdot x' = (uud' - 2u) \cdot d' + 1$, renaming $vvx' = a'$ and $uud' - 2u = b'$ and multiplying by $\gcd(x, y)$ gives:

$$a' \cdot x = b' \cdot d + \gcd(x, y) \tag{27}$$

Now substituting $d = y - cx$ gives:

$$(a' + c \cdot b') \cdot x = b'y + \gcd(x, y) \tag{28}$$

Now let $(a' + c \cdot b') = c' \cdot y + a''$, where $0 \leq a'' < y$ for the same reasons as before. Substituting this gives:

$$a'' \cdot x = (b' - c' \cdot x) \cdot y + \gcd(x, y) \tag{29}$$

We have already seen that $a'' < y$. Because of that, we see that $(b' - c' \cdot x) \cdot y \leq a'' \cdot x < y \cdot x = x \cdot y$. This means that $(b' - c' \cdot x) < x$. So a'' and $(b' - c' \cdot x)$ are examples of a, b for which $a \leq y, b \leq x(a \cdot x = b \cdot y + \gcd(x, y))$ holds. □

Theorem 7. Given a sequence of numbers $x_0, \ldots, x_{n-1}$, define:

$$\begin{aligned} m &= \max(x_0, \ldots, x_{n-1}, n)! \\ a &\equiv x_i \bmod m(i + 1) + 1 \end{aligned} \tag{30}$$

This system always has a solution and the pair (a, m) is said to *code* the sequence $x_0, \ldots, x_{n-1}$.

We use the following abbreviations: $\mathrm{rm}(x, y)$ denotes the remainder of x on division by y. $(a, m)_i$ denotes $\mathrm{rm}(a, m \cdot (i + 1) + 1)$.

Proof. The fact that a is uniquely defined, and is the same for every i, follows from the *Chinese remainder theorem.*

Theorem 8 (Chinese Remainder Theorem[37]). Suppose we are given $m = m_1 \cdots m_n$ with $\gcd(m_i, m_j) = 1$ for $i \neq j$. Let $b_1, \ldots, b_n$ be integers and consider the system of congruences:

$$\begin{aligned} x \equiv& b_1 (\text{mod } m_1) \\ &\vdots \\ x \equiv& b_n (\text{mod } m_n) \end{aligned} \tag{31}$$

This system always has solutions and solutions differ by a multiple of m.

First we see that indeed $m(i+1)+1$ and $m(j+1)+1$ are relatively prime for all $0 \leq i < j < n$. If there would be a prime number p dividing both $m(i+1)+1$ and $m(j+1)+1$, then it would divide their difference: $m(i-j)$. Because p is prime, we know that $p|ab \Rightarrow p|a \vee p|b$. So $p|m$ or $p|(j-i)$. Because $m = x!$ for $x \geq n$, we know that $(j-i)$ is a factor of m, because $(j-i) < n$. So if $p|(j-i)$ then $p|m$. We conclude that $p|m$. But also we assumed $p|m(i+1)+1$, which, together with $p|m$, is a contradiction because $p \nmid 1$. So indeed $m(i+1)+1$ and $m(j+1)+1$ are relatively prime for all $0 \leq i < j < n$.

[37] [Ireland and Rosen, 1990] p. 34

So we may apply the Chinese remainder theorem, which gives us the a from the definition. And, trivially, $x_i < (i+1)m+1$ for all i. This means that $(a,m)_i = x_i$ for all i. So there is only one sequence $x_0, \ldots, x_{n-1}$ which meets the requirements for m and a, and from (a,m) the original sequence can be deduced: this is why (a,m) *codes* the sequence. □

This coding of sequences was actually called the β-function by Gödel. The following theorem expresses some important properties for the coding of sequences defined above. They are needed for the representation of primitive recursive functions, as we have seen in section 7.1. The first one says that for every x there exists a sequence starting with x. The second one says that every sequence can be extended. The third one is a technical condition needed for the proof of theorem 40.

Theorem 9 (Properties of sequences)**.**

1. $\mathrm{PA} \vdash \forall x \exists a, m((a,m)_0 = x)$
2. $\mathrm{PA} \vdash \forall yxam \exists bn(\forall i < y((a,m)_i = (b,n)_i \wedge (b,n)_y = x)$
3. $\mathrm{PA} \vdash \forall ami((a,m)_i \leq a)$

Proof.

1. For the first one, take $m = x$ and $a = 2x+1$. Then:

$$(a,m)_0 = \mathrm{rm}(a, m \cdot (0+1)+1) = \text{2x+1, x+1} = x \tag{32}$$

2. For the second one, we prove three properties in PA:

$$\begin{aligned} &\mathrm{PA} \vdash \forall yxam \exists u(\forall i < y((a,m)_i < u) \wedge x < u \wedge y < u) \\ &\mathrm{PA} \vdash \forall u \exists v \geq 1 \forall i \leq u(i \geq 1 \rightarrow i|v) \\ &\mathrm{PA} \vdash \forall uv(\forall i \leq u(i \geq 1 \rightarrow i|v) \\ &\quad \rightarrow \forall ij(0 \leq i < j \leq u \rightarrow \gcd((i+1)\cdot v+1, (j+1)\cdot v+1) = 1)) \end{aligned} \tag{33}$$

   - For the first property we use induction on y. For $y = 0$, take $u = x+1$. Then there are no $i < y$ so the first condition trivially holds, $0 < x+1$ and $x < x+1$.
     Now assume for y that for any x, a, m we have

$$\exists u(\forall i < y((a,m)_i < u) \wedge x < u \wedge y < u) \tag{34}$$

     Take $u' = u + (a,m)_y$. Observe that $u > 0$ because $(a,m)_i < u$. If $i < y+1$ we have $i < y$ or $i = y$. For $i < y$ we know by our induction hypothesis that

$$((a,m)_i < u < u') \wedge (x < u < u') \wedge (y < u < u') \tag{35}$$

     and for $i = y$ we have $(a,m)_y < (a,m)_y + u = u'$. So for u' we see

$$\forall i < y+1((a,m)_i < u') \wedge x < u' \wedge y < u' \tag{36}$$

- For the second property we use induction on u. For $u = 0$, the only $i \leq u$ is $i = 0$. So the implication trivially holds because $i < 1$ for any v. Now assume for u we have

  $$\exists v \geq 1 \forall i \leq u(i \geq 1 \rightarrow i|v) \tag{37}$$

  Now for $u + 1$ take $v' = v \cdot (u + 1)$. Then for $1 \leq i \leq u$ we have $i|v \cdot (u + 1)$ because $i|v$ by the induction hypothesis. For $i = u + 1$ we have $(u + 1)|v \cdot (u + 1)$. So

  $$\forall i \leq (u + 1)i \geq 1 \rightarrow i|v') \tag{38}$$

- For the third property we observe that this is just the formal statement of a part of the proof of theorem 7. We proved there that $(i+i) \cdot m+1$ and $(j+1) \cdot m+1$ are relatively prime for $0 \leq i < j \leq n$, using that for all $1 \leq (i - j) \leq n$ we had that $(i - j)|m$, and other properties of gcd. This property states that this proof can be carried out in PA. I will not proof this here.

Now using these three properties, we will prove the second statement of the theorem. Given y, x, a, m, take u satisfying the first property and for that u, take a v satisfying the second. Put $n = v$. Then we have:

$$\begin{aligned}
&\forall i < y((a, m)_i < (i + 1) \cdot n + 1) \\
&x < (y + 1) \cdot n + 1 \\
&\forall ij(0 \leq i < j \leq y \rightarrow \gcd((i + 1) \cdot n + 1, (j + 1) \cdot n + 1) = 1
\end{aligned} \tag{39}$$

The first follows from the first property, using that $(i + 1) \cdot n + 1 > n = v \geq u$, which follows from a special case of the second property: if $u \neq 0$, then $u|v$ which implies that $u \leq v$.

The second follows from the same reasoning, only now using $(y+1) \cdot n+1 > n$.

The third is just the third property, only using that $y < u$ from the first property.

Now we want to find b such that

$$(\forall i < y((a, m)_i = (b, n)_i)) \wedge x = (b, n)_y \tag{40}$$

We use induction on k. Suppose for $k < y$ we have b' satisfying the following:

$$(\forall i < k((a, m)_i = (b', n)_i)) \wedge x = (b', n)_y \tag{41}$$

Then we want to find b, such that the following holds:

$$(\forall i \leq k((a, m)_i = (b, n)_i)) \wedge x = (b, n)_y \tag{42}$$

To do so we will prove the following:

$$\exists w((y+1) \cdot n+1|w \wedge \forall i < k((i+1) \cdot n+1|w) \wedge \gcd(w, (k+1) \cdot n+1) = 1) \tag{43}$$

We do this by induction on k. For $k = 0$, there is no $i < k$ so we have to prove

$$\exists w((y+1)\cdot n+1|w \wedge \gcd(w, n+1) = 1) \tag{44}$$

Let $w = (y+1)\cdot n+1$, then $(y+1)\cdot n+1|w$ and by the third property we have for n we see that, taking $i = 0$ and $j = y$, that $\gcd(w, n+1) = \gcd((y+1)\cdot n+1, (0+1)\cdot n+1) = 1$.

Now assume such w exists for $k-1$, where $0 < k < y$. For k, let $w' = w\cdot(k\cdot n+1)$.

Then we know $(y+1)\cdot n+1|w'$, because $(y+1)\cdot n+1|w$ (by induction hypothesis).

Also we see that $\gcd(w\cdot(k\cdot n+1), (k+1)\cdot n+1) = 1$: We have $\gcd(w, (k+1)\cdot n+1) = 1$, because $\gcd(w, k\cdot n+1) = 1$ by induction hypothesis and $\gcd((k+1)\cdot n+1, k\cdot n+1) = 1$ by the third property. Also we have $\gcd(k\cdot n+1, (k+1)\cdot n+1) = 1$, again by the third property, so the required property follows.

And we see that $\forall i < k((i+1)\cdot n+1|w')$ because $w' = w\cdot(k\cdot n+1)$: for $i < k-1$, we have by induction hypothesis that $(i+1)\cdot n+1|w$, and for $i = k-1$ we have $(k-1+1)\cdot n+1|k\cdot n+1$.

So by induction we proved that such w always exist. Now take such w. Now apply theorem 6, for $x = w$, $y = (k+1)\cdot n+1$. The theorem gives us f, g such that $f\cdot w = g\cdot((k+1)\cdot n+1) + \gcd(w, (k+1)\cdot n+1)$. We know that for all k, $\gcd(w, (k+1)\cdot n+1) = 1$. So we have:

$$\mathrm{rm}(f\cdot w, (k+1)\cdot n+1) = \gcd(w, (k+1)\cdot n+1) = 1 \tag{45}$$

Now take $b = b' + f\cdot w\cdot(b'\cdot n\cdot(k+1) + (a,m)_k)$. We want to show $(b,n)_y = x$ and $\forall i \leq k((b,n)_i) = (a,m)_i)$.

We see $(b,n)_y = (b',n)_y = x$, because by definition this means $\mathrm{rm}(b, (y+1)\cdot n+1) = \mathrm{rm}(b', (y+1)\cdot n+1)$ and this is true because $b = b' + w\cdot A$, for some term A, and by choice of w we know $(y+1)\cdot n+1|w$ so $\mathrm{rm}(w\cdot A, (y+1)\cdot n+1) = 0$.

For $i < k$ we have $(b,n)_i = (b',n)_i = (a,m)_i$, because also by choice of w we have $(i+1)\cdot n+1|w$, so we may use the same reasoning as before.

Now for $(b,n)_k$ we have:

$$\begin{aligned}(b,n)_k &= \mathrm{rm}(b, (k+1)\cdot n+1)\\ &= \mathrm{rm}(b' + b'\cdot n\cdot(k+1) + (a,m)_k, (k+1)\cdot n+1)\\ &= \mathrm{rm}(b'((k+1)\cdot n+1 + (a,m)_k, (k+1)\cdot n+1)\\ &= (a,m)_k\end{aligned} \tag{46}$$

This completes the induction and we see that indeed $\forall yxam\exists bn(\forall i < y((a,m)_i = (b,n)_i \wedge (b,n)_y = x)$

3. For the third one: $(a,m)_i$ is the remainder of a divided by some integer ≥ 1. This means that $(a,m)_i \leq a$.

□

4 Introduction to Logic

To be able to understand the proofs given later, it is important to see some definitions used in logic.

First of all, it is important to understand what logic actually is. However, there are several conceptions of logic. The one we will study here is logic in the sense of studying certain mathematical properties of artificial formal languages.[38] This roughly means that one can define a formal language, in which certain properties can be expressed. Now one can find models of a language; in a model all the properties that are true in the language, are true. This means that once something has been proved in the formal language, it is true in all models of that language. This might seem rather abstract, but we will explain all of this in detail below.

The first important notion is that of a *language*. A language is a set of symbols. These symbols can be devided into three groups: *constants, function symbols* and *relation symbols*.

Functions symbols and relation symbols have an *arity*. The arity specifies the number of arguments the function symbol or relation symbol takes.

From a language, *terms* and *formulae* can be deduced. Terms denote elements and formulae state properties. To deduce these, auxiliary symbols are used. These are the following:

- *Variables*, this is an infinite set of symbols, usually left unspecified.
- The equality symbol $=$.
- The absurdity symbol $\bot$.
- Connectives: $\wedge$, $\vee$, $\rightarrow$, and $\neg$. These are the symbols for respectively conjunction ('and'), disjunction ('or'), implication ('implies') and negation ('not').
- Quantifiers: $\exists$ and $\forall$. These are symbols for respectively the existential quantifier ('there exists') and the universal quantifier ('for all').
- Readability symbols like brackets.

Definition 10. *Terms* of a language are defined inductively:

- Every constant of the language is a term of the language.
- Every variable is a term of the language.
- If $t_1, \ldots, t_n$ is an n-tuple of terms of the language, and f is an n-ary function symbol of the language, then $f(t_1, \ldots, t_n)$ is a term of the language.

Definition 11. *Formulae* of a language are defined inductively as well:

- If t and s are terms of the language, then $(t = s)$ is a formula of the language.
- If $t_1, \ldots, t_n$ is an n-tuple of terms of the language, and R is an n-ary relation symbol of the language, then $R(t_1, \ldots, t_n)$ is a formula of the language.

[38] See [Hofweber, 2013] for a full explanation of all conceptions of logic.

- $\perp$ is a formula of the language.
- If φ and ψ are formulae of the language, then $(\varphi \wedge \psi)$, $(\varphi \vee \psi)$, $(\varphi \rightarrow \psi)$ and $(\neg\varphi)$ are also formulae of the language.
- If φ is a formula of the language and x is a variable, then $\forall x\varphi$ and $\exists x\varphi$ are formulae of the language as well.

A variable is called *bound* if it bound by a quantifier, like x in $\forall x\varphi$. If a variable is not bound, it is called *free*. In logic one usually uses the *Convention of variables*, which says: A variable in a formula will always be either bound or free, and not both. If it is bound, it is only bound once. If in a term or a formula, all variables are bound, we call it, respectively, a *closed term* and a *closed formula*.

One uses the convention of variables in the following definition, from [van Oosten, 2009]:

Definition 12. Suppose φ is a formula of the language, and t is a term of the language. By the *substitution* $\varphi[t/x]$ we mean the formula which results by replacing each occurence of the variable x by the term t, provided x is a free variable in φ, and no variable in the term t becomes bound in φ.

In section 3 it was said that PA was a system of *first-order logic*. In first order logic, variables denote elements of structures. This means we can only say something about all elements of a structure, and not, for example, something about subsets of all elements, or sequences, or all possible subsets. To be able to do this, one needs higher order logic. In this thesis we will restrict ourselves to first order logic.

Say we have a first order language, which we name L. Then we can have a *structure for* L, which is defined in [van Oosten, 2009] as follows:

Definition 13. An *L-structure* M consists of a nonempty set, also denoted M, together with the following data:

- for each contstant c of L, an element c^M of M;
- for each n-ary function symbol f of L, a function
$$f^M : M^n \rightarrow M \tag{47}$$
- for each n-ary relation symbol R of L, a subset
$$R^M \subseteq M^n \tag{48}$$

In this, c^M is called the *interpretation* of c, and similarly f^M and R^M are called the interpretations of f and R, respectively.

Now we can define a new language L_M. Given a language L and an L-structure M, we can define L_M to be L, together with, for every element m of M, an extra constant, which is also denoted m. Now M is also an L_M-structure, if we say that the interpretation of each constant m is just the element m.

Definition 14. We can define the *interpretation of closed terms* of the language L_M in M inductively. If t is a constant, then its interpretation is already defined, because t was already in L and M is an L-structure. If $t_1, \ldots, t_n$ are terms which have interpretations $t_1^M, \ldots, t_n^M$, then if $t = f(t_1, \ldots, t_n)$ then $t^M = f^M(t_1^M, \ldots, t_n^M)$.

Definition 15. We define *'φ is true in M'*, or $M \models \varphi$, by induction on φ:

- For closed terms $t, s, t_1, \ldots, t_n$:

$$\begin{gathered} M \models \bot \text{ never holds} \\ M \models (t = s) \text{ if and only if } t^M = s^M \\ M \models R(t_1, \ldots, t_n) \text{ if and only if } (t_1^M, \ldots, t_n^M) \in R^M \end{gathered} \tag{49}$$

- If φ is of the form $\varphi_1 \wedge \varphi_2$:

$$M \models \varphi \text{ if and only if } M \models \varphi_1 \text{ and } M \models \varphi_2 \tag{50}$$

- If φ is of the form $\varphi_1 \vee \varphi_2$, we define the following, where 'or' is the inclusive or:

$$M \models \varphi \text{ if and only if } M \models \varphi_1 \text{ or } M \models \varphi_2 \tag{51}$$

- If φ is of the form $\varphi_1 \to \varphi_2$:

$$M \models \varphi \text{ if and only if } M \models \varphi_2 \text{ whenever } M \models \varphi_1 \tag{52}$$

- If φ is of the form $\neg\psi$, we define the following, where $\not\models$ is 'not $\models$':

$$M \models \varphi \text{ if and only if } M \not\models \psi \tag{53}$$

- If φ is of the form $\forall x \psi$:

$$M \models \varphi \text{ if and only if } M \models \psi[m/x] \text{ for all } m \in M \tag{54}$$

- If φ is of the form $\exists x \psi$:

$$M \models \varphi \text{ if and only if } M \models \psi[m/x] \text{ for some } m \in M \tag{55}$$

Lemma 16 (Prenex normal form)**.** For every formula φ, φ is equivalent to a formula which starts with a string of quantifiers, followed by a formula in which no quantifiers occur. Such a formula is said to be written in *prenex normal form.*

Definition 17. If L is a language, an *L-theory* is a set of closed L-formulae.

For example, PA is a theory in the language $L = \{0, 1, +, \cdot\}$

Definition 18. If Γ is an L-theory, then an L-structure M is called a *model* of Γ if $M \models \varphi$ for all $\varphi \in \Gamma$.

For example, the natural numbers, $\mathbb{N}$, are a model of PA.

Definition 19. An L-theory Γ is called *consistent* if Γ has a model.

If $M \models \varphi$ for every model M of Γ then we write $\Gamma \models \varphi$.

Theorem 20 (Compactness Theorem)**.** Let Γ be a theory in a language L. If every finite $\Gamma' \subseteq \Gamma$ is consistent, then so is Γ.

In this, $\Gamma' \subseteq \Gamma$ has the intuitive meaning; Γ' contains some or all closed formulae in Γ, and no formulae that are not in Γ. The compactness theorem was stated and proved by Gödel in 1930, using results that he had found in 1930 as well, as we will see in section 4.2.[39]

4.1 Models of PA

We've seen already that $\mathbb{N}$ is a model of $\mathcal{L}_{PA}$, if we use the usual addition and multiplication and the constants $0, 1$. We call this the *standard model.* We denote this model by $\mathcal{N}$. However, PA also has non-standard models.

For every $n \in \mathbb{N}$, define a term $\overline{n}$ (which we call a *numeral*) of $\mathcal{L}_{PA}$ by recursion: $\overline{0} = 0$ and $\overline{n+1} = \overline{n} + 1$. Let c be a constant, different from the ones we already had, and consider in the language $\mathcal{L}_{PA} \cup \{c\}$ the set of axioms:

$$\{\text{axioms of PA}\} \cup \{\neg(c = \overline{n}) | n \in \mathbb{N}\} \tag{56}$$

These, infinitely many, axioms state that c is not equal to any $\overline{n}$ in $\mathcal{L}_{PA}$. Every finite subset of this theory has an interpretation in $\mathbb{N}$, because then c can be chosen to equal an n which is not included in any of the (now finitely many) added axioms. So, by the compactness theorem, this is a consistent set of axioms and therefore has a model $\mathcal{M}$, which has a *nonstandard element* $c^{\mathcal{M}}$.

4.2 Proofs

Proofs can be formalised using *proof trees.* We will now give a rough idea of how proof trees are defined.

Definition 21. A *tree* is a *partial order*, which means there is a set T and a relation $\leq$ on T which relates some elements of T. This partial order has a *least element*, which means there is an element a for which the following holds for every x in T: if $x \leq a$ then $x = a$. This partial order is such that for every $x \in T$ the set $\{y \in T | y \leq x\}$ is a *finite linear order*, which means that for any a, b in the set, $a \leq b$ and/or $b \leq a$. The maximal elements of the tree are called *leaves.*

Now elements of a tree can be *labelled.* Roughly speaking a *labelled tree* is a finite tree T, together with a function f from T to the set which is the disjoint union of the set of formulae and the set of labelled formulae, called the *labelling function*, such that the only elements of T which are marked by f are leaves of T. An operation on trees is the labelling of the leaves, and labelled leaves are then called *eliminated assumptions.*

One can *join* labelled trees, by adding a root. If T and S are labelled trees and φ a formula, then we can join these by adding a root r for which $f(r) = \varphi$. We denote this by $\Sigma(T, S; \varphi)$

[39] A proof of the compactness theorem can be found in [Moerdijk and van Oosten, 2000] p. 95.

Now the set $\mathcal{P}$ of *proof trees* is the smallest set of labelled trees, satisfying a list of properties, e.g.:

- Ass: For every formula φ, the tree with one element r and labelling function $f(r) = \varphi$, is an element of $\mathcal{P}$. This tree is called the *assumption tree.*
- $\wedge I$: If T is an element of $\mathcal{P}$ with conclusion φ, and S is an element of $\mathcal{P}$ with conclusion ψ, then $\Sigma(T, S; \varphi \wedge \psi)$ is an element of $\mathcal{P}$. This is formed by $\wedge$*-introduction.*
- $\forall E$: Suppose T is an element of $\mathcal{P}$ with conclusion $\forall u\varphi$, and t is a term such that the substitution $\varphi[t/u]$ is defined. Then $\Sigma(T; \varphi[t/u])$ is an element of $\mathcal{P}$. This is called $\forall$*-elimination.*

These properties are cited from [van Oosten, 2009], and for a complete list of this properties and a formal definition of proof trees, see [Moerdijk and van Oosten, 2000].

A proof tree uses the axioms of a theory as unmarked assumptions, and the least element of a proof tree is the formula that is proved. If there is a proof tree with conclusion φ and unmarked assumptions that are either axioms of a theory Γ or of the form $\forall x(x = x)$, then we write $\Gamma \vdash \varphi$.

An example of a proof tree is the following:

$$\dfrac{\forall E\,\dfrac{\forall x\varphi(x)}{\varphi[t/x]} \qquad \psi}{\varphi[t/x] \wedge \psi}\,\wedge I$$

Now we present two fundamental theorems in logic.

Theorem 22 (Soundness Theorem)**.** If $\Gamma \vdash \varphi$ then $\Gamma \models \varphi$

Theorem 23 (Completeness Theorem)**.** If $\Gamma \models \varphi$ then $\Gamma \vdash \varphi$

The completeness theorem was formulated and proved by Gödel in 1930, just before he presented his incompleteness theorems. The completeness theorem and the soundness theorem together are equivalent to the following assertion: A theory Γ is consistent if and only if $\Gamma \nvdash \perp$. Also the completeness theorem and the soundness theorem together prove the compactness theorem, which was also stated and proved by Gödel in 1930. For a proof of these theorems, see [Moerdijk and van Oosten, 2000].

5 Primitive Recursive Functions

The proof of Gödel's incompleteness theorems is based on the fact that some special types of functions, *primitive recursive* functions, *partial recursive* functions and *total recursive* functions are *representable* in PA, and that they are *computable*: one can find an algorithm that computes the value of these functions at any given time. Some of the new concepts introduced here will be formalised in this section. The definitions of these concepts will be given and some important theorems. For the proof of these theorems and a more detailed explanation of these concepts, see [van Oosten, 2009]. The application of these definitions and theorems can be found in section 7.1.

We will use the so-called λ-*notation*: if $\vec{x}$ is a sequence of variables $x_1, \ldots, x_k$ which might occur in the expression G, then $\lambda\vec{x}.G$ denotes the function which assigns to the k-tuple $n_1, \ldots, n_k$ the value $G(n_1, \ldots, n_k)$.

Definition 24. The class of primitive recursive functions $\mathbb{N}^k \to \mathbb{N}$ is defined in the following way:

- The *zero function* $Z = \lambda x.0$ is primitive recursive.
- The *successor function* $S = \lambda x.x + 1$ is primitive recursive.
- The *projections* $\Pi_i^k = \lambda x_1 \cdots x_k.x_i$ (for $1 \leq i \leq k$) are primitive recursive.
- If $G_1, \ldots, G_l : \mathbb{N}^k \to \mathbb{N}$ and $H : \mathbb{N}^l \to \mathbb{N}$ are primitive recursive, then so is $\lambda\vec{(x)}.H(G_1(\vec{x}), \ldots, G_l(\vec{x}))$. This function is said to be defined from $G_1, \ldots, G_l$ and H by *composition.*
- If $G : \mathbb{N}^k \to \mathbb{N}$ and $H : \mathbb{N}^{k+2} \to \mathbb{N}$ are primitive recursive, then so is the function F defined from G and H by *primitive recursion*:

$$\begin{aligned} F(0, \vec{x}) &= G(\vec{x}) \\ F(y + 1, \vec{x}) &= H(y, F(y, \vec{x}), \vec{x}) \end{aligned} \tag{57}$$

A relation $A \subseteq \mathbb{N}^k$ can also be primitive recursive: A is primitive recursive if its *characteristic function* χ_A is primitive recursive, where χ_A is defined as follows:

$$\chi_A(\vec{x}) = \begin{cases} 0 \text{ if } \vec{x} \in A \\ 1 \text{ else} \end{cases} \tag{58}$$

Definition 25. The function $j : \mathbb{N}^2 \to \mathbb{N}$, the *diagonal enumeration* of $\mathbb{N}$, is given as follows:

$$j(n, m) = \frac{1}{2}(n + m)(n + m + 1) + n \tag{59}$$

This function is a bijection. Define inverse functions j_1 and j_2 as follows:

$$\begin{aligned} j_1(z) &= \mu x \leq z.[\exists y \leq z.j(x, y) = z] \\ j_2(z) &= \mu y \leq z.[\exists x \leq z.j(x, y) = z] \end{aligned} \tag{60}$$

The μ in this function is a notation for a primitive recursive function, in this case it expresses the function that 'picks the lowest x or y smaller than z such that there exists respectively an y or x such that $j(x,y)=z$. A more formal definition of this μ can be found in Definition 31. j_1 and j_2 are primitive recursive and with this definition the following holds:

$$\begin{aligned} x &\leq j(x,y) \\ y &\leq j(x,y) \\ j(j_1(z),j_2(z)) &= z \end{aligned} \tag{61}$$

Theorem 26. The functions j^m and j_i^m defined below are primitive recursive.

The bijections $j^m : \mathbb{N}^m \to \mathbb{N}$ are defined as follows:

$$\begin{aligned} &j^1 \text{ is the identity function} \\ j^{m+1}(x_1,\ldots,x_m,x_{m+1}) &= j(j^m(x_1,\ldots,x_m),x_{m+1}) \end{aligned} \tag{62}$$

The *projection functions* $j_i^m : \mathbb{N} \to \mathbb{N}$ for $1 \leq i \leq m$, are defined as follows:

$$\begin{aligned} j_1^1(z) &= z \\ j_i^{m+1}(z) &= \begin{cases} j_i^m(j_1(z)) \text{ if } 1 \leq i \leq m \\ j_2(z) \text{ if } i = m+1 \end{cases} \end{aligned} \tag{63}$$

These functions can be used to code sequences:

Theorem 27. Define the *code of the sequence* as follows:

$$\begin{aligned} \langle\rangle &= 0 \\ \langle x_0,\ldots,x_{m-1}\rangle &= j(m-1,j^m(x_0,\ldots,x_{m-1}))+1 \text{ if } m>0 \end{aligned} \tag{64}$$

For every $y \in \mathbb{N}$ it holds that either $y=0$ or there is a unique $m>0$ and a unique sequence $(x_0,\ldots,x_{m-1})$ such that $y = \langle x_0,\ldots,x_{m-1}\rangle$.

There are also functions $lh(x)$, giving the length of a sequence with code x, and $(x)_i$, giving the i'th element of a sequence with code x (this holds for $0 \leq i \leq lh(x)$, for $i > lh(x)$ it is 0). These functions are primitive recursive.

Primitive recursive functions are defined in terms of the previous value of the function. It is also possible to define functions in terms of all previous values, this is called *course-of-values recursion.*

Definition 28. Let $G : \mathbb{N}^k \to \mathbb{N}$ and $H : \mathbb{N}^{k+2} \to \mathbb{N}$ be functions. $F : \mathbb{N}^{k+1} \to \mathbb{N}$ is defined from G and H by course-of-values recursion if it is defined as follows:

$$\begin{aligned} F(0,\vec{x}) &= G(\vec{x}) \\ F(y+1,\vec{x}) &= H(y,j^{y+1}(F(0,\vec{x}),\ldots,F(y,\vec{x})),\vec{x}) \end{aligned} \tag{65}$$

If G and H in the above definition are primitive recursive, then F, defined by course-of-values recursion is also primitive recursive.

Theorem 29. If G_1, G_2 and H are primitive recursive functions $\mathbb{N}^n \to \mathbb{N}$, then so is the function F, defined by

$$F(\vec{x}) = \begin{cases} G_1(\vec{x}) \text{ if } H(\vec{x}) = 0 \\ G_2(\vec{x}) \text{ else} \end{cases} \tag{66}$$

There are also *partial recursive functions*, which is a weaker than primitive recursion. First we define some concepts for functions:

Definition 30. A *partial function* F from X to Y, where X and Y are sets, is a function $F : U \to Y$ where $U \subseteq X$. We write $F : X \rightharpoonup Y$ to indicate that F is partial, and we write $U = \mathrm{dom}(F)$. If $U = X$ then F is called a *total function.* If $F : X \rightharpoonup Y$ and $G : Y \rightharpoonup Z$ then $GF : X \rightharpoonup Z$ is the composition with $\mathrm{dom}(GF) = \{x \in X | x \in \mathrm{dom}(F) \text{ and } F(x) \in \mathrm{dom}(G)\} \subseteq X$. Define $F(x) \simeq G(x)$ as follows: $F(x)$ is defined precisely when $G(x)$ is defined, and when they are defined, $F(x) = G(x)$.

Definition 31. The class of partial recursive functions is defined in the following way:

- All primitive recursive functions are partially recursive.
- The partial recursive functions are closed under composition: if the functions $G_1, \ldots, G_l : \mathbb{N}^k \rightharpoonup \mathbb{N}$ and $H : \mathbb{N}^l \rightharpoonup \mathbb{N}$ are partial recursive, then so is the function $\lambda\vec{x}.H(G_1(\vec{x}), \ldots, G_l(\vec{x}))$, with domain $\{\vec{x} \in \bigcap_{i=1}^{l} \mathrm{dom}(G_i) | (G_1(\vec{x}), \ldots, G_l(\vec{x})) \in \mathrm{dom}(H)\}$.
- If $G : \mathbb{N}^{k+1} \rightharpoonup \mathbb{N}$ is partial recursive, then so is the function F defined from G by *minimisation*: $F(\vec{x}) \simeq \mu y.G(\vec{x}, y) = 0$. $F(\vec{x})$ is defined precisely when there exists a y such that $\forall i \leq y.(\vec{x}, i) \in \mathrm{dom}(G)$ and $G(\vec{x}, y) = 0$. Then $F(\vec{x})$ is the least y with this property.

A partial recursive function is *total recursive* if it is a total function. A relation $A \subseteq \mathbb{N}^k$ is recursive if its characteristic function χ_A is partial recursive.

6 Gödels Theorems

Gödel originally stated his theorems about PA. Later, Rosser showed that the theorems could be extended to apply also to other systems - as long as enough elementary arithmetic could be carried out, e.g. to formalise primitive recursion. These final theorems have been formulated in [Franzén, 2005] in an intuitive manner:

> **Theorem 32** (First Incompleteness Theorem)**.** Any consistent formal system S within which a certain amount of elementary arithmetic can be carried out is incomplete with regard to statements of elementary arithmetic: there are such statements which can neither be proved, nor disproved in S.[40]

> **Theorem 33** (Second Incompleteness Theorem)**.** For any consistent formal system S within which a certain amount of elementary aritmetic can be carried out, the consistency of S cannot be proved in S itself.[41]

The 'certain amount of elementary arithmetic' that is needed in order for the theorems to apply, is not the same amount for both theorems. Surely, both theorems apply to PA, as this is what Gödel proved.

For the first incompleteness theorem, we could also take any system including PA, or any system in which the language of arithmetic can be defined (the language of arithmetic is the language of PA, see section 3) and in which the first six axioms of PA (again, see section 3) can be proved. However, they only need to be proved in a restricted version, applying to the objects satisfying some formula $N(x)$. This formula could be saying 'x is a natural number'. Set-theoretically this is very well possible. This means that PA can be represented in ZFC, which means that the first incompleteness theorem also applies to ZFC, that is, to the *arithmetical component* of ZFC, the part representing PA.

The last comment touches upon a common misunderstanding of the incompleteness theorems: whenever enough arithmetic can be represented in a system, the incompleteness theorems apply to it. However, it only applies to the part that represents arithmetic. An example, taken from Franzén [2005]: it has been stated that we will never know whether or not there exist ghosts, because of the first incompleteness theorem. The argument used is that we could add arithmetic to a theory about ghosts, so that the first incompleteness theorem applies to it and therefore, there are undecidable sentences. But even if we would add enough arithmetic so that the first incompleteness theorem would indeed apply to the ghost-arithmetic system, it would only apply to the part concerning the arithmetic, not the part about ghosts. The undecidable sentence that does indeed exist, is an arithmetical sentence. So it would not be possible to conclude anything about ghosts using Gödel's theorems.

For the second incompleteness theorem, the amount of arithmetic needed is slightly different. To prove the second incompleteness theorem, part of the

[40] [Franzén, 2005] p. 16
[41] [Franzén, 2005] p. 34

proof of the first incompleteness theorem needs to be formalised in the theory. This means a larger amount of arithmetic is needed. We see the use of this in the treatment of Formalised Σ_1-completeness (theorem 50).

An example of a theory in in which enough arithmetic can be carried out to for the first incompleteness theorem to apply to it, but not enough for the Second to apply to it, is *Robinson Arithmetic*. Robinson Arithmetic is a finite part of PA: it is the first six axioms, together with the axiom $\forall x(x = 0 \vee \exists y(y + 1 = x)$. Alternatively, one could say it is the first six axioms of PA together with the definition of $<$ as strict total order.

7 Proof of the Theorems

Now for the formal statement of the theorems and the proof, we will restrict ourselves to the versions stated and proved by Gödel. (In fact only the first incompleteness theorem was proved by Gödel in detail, the second incompleteness theorem was proved in the second part of the *Grundlagen der Mathematik* of Hilbert and Bernays). Gödel only gave arguments about how the proof should work.) These versions only apply to PA. The theorems stated in a mathematical way look like this:

Theorem 34 (First Incompleteness Theorem)**.** There is a Π_1-sentence G such that $\mathrm{PA} \vdash G \leftrightarrow \neg\Box G$. G is independent of PA and thus PA is incomplete.

Theorem 35 (Second Incompleteness Theorem)**.**

$$\mathrm{PA} \nvdash \mathrm{Con}_{\mathrm{PA}} \tag{67}$$

This formulation of course raises questions, as there is a lot of notation not introduced yet. In order to fully understand the theorems and the proofs, we need to answer these questions, e.g.: How did we find this sentence G, what is the $\Box$, what is a Π_1 sentence and what is $\mathrm{Con}_{\mathrm{PA}}$?

To be able to answer these questions, we need a few more definitions, the *Diagonalisation lemma*, and the representability and some features of primitive recursive functions in PA. We will introduce these, before we explain in detail the first incompleteness theorem, its proof, and the second incompleteness theorem.

7.1 Primitive Recursion in PA

We have seen that a lot of arithmetic can be carried out in PA, in section 3.1. Now we will see that also primitive recursive functions can be represented in PA, by a special sort of formulae, which will be defined shortly. The fact that primitive recursive functions (or, actually, all recursive functions) can be represented in PA is important in the proof of the *Diagonalisation lemma* (47) and in the construction of the *Gödel sentence* G that appeared in the first incompleteness theorem. We will start with some definitions.

Definition 36. An $\mathcal{L}_{\mathrm{PA}}$-formula φ is called a Δ_0-formula if all quantifiers are bounded in φ, that is of the form $\forall x < t$ or $\exists x < t$, for a term t not containing the variable x. We write $\varphi \in \Delta_0$.

An $\mathcal{L}_{\mathrm{PA}}$-formula φ is called a Σ_1-formula if it is of the form $\exists y_1 \dots y_t \psi$ with ψ a Δ_0-formula. We write $\varphi \in \Sigma_1$.

An $\mathcal{L}_{\mathrm{PA}}$-formula φ is called a Δ_1-formula if both φ and $\neg\varphi$ are equivalent in PA to a Σ_1-formula. We write $\varphi \in \Delta_1$.

An $\mathcal{L}_{\mathrm{PA}}$-formula φ is called a Π_1-formula if it is of the form $\forall y_1 \dots y_t \psi$ with $\psi \in \Delta_0$.

First we will state an important property of Σ_1-formulae. Once we have proved that, we will look at the representation of recursive functions.

Lemma 37 (Σ_1-Completeness of PA)**.** A closed Σ_1-formula is provable in PA if and only if it is true in $\mathbb{N}$.

Proof. We will first prove the completeness of PA with respect to Δ_0-formulae, with induction to Δ_0-formula.

- First we will show the equivalence for atomic formulae. Remember, as we have seen in section 4.1, that numerals are defined inductively as follows:

$$\overline{0} = 0 \qquad\qquad \overline{n+1} = \overline{n} + 1 \tag{68}$$

  - First we show $(\text{PA} \vdash \overline{n} + \overline{m} = \overline{k}) \Leftrightarrow n + m = k$, for all $n, m, k \in \mathbb{N}$ with induction on m.

    For $m = 0$ we have to prove $\text{PA} \vdash (\overline{n} + \overline{0} = \overline{k}) \Leftrightarrow n = k$ Because we know $\text{PA} \vdash (\overline{n} + \overline{0} = \overline{n})$, this means we have to show that $\text{PA} \vdash (\overline{n} = \overline{k}) \Leftrightarrow n = k$ This can be done by induction on k:

    For $k = 0$ we have $\text{PA} \vdash (\overline{n} = \overline{0} = 0) \Leftrightarrow n = 0$.

    For $k + 1$ we have $\text{PA} \vdash (\overline{n} = \overline{k+1} = \overline{k} + 1) \Leftrightarrow n = k + 1$, because $\text{PA} \vdash (\overline{n} = \overline{k}) \Leftrightarrow n = k$.

    Now assume $(\text{PA} \vdash \overline{n} + \overline{m} = \overline{k}) \Leftrightarrow n + m = k$ holds for m, then for $m + 1$ we have: $(\text{PA} \vdash \overline{n} + \overline{m+1} = \overline{n} + \overline{m} + 1 = \overline{k} + 1)$. With our induction hypothesis we now know: $(\text{PA} \vdash \overline{n} + \overline{m} + 1 = \overline{k} + 1) \Leftrightarrow n + m + 1 = k + 1$.

  - Now we show $(\text{PA} \vdash \overline{n} \cdot \overline{m} = \overline{k}) \Leftrightarrow n \cdot m = k$, for all $n, m, k \in \mathbb{N}$ with induction on m.

    For $m = 0$ we know that $\overline{n} \cdot \overline{0} = \overline{n} \cdot 0 = 0$, by one of the axioms for PA. So we have $(\text{PA} \vdash \overline{k} = 0) \Leftrightarrow k = 0$, which holds for all $n \in \mathbb{N}$.

    Now assume $(\text{PA} \vdash \overline{n} \cdot \overline{m} = \overline{k}) \Leftrightarrow n \cdot m = k$ holds for m, then for $m + 1$ we have: $(\text{PA} \vdash \overline{n} \cdot \overline{m+1} = \overline{n} \cdot (\overline{m} + 1) = \overline{k} + \overline{n})$. With our induction hypothesis and our previous result for addition we now know: $(\text{PA} \vdash \overline{n} \cdot (\overline{m} + 1) = \overline{k} + \overline{n} \Leftrightarrow n \cdot (m + 1) = n \cdot m + n$. for all $n, m, k \in \mathbb{N}$ with induction on m.

  - Now we show $(\text{PA} \vdash \overline{n} < \overline{m}) \Leftrightarrow n < m$ for all $n, m \in \mathbb{N}$ by induction on m.

    For $m = 0$ the equivalence automatically holds because in $\mathbb{N}$ there are no n such that $n < 0$ and $\text{PA} \vdash \forall x \neg(x + 1 = 0)$ which implies that there are no x such that $x < 0$.

    Now assume $(\text{PA} \vdash \overline{n} < \overline{m}) \Leftrightarrow n < m$ holds for m. Then for $m + 1$ we have: $(\text{PA} \vdash \overline{n} < \overline{m+1} = \overline{m} + 1)$. With our induction hypothesis we can now say that if $\overline{n} < \overline{m}$ then this is equivalent to $n < m$ for all $n, m \in \mathbb{N}$. If $\overline{n} = \overline{m}$ then of course $n = m$.

  - The last to show is $\text{PA} \vdash \forall x(x < \overline{n} \leftrightarrow x = \overline{0} \vee \ldots \vee x = \overline{n-1})$ for all $n \in \mathbb{N}$. For $n = 0$ the implication trivially holds because there are no $x < 0$.

    Now assume $\text{PA} \vdash \forall x(x < \overline{n} \leftrightarrow x = \overline{0} \vee \ldots \vee x = \overline{n-1})$ holds for some n. Then for $n + 1$ we have that $x < \overline{n} + 1 \rightarrow (x < \overline{n} \vee x = \overline{n}$. If $x < \overline{n}$ with our induction hypothesis we have $x = \overline{0} \vee \ldots \vee x = \overline{n-1})$. So for any $x < \overline{n} + 1$ we have $x = \overline{0} \vee \ldots \vee x = \overline{n-1} \vee x = \overline{n}$. So we see that $\text{PA} \vdash \forall x(x < \overline{n} + 1 \leftrightarrow x = \overline{0} \vee \ldots \vee x = \overline{n})$.

- Now we will prove, by induction on terms, that PA $\vdash t(\overline{n_1}, \ldots, \overline{n_k}) = \overline{t^{\mathcal{N}}(n_1, \ldots, n_k)}$.
  - If $t = c$ for some constant c then $t = 0 \vee t = 1$, as these are the only constants. Now $\overline{0} = 0$ and $\overline{1} = 0 + 1$, and of course $0^{\mathcal{N}} = 0$ and $1^{\mathcal{N}} = 1$, so $\overline{0^{\mathcal{N}}} = 0$ and $\overline{1^{\mathcal{N}}} = 0 + \overline{1}$.
  - If $t = x$ for some variable x, then if $x^{\mathcal{N}}$ is its interpretation in $\mathcal{N}$, then by definition $\overline{x^{\mathcal{N}}} = x$.
  - If t_1, t_2 are terms for which PA $\vdash t(\overline{n_1}, \ldots, \overline{n_k}) = \overline{t^{\mathcal{N}}(n_1, \ldots, n_k)}$ holds. Then PA $\vdash (\overline{t_1} + \overline{t_2}) = \overline{t_1 +^{\mathcal{N}} t_2}$ and PA $\vdash (\overline{t_1} \cdot \overline{t_2}) = \overline{t_1 \cdot^{\mathcal{N}} t_2}$, with the obvious interpretation of + and · in $\mathcal{N}$.
- Now we prove, by induction on Δ_0 formulae that PA is Δ_0 complete:
  - For atomic formulae it holds by the first part of the proof.
  - If φ and ψ are formulae for which PA $\vdash \varphi(n_1, \ldots, n_k) \Leftrightarrow \mathbb{N} \models \varphi(n_1, \ldots, n_k)$ and PA $\vdash \psi(m_1, \ldots, m_l) \Leftrightarrow \mathbb{N} \models \psi(m_1, \ldots, m_l)$ holds. Then we see that also the following hold:

$$\begin{aligned} \text{PA} \vdash & \varphi(n_1, \ldots, n_k) \wedge \psi(m_1, \ldots, m_l) \\ & \quad \Leftrightarrow \mathbb{N} \models \varphi(n_1, \ldots, n_k) \wedge \psi(m_1, \ldots, m_l) \\ \text{PA} \vdash & \varphi(n_1, \ldots, n_k) \vee \psi(m_1, \ldots, m_l) \\ & \quad \Leftrightarrow \mathbb{N} \models \varphi(n_1, \ldots, n_k) \vee \psi(m_1, \ldots, m_l) \\ \text{PA} \vdash & \varphi(n_1, \ldots, n_k) \to \psi(m_1, \ldots, m_l) \\ & \quad \Leftrightarrow \mathbb{N} \models \varphi(n_1, \ldots, n_k) \to \psi(m_1, \ldots, m_l) \end{aligned} \tag{69}$$

    For the first one we see that for the conjunction to be true in $\mathbb{N}$, they both need to be true in $\mathbb{N}$, which is equivalent (by induction hypothesis) with both of them being provable in PA, which is equivalent with the conjunction being provable in PA.

    For the second one we see that for the disjunction to be true in $\mathbb{N}$, at least one of them needs to be true in $\mathbb{N}$, which is equivalent (by induction hypothesis) with at least one of them being provable in PA, which is equivalent with the disjunction being provable in PA.

    For the last one we see that for the implication to be true in $\mathbb{N}$, ψ needs to be true in $\mathbb{N}$ (which is equivalent, by induction hypothesis, to ψ being provable in PA) or $\neg\varphi$ should be true in $\mathbb{N}$ (which is equivalent, by induction hypothesis, to $\neg\varphi$ being provable in PA), so this is equivalent with the implication being provable in PA.
  - Now only the induction step for bounded quantifiers is left. We know that PA $\vdash t = \overline{t^{\mathcal{N}}}$. So we may substitute this in PA $\vdash \forall x(x < \overline{n} \leftrightarrow x = \overline{0} \vee \ldots \vee x = \overline{n-1})$, to find that PA $\vdash \forall x(x < t \leftrightarrow x = \overline{0} \vee \ldots \vee y = \overline{t^{\mathcal{N}} - 1})$.

    Let $\varphi = \forall x < t\psi$, with PA $\vdash \psi \Leftrightarrow \mathbb{N} \models \psi$ and x is not a free variable of t. Then we know PA $\vdash \forall x < t\psi \leftrightarrow \psi(\overline{0}) \wedge \ldots \wedge \psi(\overline{t^{\mathcal{N}} - 1})$. Now we have seen that PA $\vdash \varphi(n_1, \ldots, n_k) \wedge \psi(m_1, \ldots, m_l) \Leftrightarrow \mathbb{N} \models \varphi(n_1, \ldots, n_k) \wedge \psi(m_1, \ldots, m_l)$ and repeated application of this gives us

$$\begin{aligned} \mathrm{PA} \vdash \varphi \Leftrightarrow \mathrm{PA} \vdash \forall x < t\psi \Leftrightarrow \mathrm{PA} \vdash \psi(\overline{0}) \wedge \ldots \wedge \psi(\overline{t^{\mathcal{N}} + 1}) \\ \Leftrightarrow \mathbb{N} \models \psi(\overline{0}) \wedge \ldots \wedge \psi(\overline{t^{\mathcal{N}} + 1}) \\ \Leftrightarrow \mathbb{N} \models \forall x < t\psi \\ \Leftrightarrow \mathbb{N} \models \varphi \end{aligned} \tag{70}$$

Now let $\varphi = \exists x < t\psi$, with $\mathrm{PA} \vdash \psi \Leftrightarrow \mathbb{N} \models \psi$ and x is not a free variable of t. Then we know $\mathrm{PA} \vdash \exists x < t\psi \leftrightarrow \psi(\overline{0}) \vee \ldots \vee \psi(\overline{t^{\mathcal{N}} - 1})$. Now we have seen that $\mathrm{PA} \vdash \varphi(n_1, \ldots, n_k) \vee \psi(m_1, \ldots, m_l) \Leftrightarrow \mathbb{N} \models \varphi(n_1, \ldots, n_k) \vee \psi(m_1, \ldots, m_l)$ and repeated application of this gives us

$$\begin{aligned} \mathrm{PA} \vdash \varphi \Leftrightarrow \mathrm{PA} \vdash \exists x < t\psi \Leftrightarrow \mathrm{PA} \vdash \psi(\overline{0}) \vee \ldots \vee \psi(\overline{t^{\mathcal{N}} + 1}) \\ \Leftrightarrow \mathbb{N} \models \psi(\overline{0}) \vee \ldots \vee \psi(\overline{t^{\mathcal{N}} + 1}) \\ \Leftrightarrow \mathbb{N} \models \exists x < t\psi \\ \Leftrightarrow \mathbb{N} \models \varphi \end{aligned} \tag{71}$$

- Now we prove that this also holds for Σ_1 formulae by proving this for unbounded $\exists$. Let $\varphi = \exists x\psi$, where $\psi \in \Delta_0$. Then, because $\mathcal{N}$ is a model of PA, we trivially have that $\mathrm{PA} \vdash \varphi \Rightarrow \mathbb{N} \models \varphi$. Now for the other implication we have the following, using that $\mathbb{N} \models \psi \Rightarrow \mathrm{PA} \vdash \psi$, which we know because $\psi \in \Delta_0$:

$$\begin{aligned} \mathbb{N} \models \varphi \Rightarrow \mathbb{N} \models \exists x\psi(x) \\ \Rightarrow \exists n \mathbb{N} \models \psi(\overline{n}) \\ \Rightarrow \exists n \mathrm{PA} \vdash \psi(\overline{n}) \\ \Rightarrow \mathrm{PA} \vdash \exists x\psi(x) \\ \Rightarrow \mathrm{PA} \vdash \varphi \end{aligned} \tag{72}$$

□

Definition 38. An $\mathcal{L}_{\mathrm{PA}}$-formula $\varphi(x_1, \ldots, x_k)$ is said to *represent (numeralwise)* the k-ary relation $A \subseteq \mathbb{N}^k$ if for all $n_1, \ldots, n_k \in \mathbb{N}$ we have:

$$\begin{aligned} (n_1, \ldots, n_k) \in A \Rightarrow \mathrm{PA} \vdash \varphi(\overline{n_1}, \ldots, \overline{n_k}) \\ (n_1, \ldots, n_k) \notin A \Rightarrow \mathrm{PA} \vdash \neg\varphi(\overline{n_1}, \ldots, \overline{n_k}) \end{aligned} \tag{73}$$

An $\mathcal{L}_{\mathrm{PA}}$-formula $\varphi(x_1, \ldots, x_k, z)$ is said to *represent (numeralwise)* the k-ary function $F : \mathbb{N}^k \to \mathbb{N}$ if for all $n_1, \ldots, n_k \in \mathbb{N}$ we have:

$$\begin{aligned} \mathrm{PA} \vdash \varphi(\overline{n_1}, \ldots, \overline{n_k}, \overline{F(n_1, \ldots, n_k)}) \\ \mathrm{PA} \vdash \exists!\, z\varphi(\overline{n_1}, \ldots, \overline{n_k}, z) \end{aligned} \tag{74}$$

A relation or function is Σ_1-represented if the formula φ representing it is a Σ_1-formula.

Lemma 39 (Collection Principle)**.** In PA we have the following principle:

$$\mathrm{PA} \vdash \forall i < t\exists v\psi(i, v) \to \exists w\forall i < t\exists v < w\psi(i, v) \tag{75}$$

and because of the above, we have that if φ is equivalent to a Σ_1-formula, so is $\forall i < t\varphi$.

Theorem 40 (Δ_1 provable recursion)**.** For every primitive recursive function $F : \mathbb{N}^k \to \mathbb{N}$ there is a Δ_1-formula $\varphi_F(x_1, \dots, x_{k+1})$ which represents F and is such that

$$\mathrm{PA} \vdash \forall x_1, \dots, x_k \exists!\, x_{k+1} \varphi_F(x_1, \dots, x_{k+1}) \tag{76}$$

Proof. We prove this by induction on primitive recursive functions.

- The basic functions: $\lambda x_1 \cdots x_k.0$ is represented by the formula $x_{k+1} = 0$. We see that indeed for every $x_1, \dots, x_k$ there is a unique x_{k+1} such that $x_{k+1} = 0$. This is trivial, as we can always pick $x_{k+1} = 0$ and 0 is unique.

  $\lambda x_1.x_1 + 1$ is represented by the formula $x_2 = x_1 + 1$. For every x_1 we know that $x_1 + 1$ exists. It is trivially also unique.

  $\lambda x_1 \cdots x_k.x_i$ is represented by the formula $x_{k+1} = x_i$. Indeed for every $x_1, \dots, x_k$ there is a x_{k+1} such that $x_{k+1} = x_i$ and of course it is unique.

- composition: let F be defined by composition, so:

  $$F(\vec{x}) = G(H_1(\vec{x}), \dots, H_m(\vec{x})) \tag{77}$$

  with G, $H_1, \dots, H_m$ Δ_1 provable recursive (by induction hypothesis) and represented by Δ_1-formulae $\chi, \psi_1, \dots, \psi_m$ respectively. Then F is represented by the following formula:

  $$\varphi(\vec{x}, z) \equiv \exists z_1 \cdots z_m(\psi_1(\vec{x}, z_1) \wedge \cdots \wedge \psi_m(\vec{x}, z_m) \wedge \chi(z_1, \dots, z_m, z)) \tag{78}$$

  Now φ is equivalent to a Σ_1-formula. We know that $\psi_1, \dots, \psi_m, \chi$ are all equivalent to Σ_1-formulae, this is a finite amount of formula, and they all have a finite amount of existential quantifiers. By applying the following equivalence finite number of times, we obtain a Σ_1-formula:

  $$\varphi \wedge \exists x \psi \leftrightarrow \exists x(\varphi \wedge \psi) \tag{79}$$

  Now we still have to prove that $\neg\varphi$ is also equivalent to a Σ_1-formula.

  We know $\psi_i(\vec{x}, z_i) \leftrightarrow \exists \vec{y_i} \zeta_i(\vec{y_i}, \vec{x}, z_i)$ with ζ_i a Δ_0-formula, for all $1 \leq i \leq m$, and $\chi(z_1, \dots, z_m, z) \leftrightarrow \forall \vec{h} \eta(\vec{h}, z_1, \dots, z_m, z)$ with η a Δ_0-formula. The equivalences for ψ_i follow from the fact that all ψ_i are Σ_1, the equivalence for χ follows from the fact that χ is Δ_1, which in particular means that $\neg\chi$ is Σ_1, the equivalence stated above then follows from the property $\neg\exists x\varphi \leftrightarrow \forall x \neg\varphi$ and that the negation of a Δ_0-formula is again a Δ_0-formula. Then we have the following equivalence:

  $$\begin{aligned}\varphi(\vec{x}, z) \equiv& \exists z_1 \cdots z_m(\psi_1(\vec{x}, z_1) \wedge \cdots \wedge \psi_m(\vec{x}, z_m) \wedge \chi(z_1, \dots, z_m, z)) \\ \leftrightarrow& \forall z_1, \dots, z_m(\psi_1(\vec{x}, z_1) \wedge \cdots \wedge \psi_m(\vec{x}, z_m) \to \chi(z_1, \dots, z_m, z)) \\ \leftrightarrow& \forall z_1, \dots, z_m(\exists \vec{y_1} \zeta_1(, \vec{y_1}, \vec{x}, z_1) \wedge \cdots \wedge \exists \vec{y_m} \zeta_m(\vec{y_m}, \vec{x}, z_m) \to \\ & \forall \vec{h} \eta(\vec{h}, z_1, \dots, z_m, z)) \\ \leftrightarrow& \forall z_1, \dots, z_m, \vec{y_1}, \dots, \vec{y_m}, \vec{h}(\zeta_1(\vec{y_1}, \vec{x}, z_1) \wedge \cdots \wedge \zeta_m(\vec{y_m}, \vec{x}, z_m) \to \\ & \eta(\vec{h}, z_1, \dots, z_m, z))\end{aligned} \tag{80}$$

So by the equivalence $\neg\forall x\varphi \leftrightarrow \exists x\neg\varphi$ we see that indeed the negation of φ is again a Σ_1-formula:

$$\begin{gathered}\neg\varphi(\vec{x},z) \leftrightarrow \exists z_1,\dots,z_m,\vec{y_1},\dots,\vec{y_m},\vec{h} \\ (\neg\eta(\vec{h},z_1,\dots,z_m,z) \wedge \zeta_1(\vec{y_1},\vec{x},z_1) \wedge \dots \wedge \zeta_m(\vec{y_m},\vec{x},z_m))\end{gathered} \tag{81}$$

So φ is equivalent to a Δ_1-formula. We know that F is Δ_1 provable recursive if we show that PA $\vdash \forall\vec{x}\exists!\, z\varphi(\vec{x},z)$.

Say there is no z such that $\varphi(\vec{x},z)$, for some $\vec{x}$. We know that $\exists!\, z_i\psi_i(\vec{x},z_i)$ for $1 \leq i \leq m$ by induction hypothesis for ψ_i. Now our assumption is that for $z_1,\dots,z_m$ we have $\forall z\neg\chi(z_1,\dots,z_m,z)$. This is in contradiction with our induction hypothesis for χ that $\forall\vec{x}\exists!\, z\chi(\vec{x},z)$. Conclude $\forall\vec{x}\exists z\varphi(\vec{x},z)$.

Now suppose we have z and w, $w \neq z$ and $\varphi(\vec{x},z)\wedge\varphi(\vec{x},w)$ for some $\vec{x}$. By induction hypothesis for ψ_i we know that all z_i are uniquely defined. So we have $\chi(z_1,\dots,z_m,z)\wedge\chi(z_1,\dots,z_m,w)$ which is in contradiction with our hypothesis for χ that z is uniquely defined. We conclude $\forall\vec{x}\exists!\, z\varphi(\vec{x},z)$. So F is Δ_1 provable recursive.

- primitive recursion: let F be defined by primitive recursion, so:

$$F(\vec{x},0) = G(\vec{x}) \text{ and } F(\vec{x},y+1) = H(\vec{x},F(\vec{x},y),y) \tag{82}$$

with G and H both Δ_1 provable recursive, represented by $\chi(\vec{x},z)$ and $\psi(\vec{x},u,v,w)$ respectively. Then F is represented by the following formula Φ:

$$\begin{aligned}\Phi \equiv \exists am(&\chi(\vec{x},(a,m)_0)\wedge \\ &\forall i < y\psi(\vec{x},(a,m)_i,i,(a,m)_{i+1})\wedge(a,m)_y = u)\end{aligned} \tag{83}$$

As before, this should just be seen as an *abbreviation*, as there is no term $(a,m)_i$ in $\mathcal{L}_{\text{PA}}$. These abbreviations have been introduced in section 3.2.

First we will show that this formula is indeed the representation of F. For this we have to prove that for every $\vec{n},p$:

$$\begin{aligned}&\text{PA} \vdash \Phi(\vec{\overline{n}},\overline{k},\overline{p}) \Leftrightarrow F(\vec{n},k) = p \\ &\text{PA} \vdash \exists!\, u\Phi(\vec{\overline{n}},\overline{k},u)\end{aligned} \tag{84}$$

For the first property, we will do this by induction on k. The second property will be proved while proving that F is Δ_1 provable recursive (it is a weaker condition than the condition for being Δ_1 provable recursive). For $k = 0$ we see $\Phi(\vec{\overline{n}},\overline{0},\overline{p}) = \exists am(\chi(\vec{\overline{n}},(a,m)_0)\wedge(a,m)_0 = \overline{p})$ which means we have $\chi(\vec{\overline{n}},p)$. So PA $\vdash \Phi(\vec{\overline{n}},\overline{0},\overline{p}) \Leftrightarrow$ PA $\vdash \chi(\vec{\overline{n}},p)$, so $G(\vec{n}) = p$, so $F(\vec{n},0) = p$.

For the other implication, say $F(\vec{n},0) = p$, then $G(\vec{n}) = p$ so PA $\vdash \chi(\vec{\overline{n}},\overline{p})$. By the first property of theorem 7, we know that for every k there is a,m such that $(a,m)_0 = k$. This means we have PA $\vdash \chi(\vec{\overline{n}},\overline{p}) \Rightarrow$ PA $\vdash \exists am(\chi(\vec{\overline{n}},(a,m)_0)\wedge(a,m)_0 = \overline{p})$, so PA $\vdash \Phi(\vec{\overline{n}},\overline{0},\overline{p})$.

Now for the induction step, suppose we have PA $\vdash \Phi(\vec{\overline{n}},\overline{k},\overline{p}) \Leftrightarrow F(\vec{n},k) = p$.

Suppose $F(\vec{n}, k+1) = q$. We know that PA $\vdash \exists am(\chi(\vec{\overline{n}}, (a,m)_0) \wedge \forall i < k\psi(\vec{\overline{n}}, (a,m)_i, i, (a,m)_{i+1}) \wedge (a,m)_k = \overline{p})$.

By the second property of theorem 7, we know that there are a', m' such that $(a,m)_i = (a',m')_i$ for all $i < k+1$ and $(a',m')_{k+1} = \overline{q}$.

Because $F(\vec{n}, k+1) = q$, we have $H(\vec{n}, p, k) = q$, which means we have PA $\vdash \psi(\vec{\overline{n}}, \overline{p}, k, \overline{q})$,so we know that PA $\vdash \psi(\vec{\overline{n}}, (a',m')_k, k, (a',m')_{k+1})$.

So PA $\vdash \exists a'm'(\chi(\vec{\overline{n}}, (a',m')_0) \wedge \forall i < k+1\psi(\vec{\overline{n}}, (a',m')_i, i, (a',m')_{i+1}) \wedge (a',m')_{k+1} = \overline{q})$.

For the other implication, suppose we have PA $\vdash \exists a'm'(\chi(\vec{\overline{n}}, (a',m')_0) \wedge \forall i < k+1\psi(\vec{\overline{n}}, (a',m')_i, i, (a',m')_{i+1}) \wedge (a',m')_{k+1} = \overline{q})$.

Because $(a',m')_0$ and $(a',m')_{i+1}$ are uniquely determined for every i (as we will see below, while proving that F is provable Δ_1 represented), we know that $(a',m')_{k+1}$ is also uniquely determined. This means we have PA $\vdash \psi(\vec{\overline{n}}, (a',m')_k, k, \overline{q})$, so $H(\vec{n}, p, k) = q$, so $F(\overline{n}, k+1) = q$.

Now we prove that this formula is a Δ_1-formula. G and H are represented by Δ_1-formulae, so there are Δ_0-formulae χ' and ψ' such that $\chi(\vec{x}, z) \leftrightarrow \exists r\chi'(\vec{x}, z, r)$ and $\psi(\vec{x}, u, v, w) \leftrightarrow \exists s\psi'(\vec{x}, u, v, w, s)$. Now Φ becomes:

$$\begin{aligned} &\exists am(\exists r\chi'(\vec{x}, (a,m)_0, r) \\ &\qquad \wedge \forall i < y\exists s\psi'(\vec{x}, (a,m)_i, i, (a,m)_{i+1}, s) \wedge (a,m)_y = u) \end{aligned} \tag{85}$$

which is equivalent, using the Collection Principle (39) to:

$$\begin{aligned} &\exists amrs(\chi'(\vec{x}, (a,m)_0, r) \wedge \\ &\qquad \forall i < y\psi'(\vec{x}, (a,m)_i, i, (a,m)_{i+1}, s) \wedge (a,m)_y = u) \end{aligned} \tag{86}$$

Now χ' and ψ' are Δ_0, $\forall i < y\psi'$ is also Δ_0 as the quantifier is bounded, and $(a,m)_y = u$ is Δ_0 as this is just an abbreviation of an arithmetical statement. The conjunction of the above is therefore also Δ_0, which means Φ is equivalent to a Σ_1-formula.

Because G and H are Δ_1 provable recursive, we have PA $\vdash \forall\vec{x}\exists!\, z\chi(\vec{x}, z)$ and PA $\vdash \forall\vec{x}uv\exists!\, w\psi(\vec{x}, u, v, w)$. This means we also have:

$$\begin{aligned} &\text{PA} \vdash \forall fg(\chi(\vec{x}, f) \wedge \chi(\vec{x}, g) \to f = g) \\ &\text{PA} \vdash \forall fg(\psi(\vec{x}, u, v, f) \wedge \psi(\vec{x}, u, v, g) \to f = g) \end{aligned} \tag{87}$$

This means that also a, m are uniquely determined, so Φ is equivalent to the following formula:

$$\forall am(\chi(\vec{x}, (a,m)_0) \wedge \forall i < y\psi(\vec{x}, (a,m)_i, i, (a,m)_{i+1}) \to (a,m)_y = u) \tag{88}$$

Now, using again the Collection Principle and the basic rule $(\exists xf(x) \to g) \leftrightarrow (\forall x(f(x) \to g))$, we have that Φ is also equivalent to:

$$\begin{aligned} &\forall amrs(\chi'(\vec{x}, (a,m)_0, r) \wedge \\ &\qquad \forall i < y\psi'(\vec{x}, (a,m)_i, i, (a,m)_{i+1}, s) \to (a,m)_y = u) \end{aligned} \tag{89}$$

This is a Π_1-formula, which means the negation of Φ is a Σ_1-formula, which means Φ is indeed a Δ_1-formula.

And, to show that F is Δ_1 provable recursive, we will show that PA $\vdash$ $\forall \vec{x}, y \exists!\, u \varphi(\vec{x}, y, u)$.

Let $y = 0$. Then we know that $\exists!\, z \chi(\vec{x}, z)$ because G is Δ_1 provable recursive. By the first property of theorem 7 we know that $\forall z \exists am((a,m)_0 = z)$. Now take $u = z$, then u is uniquely defined and we have (as there are no $i < 0$) that:

$$\exists am(\chi(\vec{x}, (a,m)_0) \wedge \forall i < 0 \psi(\vec{x}, (a,m)_i, i, (a,m)_{i+1}) \wedge (a,m)_0 = u \quad (90)$$

Now say the formula holds for y. We want to show it also holds for $y+1$. Use that for H we know that for $\vec{x}, (a,m)_y, y$ there exists a unique z such that $\psi(\vec{x}, (a,m)_y, y, z)$. Let $u = z$ then u exists and is uniquely determined. For this we use the second property of theorem 7. By this property we know that $\forall amuy$ we can find b, n such that $\forall i \leq y((a,m)_i = (b,n)_i) \wedge (b,n)_{y+1} = u$. So we have:

$$\exists bn(\chi(\vec{x}, (b,n)_0) \wedge \forall i \leq y \psi(\vec{x}, (b,n)_i, i, (b,n)_{i+1}) \wedge (b,n)_{y+1} = u \quad (91)$$

So by induction we see that such a u always exists and is uniquely determined because G and H are Δ_1 provable recursive. So F is Δ_1 provable recursive.

□

7.2 Coding

Gödel presented a coding of terms, formulae and proofs in PA. This coding will be used to prove the *Diagonalisation lemma*, an essential property in the proof of the first incompleteness theorem.

7.2.1 Coding of Terms and Formulae

Remember the definition of the *coding of sequences* given in theorem 27, which was primitive recursive. We use this to assign to every formula φ in $\mathcal{L}_{\text{PA}}$ a code $\ulcorner n \urcorner \in \mathbb{N}$. We will see that this is done in a way which translates all relevant operations on formulae into primitive recursive functions on codes.

We assume that variables are numbered $(v_0, v_1, \dots)$, and we take $<$ as a primitive symbol of $\mathcal{L}_{\text{PA}}$. We define the following:

0	1	v	$+$	$\cdot$	$=$	$<$	$\wedge$	$\vee$	$\rightarrow$	$\neg$	$\forall$	$\exists$
0	1	2	3	4	5	6	7	8	9	10	11	12

First we define the coding of terms, by recursion on the term t:

- $\ulcorner 0 \urcorner = \langle 0 \rangle$
- $\ulcorner 1 \urcorner = \langle 1 \rangle$
- $\ulcorner v_i \urcorner = \langle 2, i \rangle$
- $\ulcorner t + s \urcorner = \langle 3, \ulcorner t \urcorner, \ulcorner s \urcorner \rangle$
- $\ulcorner t \cdot s \urcorner = \langle 4, \ulcorner t \urcorner, \ulcorner s \urcorner \rangle$

We define the coding of formulae by recursion as well:

- $\ulcorner t = s \urcorner = \langle 5, \ulcorner t \urcorner, \ulcorner s \urcorner \rangle$
- $\ulcorner t < s \urcorner = \langle 6, \ulcorner t \urcorner, \ulcorner s \urcorner \rangle$
- $\ulcorner \varphi \wedge \psi \urcorner = \langle 7, \ulcorner \varphi \urcorner, \ulcorner \psi \urcorner \rangle$
- $\ulcorner \varphi \vee \psi \urcorner = \langle 8, \ulcorner \varphi \urcorner, \ulcorner \psi \urcorner \rangle$
- $\ulcorner \varphi \to \psi \urcorner = \langle 9, \ulcorner \varphi \urcorner, \ulcorner \psi \urcorner \rangle$
- $\ulcorner \neg \varphi \urcorner = \langle 10, \ulcorner \varphi \urcorner \rangle$
- $\ulcorner \forall v_i \varphi \urcorner = \langle 11, \ulcorner v_i \urcorner, \ulcorner \varphi \urcorner \rangle$
- $\ulcorner \exists v_i \varphi \urcorner = \langle 12, \ulcorner v_i \urcorner, \ulcorner \varphi \urcorner \rangle$

Now many properties are primitive recursive, we will show that for the following example, the proof of which also includes many other examples:

Lemma 41 (Substitution)**.** There is a primitive recursive function Sub, such that

$$Sub(x, y, i) = \begin{cases} \ulcorner \varphi[s/v_i] \urcorner \text{ if } y = \ulcorner \varphi \urcorner \text{ and } x = \ulcorner s \urcorner \\ \qquad 0 \text{ else} \end{cases} \tag{92}$$

Proof. We will prove that 'y codes a formula', 'x codes a term', 'v_i is free in φ and variables in s are not bound in φ when s is substituted for v_i' are primitive recursive. Then the set of x, y for which x codes a term and y codes a formula is primitive recursive as well: it is the intersection of two primitive recursive sets, so we may add their characteristic function to obtain its characteristic function, and this is primitive recursive because it is the composition of addition (which is primitive recursive) and two primitive recursive functions. So that set is primitive recursive and 'v_i is free in φ and variables in s are not bound in φ when s is substituted for v_i' as well, and by applying theorem 29 we see that Sub is primitive recursive.

Before we prove the properties, let us observe that if a is defined by a code of a sequence with at least two elements, e.g. $a = \langle b, c \rangle$, then we always have $a > b \wedge a > c$.

First we prove that 'x codes a term' is a primitive recursive property. This is equivalent with proving that the characteristic function $\chi_t(x)$ is primitive recursive. Now we have the following equivalence, which follows from the recursive definition of terms:

$$\begin{aligned} \chi_t(x) \Leftrightarrow & x = \langle 0 \rangle \vee x = \langle 1 \rangle \\ & \vee \exists i < x (x = \langle 2, i \rangle) \\ & \vee \exists ij < x (\chi_t(i) \wedge \chi_t(j) \wedge x = \langle 3, i, j \rangle) \\ & \vee \exists ij < x (\chi_t(i) \wedge \chi_t(j) \wedge x = \langle 4, i, j \rangle) \end{aligned} \tag{93}$$

We know that $\langle 0 \rangle, \langle 1 \rangle, \langle 2, i \rangle, \langle 3, i, j \rangle$ and $\langle 4, i, j \rangle$ are primitive recursive by the properties of $\langle \rangle$. So we see that $\chi_t(x)$ is defined by course-of-values recursion (see definition 28), from the composition of primitive recursive functions. So it is primitive recursive.

Now we will prove that 'y codes a formula' is primitive recursive. For the characteristic function χ_f of the set of all y that code formulae, we have the following equivalence, which follows from the recursive definition of formulae:

$$\begin{aligned}\chi_f(y) \Leftrightarrow & \exists vw < y(\chi_t(v) \wedge \chi_t(w) \wedge y = \langle 5, v, w\rangle) \\ & \vee \exists vw < y(\chi_t(v) \wedge \chi_t(w) \wedge y = \langle 6, v, w\rangle) \\ & \vee \exists vw < y(\chi_f(v) \wedge \chi_f(w) \wedge y = \langle 7, v, w\rangle) \\ & \vee \exists vw < y(\chi_f(v) \wedge \chi_f(w) \wedge y = \langle 8, v, w\rangle) \\ & \vee \exists vw < y(\chi_f(v) \wedge \chi_f(w) \wedge y = \langle 9, v, w\rangle) \\ & \vee \exists v < y(\chi_f(v) \wedge y = \langle 10, v\rangle) \\ & \vee \exists iv < y(\exists j < i(i = \langle 2, i\rangle) \wedge \chi_f(v) \wedge y = \langle 11, i, v\rangle) \\ & \vee \exists iv < y(\exists j < i(i = \langle 2, i\rangle) \wedge \chi_f(v) \wedge y = \langle 12, i, v\rangle)\end{aligned} \tag{94}$$

We know that χ_t is primitive recursive, as well as $\langle\rangle$, so we see that also $\chi_f(y)$ is defined by course-of-values recursion from the composition of primitive recursive functions, and is therefore primitive recursive.

The last one to prove is that 'v_i is free in φ and no free variables in s are bound in φ when s is substituted for v_i' is primitive recursive. We see that these are in fact two properties, and we can show they are primitive recursive seperately, as we can just add their characteristic functions to obtain the characteristic function of the property asked for. We will define the characteristic function χ_b for v_i is bound in φ, its negation will then be v_i is free in φ.

$$\begin{aligned}\chi_b(v_i, \varphi) \Leftrightarrow & \varphi = \langle 11, \ulcorner v_i \urcorner, \ulcorner \psi \urcorner\rangle \vee \varphi = \langle 12, \ulcorner v_i \urcorner, \ulcorner \psi \urcorner\rangle \\ & \vee (\varphi = \langle 7, \ulcorner \psi \urcorner, \ulcorner \psi' \urcorner\rangle \wedge (\chi_b(v_i, \psi) \vee \chi_b(v_i, \psi'))) \\ & \vee \ldots \vee (\varphi = \langle 10, \ulcorner \psi \urcorner\rangle \wedge \chi_b(v_i, \psi))\end{aligned} \tag{95}$$

We see that this function is primitive recursive, as it is defined from primitive recursive functions by course-of-values recursion. So its negation, $\chi_f(v_i, \varphi)$ for v_i is free in φ, is also primitive recursive.

For the second part, no free variables in s are bound φ when s is substituted for v_i we define the characteristic function χ_s:

$$\chi_s(s, \varphi) \Leftrightarrow (\chi_f(v_j, s) \rightarrow \chi_f(v_j, \varphi)) \tag{96}$$

We see that this is a composition of primitive recursive functions, so it is primitive recursive. The characteristic function for the property as a whole becomes

$$\chi_f(v_i, \varphi) + \chi_s(s, \varphi) \tag{97}$$

which is again a primitive recursive function.

□

7.2.2 Coding of Proofs

Also for the construction steps of proof trees one can find a coding like the one above. I will only discuss the construction steps I have introduced before. For simplicity I will just number them from 0 again. For a more complete overview of the coding of proofs, see [van Oosten, 2009].

Ass	$\wedge I$	$\forall E$
0	1	2

The coding of proof trees is done by induction on the construction of proof trees. This means for example that we have

- $\langle 0, \ulcorner\varphi\urcorner\rangle$ is the code of the tree with one node, the assumption tree.
- $\langle 1, \ulcorner D_1\urcorner, \ulcorner D_2\urcorner, \ulcorner\varphi\wedge\psi\urcorner\rangle$ is the code of the tree resulting from the trees D_1 and D_2 with roots φ and ψ respectively, by applying $\wedge I$.
- $\langle 2, \ulcorner D\urcorner, \ulcorner\varphi[s/t]\urcorner\rangle$, where the root of D is $\forall t\varphi(t)$ is the code of the tree resulting from D by applying $\forall E$.

Remember the example of the proof tree given in section 4:

$$\frac{\forall E\,\dfrac{\forall x\varphi(x)}{\varphi[t/x]} \qquad \psi}{\varphi[t/x]\wedge\psi}\,\wedge I$$

For this proof tree I will present its coding as an example to convince you of this method. The code of this proof tree D is:

$$\begin{aligned}
\ulcorner D\urcorner &= \langle 1, \ulcorner D_1\urcorner, \ulcorner D_2\urcorner, \ulcorner\varphi[t/x]\wedge\psi\urcorner\rangle\\
\text{where } \ulcorner D_2\urcorner &= \langle 0, \ulcorner\psi\urcorner\rangle,\\
\text{and } \ulcorner D_1\urcorner &= \langle 2, \ulcorner D_3\urcorner, \ulcorner\varphi[t/x]\urcorner\rangle,\\
\text{where } \ulcorner D_3\urcorner &= \langle 0, \ulcorner\forall x\varphi(x)\urcorner\rangle.
\end{aligned} \tag{98}$$

We see that this coding is also primitive recursive. Using this, one can prove that some special properties that we will define here are also primitive recursive. These properties we will use in section 7.3.

Definition 42. We define functions OA, giving the set of undischarged (not eliminated) assumptions of a proof tree, NDT (natural deduction tree) and Ax:

- $\mathrm{OA}(x, y) \leftrightarrow x$ is the code of a natural deduction tree and y is the code of an undischarged assumption of the tree coded by x.
- $\mathrm{NDT}(x, y) \leftrightarrow y$ is the code of a formula and x is the code of a correct natural deduction tree with root labelled by the formula coded by y.
- $\mathrm{Ax}(x) \leftrightarrow x$ is the code of an axiom of PA or the predicate calculus (governing the equality sign).

Lemma 43. The function OA and the predicates NDT and Ax are primitive recursive.

Proof.

- For $\mathrm{OA}(x, y)$, its primitive recursiveness is proved by course-of-values recursion. We will, again, not show all construction steps for proof trees. We will show the steps presented above and a few steps with additional difficulties. These are negation introduction ($\neg I$) and implication introduction ($\rightarrow I$). These are examples of steps in which assumptions are being discharged, which means the set of undischarged assumptions changes.

We have the following partial definition for OA:

$$\mathrm{OA}(x,y) \leftrightarrow \begin{array}{l}(x = \langle 0, \ulcorner\varphi\urcorner\rangle \wedge y = \ulcorner\varphi\urcorner)\vee \\ (x = \langle 1, \ulcorner D_1\urcorner, \ulcorner D_2\urcorner, \ulcorner\varphi \wedge \psi\urcorner\rangle\wedge \\ (\mathrm{OA}(\ulcorner D_1\urcorner, y) \vee \mathrm{OA}(\ulcorner D_2\urcorner, y)))\vee \\ (x = \langle 2, \ulcorner D\urcorner, \ulcorner\varphi[s/t]\urcorner\rangle \wedge \mathrm{OA}(\ulcorner D\urcorner, y))\vee \\ (x = \langle \ulcorner\rightarrow I\urcorner, \ulcorner D\urcorner, \ulcorner\varphi \rightarrow \psi\urcorner\rangle \wedge y \neq \ulcorner\varphi\urcorner \wedge \mathrm{OA}(\ulcorner D\urcorner, y))\vee \\ (x = \langle \ulcorner\neg I\urcorner, \ulcorner D\urcorner, \ulcorner\neg\varphi\urcorner\rangle \wedge y \neq \ulcorner\varphi\urcorner \wedge \mathrm{OA}(\ulcorner D\urcorner, y))\end{array} \tag{99}$$

We see that OA is a composition of the primitive recursive coding of trees and coding of formulae, and that OA for lower numbers (because adding a root to a proof tree enlarges its Gödel number) appears in the definition. This means that OA is defined by course-of-values recursion and composition of primitive recursive functions, which means it is primitive recursive. The other construction steps can be formalised in the same manner, so OA is primitive recursive.

- The proof that NDT is primitive recursive will be given in the same way as the proof for OA. We will show its primitive recursiveness for a few construction steps, the ones presented earlier, implication introduction ($\rightarrow I$) and the introduction of the universal quantifier ($\forall I$) as these are again examples of more complicated steps: the former discharges an assumption and for the latter there are additional requirements for the open assumptions and the free variables, as we will see below.

  First we see that the characteristic function for 'the term with code x does not occur in the term with code y' is primitive recursive, defined by course-of-values recursion and composition of primitive recursive functions:

$$\begin{aligned}\chi_{nt}(x,y) \leftrightarrow x \neq y \wedge (y = \langle 0\rangle \vee y = \langle 1\rangle \vee y = \langle 2, i\rangle \\ \vee(y = \langle 3, \ulcorner t\urcorner, \ulcorner s\urcorner\rangle \wedge \chi_{nt}(x, \ulcorner s\urcorner) \wedge \chi_{nt}(x, \ulcorner t\urcorner)) \\ \vee(y = \langle 4, \ulcorner t\urcorner, \ulcorner s\urcorner\rangle \wedge \chi_{nt}(x, \ulcorner s\urcorner) \wedge \chi_{nt}(x, \ulcorner t\urcorner)))\end{aligned} \tag{100}$$

  and we see that the characteristic function for 'the term with code x does not occur in the formula with code y' is therefore also primitive recursive, as it is defined by course-of-values recursion and composition of primitive recursive functions:

$$\begin{aligned}\chi_{nf}(x,y) \leftrightarrow &(y = \langle 5, \ulcorner t\urcorner, \ulcorner s\urcorner\rangle \wedge \chi_{nt}(x, \ulcorner t\urcorner) \wedge \chi_{nt}(x, \ulcorner s\urcorner)) \vee \ldots \vee \\ &(y = \langle 12, \ulcorner v_i\urcorner, \ulcorner\varphi\urcorner\rangle \wedge \chi_{nt}(x, \ulcorner v_i\urcorner) \wedge \chi_{nf}(x, \ulcorner\varphi\urcorner)\end{aligned} \tag{101}$$

Now we give a partial definition of NDT:

$$\mathrm{NDT}(x,y) \leftrightarrow \begin{aligned}
&(x = \langle 0, y\rangle \wedge \chi_f(y))\vee \\
&(x =\langle 1, c, d, e\rangle \wedge \chi_f(e) \wedge e = \langle 7, u, v\rangle\wedge \\
&\quad \mathrm{NDT}(c, u) \wedge \mathrm{NDT}(d, v))\vee \\
&(x =\langle 2, c, d\rangle \wedge \chi_f(d) \wedge d = \mathrm{Sub}(\ulcorner s\urcorner, \ulcorner \varphi\urcorner, i)\wedge \\
&\quad \mathrm{NDT}(c, b) \wedge b = \langle 11, \ulcorner v_i\urcorner, \ulcorner \varphi\urcorner\rangle)\vee \\
&(x =\langle \ulcorner\!\rightarrow I\urcorner, c, d\rangle \wedge \chi_f(d) \wedge d = \langle 9, u, v\rangle\wedge \\
&\quad \mathrm{NDT}(c, v) \wedge \neg\mathrm{OA}(x, u))\vee \\
&(x =\langle \ulcorner\forall I\urcorner, c, d\rangle \wedge \chi_f(d) \wedge d = \langle 11, \ulcorner v_i\urcorner, \ulcorner \varphi\urcorner\rangle\wedge \\
&\quad \mathrm{NDT}(c, b) \wedge b = \mathrm{Sub}(\ulcorner s\urcorner, \ulcorner \varphi\urcorner, i))\wedge \\
&\quad \neg\mathrm{OA}(c, \ulcorner s\urcorner) \wedge \chi_{nf}(\ulcorner s\urcorner, \ulcorner d\urcorner)
\end{aligned} \tag{102}$$

We see that NDT is defined by course-of-values recursion and composition of primitive recursive functions: all functions occuring have been proved to be primitive recursive and the appearances of NDT are for lower numbers, because adding a root enlarges the Gödel number. The other construction steps can be formalised in the same manner, so NDT is primitive recursive.

- $\mathrm{Ax}(x)$ is the characteristic function of the set of the codes of all axioms of PA and the predicate calculus. The axioms of the predicate calculus are the following:
  - $\forall u(u = u)$
  - $\forall uvw(u = v \wedge v = w \rightarrow u = w)$
  - $\forall \varphi tsu(t = s \wedge \varphi[t/u] \rightarrow \varphi[s/u])$

  Now we see that for example for the first two of these properties we have:

$$\mathrm{Ax}(x) = \begin{cases} 0 \text{ if } (x)_0 = 5 \wedge (x)_1 = (x)_2 \\[2em] 0 \text{ if } & \begin{aligned} &\exists ij < x(x = \langle 9, i, j\rangle \wedge \exists kl < j(j = \langle 5, k, l\rangle)\wedge \\ &\qquad \exists mn < i(i = \langle 7, m, n\rangle) \\ &\qquad \wedge\, k < m \wedge n < l\exists t((t < m \wedge t < n) \\ &\qquad \wedge\, (m = \langle 5, k, t\rangle \wedge n = \langle 5, t, l\rangle)) \end{aligned} \\[2em] 1 \text{ else} \end{cases} \tag{103}$$

This can also be done for the axioms. For those I will also give an example. This is for the axioms $\forall x\neg(x + 1 = 0)$ and $\forall x(x \cdot 0 = 0)$

$$\mathrm{Ax}(x) = \begin{cases} 0 \text{ if } (x)_0 = 10 \wedge \exists t < (x)_1((x)_1 = \langle 5, i, \ulcorner 0\urcorner\rangle \wedge i = \langle 3, t, \ulcorner 1\urcorner\rangle) \\ 0 \text{ if } (x)_0 = 5 \wedge (x)_2 = \ulcorner 0\urcorner \wedge \exists t < (x)_1((x)_1 = \langle 4, t, 0\rangle) \\ 1 \text{ else} \end{cases}$$

(104)

Now we see that all these properties are primitive recursive, as they are compositions of $()_i$ and $\langle\rangle$ which are primitive recursive. This can also be shown for the other axioms. So if we define Ax as the characteristic function of the axioms of PA and the predicate calculus, this will also be a primitive recursive function by theorem 29.

□

7.3 The Predicate Prf(x,y)

We will define a new predicate, $\text{Prf}(x,y)$, which says that x is the code of a correct proof of the formula that y codes.

Definition 44. Define $\text{Prf}(x,y)$ with the following equivalence:

$$\text{Prf}(x,y) \leftrightarrow \text{NDT}(x,y) \wedge \forall z \in \text{OA}(x)\text{Ax}(x) \tag{105}$$

From the definition it is clear that Prf is the predicate which says x is a correct proof of y in PA: there is a natural deduction tree that proves y, for which all assumptions which are not eliminated are axioms of PA or of the predicate calculus. We will prove some properties of Prf that we will need in the proof of the first incompleteness theorem.

Lemma 45. $\text{Prf}(x,y)$ is primitive recursive.

Proof. We see $\text{Prf}(x,y)$ is just a composition of primitive recursive functions, so it is primitive recursive. □

As NDT, Ax and Prf are all primitive recursive, we know they are Δ_1-represented. Now let $\overline{\text{Prf}}$, $\overline{\text{NDT}}$ and $\overline{\text{Ax}}$ be Δ_1-formulae representing the predicates Prf, NDT and Ax in PA.

Lemma 46.

1. $\text{PA} \vdash \varphi \Rightarrow \text{PA} \vdash \exists x \overline{\text{Prf}}(x, \overline{\ulcorner\varphi\urcorner})$
2. $\text{PA} \vdash \forall xy(\overline{\text{Prf}}(x, \overline{\ulcorner\varphi \to \psi\urcorner}) \wedge \overline{\text{Prf}}(y, \overline{\ulcorner\varphi\urcorner}) \to \overline{\text{Prf}}(\langle \overline{7}, x, y, \overline{\ulcorner\psi\urcorner}\rangle, \overline{\ulcorner\psi\urcorner}))$

Proof. For the first property:

$\text{PA} \vdash \varphi$ means there is a proof of φ in PA. So there is a proof tree D with conclusion φ, say $\ulcorner D \urcorner = n$. Because it is a proof, we know that $\text{OA}(n,y) \to \text{Ax}(y)$, and that $\text{NDT}(n, \ulcorner\varphi\urcorner)$. So we have $\text{Prf}(n, \overline{\ulcorner\varphi\urcorner})$. We know $\overline{\text{Prf}}$ represents Prf, and because $\overline{\text{Prf}}$ is Σ_1 and PA is Σ_1-complete we may conclude that PA $\vdash \overline{\text{Prf}}(\overline{n}, \overline{\ulcorner\varphi\urcorner})$. So $\text{PA} \vdash \exists x \overline{\text{Prf}}(x, \overline{\ulcorner\varphi\urcorner})$.

For the second property:

This follows from the inference rule for implication-elimination. If we have $\overline{\text{Prf}}(x, \overline{\ulcorner\varphi \to \psi\urcorner}) \wedge \overline{\text{Prf}}(y, \overline{\ulcorner\varphi\urcorner})$, then there is a proof tree for $\varphi \to \psi$ and a proof tree for φ, both of which are correct natural deduction trees with undischarged assumpions consisting only of axioms. We can primitive recursively add another root using the inference rule for implication-elimination. This can be done in a same fashion as the examples given above (i.e. $\forall E$). The resulting tree will again be a correct natural deduction tree with undischarged assumptions consisting only of axioms, so it is a proof, and the implication is shown. □

We introduce the following abbreviation: $\Box\varphi \equiv \exists x\overline{\mathrm{Prf}}(x, \ulcorner\varphi\urcorner)$. Then lemma 46 becomes:

1. $\mathrm{PA} \vdash \varphi \Rightarrow \mathrm{PA} \vdash \Box\varphi$
2. $\mathrm{PA} \vdash \Box(\varphi \to \psi) \wedge \Box\varphi \to \Box\psi$

7.4 The Diagonalization Lemma

To prove the first incompleteness theorem, Gödel invented sentences that are independent of PA. These sentences have been named after him: they are called *Gödel sentences*, we will see them in section 7.6. To prove that these are indeed independent of PA, we will need an important lemma, formulated by Gödel.

We could describe this lemma in words, which will hopefully give some insight in how the lemma works.

> The result of substituting the quotation of "The result of substituting the quotation of x for 'x' in x has property P." for 'x' in "The result of substituting the quotation of x for 'x' in x has property P." has property P.[42]

We will now give the formal formulation and its proof.

Lemma 47 (Diagonalisation Lemma)**.** For any $\mathcal{L}_{\mathrm{PA}}$-formula φ with free variable v_0 there is an $\mathcal{L}_{\mathrm{PA}}$-formula ψ with the same free variables as φ except v_0, such that

$$\mathrm{PA} \vdash \psi \leftrightarrow \varphi[\ulcorner\psi\urcorner/v_0] \tag{106}$$

Moreover, if $\varphi \in \Pi_1$ then ψ can be chosen to be Π_1 as well.

Proof. We know the function $\mathrm{Sub}(x, y, i)$ is primitive recursive by lemma 41. So the function $\lambda xy.\mathrm{Sub}(x, y, 0)$ is primitive recursive as well. So it is Δ_1 provable recursive by theorem 40, so in particular it can be represented by a Σ_1-formula S. Let T be the Σ_1-formula representing the primitive recursive function $n \to \ulcorner\overline{n}\urcorner$. Then we have the following properties, these follow from the definition of representation and Δ_1 provable recursion:

1. $\forall nm \in \mathbb{N}\ \mathrm{PA} \vdash S(\overline{n}, \overline{m}, \overline{\mathrm{Sub}(n, m, 0)})$
2. $\forall n \in \mathbb{N}\ \mathrm{PA} \vdash T(\overline{n}, \overline{\ulcorner\overline{n}\urcorner})$
3. $\mathrm{PA}\vdash \forall xy\exists!\, zS(x, y, z)$
4. $\mathrm{PA}\vdash \forall x\exists!\, yT(x, y)$

Now let φ be an $\mathcal{L}_{\mathrm{PA}}$-formula with free variable v_0. We define the formula C as follows:

$$C \equiv \forall xy(T(v_0, x) \wedge S(x, v_0, y) \to \varphi[y/v_0]) \tag{107}$$

and let ψ be defined as follows:

$$\psi \equiv C[\overline{\ulcorner C\urcorner}/v_0] \tag{108}$$

[42] Cited from [Franzén, 2005] p. 41

We know that T and S are Σ_1, so they are of the form $\exists\vec{u}T'$ and $\exists\vec{w}S'$, so we may write $T(v_0, x) \wedge S(x, v_0, y) \leftrightarrow \exists\vec{u}, \vec{w}(T'(\vec{u}, v_0, x) \wedge S'(\vec{w}, x, v_0, y))$, after renaming variables bound by $\exists$ in S if they are already bound by $\exists$ in T.

Now we have the equivalence $(\exists x\varphi \to \psi) \leftrightarrow \forall x(\varphi \to \psi)$, applying this equivalence a finite number of times gives us

$$\begin{aligned}&\forall xy(T(v_0, x) \wedge S(x, v_0, y) \to \varphi[y/v_0]) \\ &\quad\leftrightarrow \forall xy\vec{u}\vec{v}(T'(\vec{u}, v_0, x) \wedge S'(\vec{w}, x, v_0, y) \to \varphi[y/v_0])\end{aligned} \tag{109}$$

where T' and S' are Δ_0-formulae.

φ is a Π_1-formula, so it is of the form $\forall\vec{z}\varphi'(\vec{z}, y/v_0)$. Because $(\varphi \to \forall x\psi) \leftrightarrow \forall x(\varphi \to \psi)$, we now obtain:

$$C \leftrightarrow \forall xy\vec{u}\vec{w}\vec{z}(T'(\vec{u}, v_0, x) \wedge S'(\vec{w}, x, v_0, y) \to \varphi'(\vec{u}, y/v_0)) \tag{110}$$

This is a Π_1-formula. Now ψ is also a Π_1 formula, as this is just C with a substitution of a variable with a certain (interpretation of a) numeral.

Now by the properties 2 and 4 stated above, we have:

$$\mathrm{PA} \vdash \forall y(\exists x(T(\overline{\ulcorner C \urcorner}, x) \wedge S(x, \overline{\ulcorner C \urcorner}, y)) \leftrightarrow S(\overline{\ulcorner \overline{\ulcorner C \urcorner} \urcorner}, \overline{\ulcorner C \urcorner}, y)) \tag{111}$$

From 2 we derive that $\mathrm{PA} \vdash T(\overline{\ulcorner C \urcorner}, \overline{\ulcorner \overline{\ulcorner C \urcorner} \urcorner})$, then from 4 we see that if $\mathrm{PA} \vdash T(\overline{\ulcorner C \urcorner})$ then $x = \overline{\ulcorner \overline{\ulcorner C \urcorner} \urcorner}$. Now the above stated property follows.

By properties 1 and 3 stated above, we have:

$$\mathrm{PA} \vdash \forall y(S(\overline{\ulcorner \overline{\ulcorner C \urcorner} \urcorner}, \overline{\ulcorner C \urcorner}, y) \leftrightarrow y = \overline{\ulcorner C[\overline{\ulcorner C \urcorner}/v_0] \urcorner}) \tag{112}$$

From 1 we derive that $\mathrm{PA} \vdash S(\overline{\ulcorner \overline{\ulcorner C \urcorner} \urcorner}, \overline{\ulcorner C \urcorner}, \overline{\mathrm{Sub}(\ulcorner \overline{\ulcorner C \urcorner} \urcorner, \ulcorner C \urcorner, 0)})$ and from 3 we know that if $\mathrm{PA} \vdash S(\overline{\ulcorner \overline{\ulcorner C \urcorner} \urcorner}, \overline{\ulcorner C \urcorner}, y)$ then $y = \overline{\mathrm{Sub}(\ulcorner \overline{\ulcorner C \urcorner} \urcorner, \ulcorner C \urcorner, 0)}$.

And from the definition of Sub we derive $\mathrm{Sub}(\ulcorner \overline{\ulcorner C \urcorner} \urcorner, \ulcorner C \urcorner, 0) = \ulcorner C[\overline{\ulcorner C \urcorner}/v_0] \urcorner$, so the above stated property follows.

Now combining these two equivalences and using the definition of ψ, we obtain the following equivalence:

$$\mathrm{PA} \vdash \forall y(\exists x(T(\overline{\ulcorner C \urcorner}, x) \wedge S(x, \overline{\ulcorner C \urcorner}, y)) \leftrightarrow y = \overline{\ulcorner \psi \urcorner}) \tag{113}$$

Now it is straightforward to derive the equivalence stated in the lemma, using the definition of ψ and the above stated equivalence:

$$\begin{aligned}\mathrm{PA} \vdash \psi &\leftrightarrow \forall y(\exists x(T(\overline{\ulcorner C \urcorner}, x) \wedge S(x, \overline{\ulcorner C \urcorner}, y)) \to \varphi[y/v_0] \\ &\leftrightarrow \forall y(y = \overline{\ulcorner \psi \urcorner} \to \varphi[y/v_0]) \\ &\leftrightarrow \varphi[\overline{\ulcorner \psi \urcorner}/v_0]\end{aligned} \tag{114}$$

□

7.5 Proof of the First Theorem

Now that we have seen enough theory to understand the first incompleteness theorem, we will state it again and prove it.

Theorem 48 (First Incompleteness Theorem)**.** There is a Π_1-sentence G such that $\text{PA} \vdash G \leftrightarrow \neg\Box G$. G is independent of PA.

Proof. Let F be the formula $\neg\exists\overline{\text{Prf}}(x, v_0)$. We know that $\overline{\text{Prf}} \in \Delta_1$, so it is equivalent to $\exists\vec{y}\varphi$ for some $\varphi \in \Delta_0$. So $\neg\overline{\text{Prf}}$ is equivalent to $\forall\vec{y}\neg\varphi$. F equals $\forall x\neg\overline{\text{Prf}}(x, v_0)$ so it is equivalent to $\forall x\forall\vec{y}\neg\varphi(\vec{y}, x, v_0)$ which is a Π_1-formula.

Now apply the Diagonalisation lemma to F. F has free variable v_0 and is Π_1, so we can find a formula G which does not have v_0 as a free variable and is also Π_1. For G we have:

$$\text{PA} \vdash G \leftrightarrow \neg\exists x\overline{\text{Prf}}(x, G) \tag{115}$$

In the abbreviation introduced before this becomes:

$$\text{PA} \vdash G \leftrightarrow \neg\Box G \tag{116}$$

Assume $\text{PA} \vdash G$. By the first property of lemma 46 we then have $\text{PA} \vdash \Box G$. From the equivalence that we have for G, we then have $\text{PA} \vdash \neg G$. So $\text{PA} \vdash G$ and $\text{PA} \vdash \neg G$, which means PA is inconsistent. This is a contradiction so PA $\nvdash G$.

Now assume $\text{PA} \vdash \neg G$. Then, by the equivalence we have for G, we have $\text{PA} \vdash \Box G$. We know $\text{PA} \vdash \varphi \Rightarrow \mathbb{N} \models \varphi$ so we know $\mathbb{N} \models \Box G$. This means that there is a proof of G. For this proof one can make a correct proof tree so PA $\vdash G$. So we have $\text{PA} \vdash \neg G$ and $\text{PA} \vdash G$, which means PA is inconsistent, which is again a contradiction so PA $\nvdash \neg G$. This means G is independent of PA, so PA is incomplete. □

The sentence G is true in $\mathbb{N}$. This is because G is a Π_1-formula, which means $\neg G$ is a Σ_1-formula. We know PA is Σ_1-complete, by lemma 37, so this would mean that if $\mathbb{N} \models \neg G$, then $\text{PA} \vdash \neg G$, and we know PA $\nvdash \neg G$ so it is not the case that $\mathbb{N} \models \neg G$, so G is true in $\mathbb{N}$.

7.6 Gödel Sentences

The sentence G is the Gödel sentence we mentioned before. This sentence has been compared with several liar paradoxes, because of the strange equivalence it involves: if it is true in all models, there is no proof, and if there is a proof it is not true in all models.

There have been developed other ways of producing self-referential sentences, like the following, developed by W. Quine:

> "yiels a sentence with property P when appended to its own quotation." yields a sentence with property P when appended to its own quotation.[43]

The Gödel sentence is a sentence G, such that the theory S proves

> G if and only if n is not the Gödel number of a theorem of S,

where n is the Gödel number of G itself.

The formulation in words of the sentences Rosser used to adapt Gödel's proof to cover also other theories than PA (*Rosser sentences*, denote this by R), would be something like this: PA proves

[43] From [Franzén, 2005] p.42

> R if and only if for every n, if n is the Gödel number of a proof of R, then there is an $m < n$ such that m is the Gödel number of a proof of not-R.

In the proof we have just seen, using the Gödel sentence, it is used that $\mathbb{N}$ is a model of PA. If one uses the Rosser sentence to prove the first incompleteness theorem, only the consistency of PA is used to prove the independency of PA of the Rosser sentence. This is why this sentence can be used to prove the first incompleteness theorem also for other theories, as we stated above. We will discuss Rosser sentences again in section 9.3.

7.7 Proof Sketch of the Second Theorem

Now we will take a second look at the second incompleteness theorem.

We know a system is consistent if $\bot$ cannot be derived in it. This means that $\neg\Box\bot$ is actually a formula expressing consistency. This is such an important property that this sentence has been given a name:

$$\mathrm{Con}_{\mathrm{PA}} = \neg\Box\bot \tag{117}$$

We now see that the second incompleteness theorem indeed states that PA does not prove its own consistency:

Theorem 49 (Second Incompleteness Theorem)**.**

$$\mathrm{PA} \nvdash \mathrm{Con}_{\mathrm{PA}} \tag{118}$$

The proof of this theorem will now be presented, assuming one property that has not been proved:

$$\mathrm{PA} \vdash \Box\varphi \to \Box\Box\varphi \tag{119}$$

This property is far from trivial. In fact it is a consequence of a more general property of PA: *Formalised* Σ_1*-completeness*. We will return to this after we have given the proof of the second incompleteness theorem.

Recall the properties of $\Box$ that we proved in section 7.3:

1. $\mathrm{PA} \vdash \varphi \Rightarrow \mathrm{PA} \vdash \Box\varphi$
2. $\mathrm{PA} \vdash \Box(\varphi \to \psi) \wedge \Box\varphi \to \Box\psi$

Note that the second property is equivalent to the following property:

$$\mathrm{PA} \vdash \Box(\varphi \to \psi) \to (\Box\varphi \to \Box\psi) \tag{120}$$

Proof. In the proof of the first incompleteness theorem, we constructed a sentence G such that $\mathrm{PA} \vdash G \leftrightarrow \neg\Box G$. We will now show that for such G we have $\mathrm{PA} \vdash G \leftrightarrow \neg\Box\bot$.

We know $\mathrm{PA} \vdash \bot \to G$. By the first property, this means we have $\mathrm{PA} \vdash \Box(\bot \to G)$. Then by the equivalent formulation of the second property we have $\mathrm{PA} \vdash \Box\bot \to \Box G$. By construction of G we now have: $\mathrm{PA} \vdash G \to \neg\Box G \to \neg\Box\bot$.

For the other implication, we will first proof that the first and second property together proof the following property:

$$\mathrm{PA} \vdash \Box\varphi \wedge \Box\psi \to \Box(\varphi \wedge \psi) \tag{121}$$

We know that for all φ and ψ we have $\mathrm{PA} \vdash \psi \to (\varphi \to (\varphi \wedge \psi))$. So by the first property this means we have $\mathrm{PA} \vdash \Box(\psi \to (\varphi \to (\varphi \wedge \psi)))$.

Now assume we know $\mathrm{PA} \vdash \Box\varphi \wedge \Box\psi$. Then we may conclude the following, using that the property stated above is always true, and using the second property of the proposition twice:

$$\begin{aligned} \mathrm{PA} \vdash &\Box\psi \wedge \Box\varphi \\ \leftrightarrow &\Box(\psi \to (\varphi \to (\varphi \wedge \psi))) \wedge \Box\psi \wedge \Box\varphi \\ \to &\Box(\varphi \to (\varphi \wedge \psi)) \wedge \Box\varphi \\ \to &\Box(\varphi \wedge \psi) \end{aligned} \tag{122}$$

Now we have $\mathrm{PA} \vdash G \to \neg\Box G$, so by the first property we have $\mathrm{PA} \vdash \Box(G \to \neg\Box G)$, by the equivalent definition of the second property we thus have $\mathrm{PA} \vdash \Box G \to \Box(\neg\Box G)$. On the other hand, by our assumption we have $\mathrm{PA} \vdash \Box G \to \Box\Box G$. These together give us $\mathrm{PA} \vdash (\Box G \to (\Box(\neg\Box G) \wedge \Box\Box G))$ and now we have, using the property we just proved:

$$\begin{aligned} \mathrm{PA} \vdash &(\Box G \to (\Box(\neg\Box G) \wedge \Box\Box G)) \\ \to &(\Box G \to \Box((\neg\Box G) \wedge \Box G)) \\ \to &(\Box G \to \Box\bot) \end{aligned} \tag{123}$$

This means we have $\mathrm{PA} \vdash \neg\Box\bot \to \neg\Box G \to G$.

So indeed we have $\mathrm{PA} \vdash G \leftrightarrow \neg\Box\bot$. As we know that G is independent of PA by the first incompleteness theorem, this equivalence shows that also $\mathrm{Con}_{\mathrm{PA}}$ is independent of PA and therefore the consistency of PA cannot be proved within PA.

□

Now for the proof of the second incompleteness theorem, we assumed a property that was a consequence of the *Formalised* Σ_1*-completeness*:

Theorem 50 (Formalised Σ_1-completeness of PA)**.** For every Σ_1-sentence of PA,

$$\mathrm{PA} \vdash \varphi \to \Box\varphi \tag{124}$$

Now the proof of this theorem is far from trivial, and I will only give an idea of how the proof can be carried out.

First, we take symbols T, representing the primitive recursive operation $n \to \ulcorner\overline{n}\urcorner$ and S_f and S_t, representing primitive recursive substitution operations on formulae and terms respectively, and prove that PA proves the recursion for those functions.

Then one can prove the following lemma:

Lemma 51. For every Δ_0-formula $\varphi(v_0, \ldots, v_{k-1})$ we have:

$$\mathrm{PA} \vdash \forall x_0 \cdots x_{k-1}(\varphi(\vec{x}) \to \exists y \overline{\mathrm{Prf}}(y, S_f(\overline{\ulcorner\varphi\urcorner}, \langle T(x_0), \ldots, T(x_{k-1})\rangle))) \tag{125}$$

To prove this lemma, one has to formalise the results obtained in lemma 37, the Σ_1-completeness of PA.

This lemma now suffices to prove the theorem, it only has to be extended to cover not only Δ_0- but also Σ_1-formulae: If ψ is a Σ_1-formula, we know $\mathrm{PA} \vdash \psi \leftrightarrow \exists x_1, \ldots, x_n \varphi(\vec{x})$ for some Δ_0-formula φ. Now for φ we know:

$$\mathrm{PA} \vdash \forall x_0 \cdots x_{k-1}(\varphi(\vec{x}) \to \exists y \overline{\mathrm{Prf}}(y, S_f(\overline{\ulcorner \varphi \urcorner}, \langle T(x_0), \ldots, T(x_{k-1})\rangle))) \tag{126}$$

So we have

$$\begin{aligned}\mathrm{PA} \vdash & \exists x_0 \cdots x_{k-1} \varphi(\vec{x}) \to \exists y \overline{\mathrm{Prf}}(y, S_f(\overline{\ulcorner \varphi \urcorner}, \langle T(x_0), \ldots, T(x_{k-1})\rangle)) \\ & \to \exists y \overline{\mathrm{Prf}}(y, S_f(\overline{\ulcorner \exists \vec{x} \varphi(\vec{x}) \urcorner}, \langle T(x_0), \ldots, T(x_{k-1})\rangle))\end{aligned} \tag{127}$$

Which is equivalent to

$$\mathrm{PA} \vdash \psi \to \exists y \overline{\mathrm{Prf}}(y, S_f(\overline{\ulcorner \psi \urcorner}, \langle T(x_0), \ldots, T(x_{k-1})\rangle)) \tag{128}$$

8 Gödel and Hilbert's Programme

In 1930, a conference was held, at which Gödels incompleteness theorems were presented. On the first day, von Neumann presented an overview of Hilbert's programme.

Gödel reacted to this in the following way: 'one cannot claim with certainty of any formal system that all conceptual considerations are representable in it.'[44] Thus, it is possible that one could prove, with the transfinite methods of classical mathematics, a sentence p in a consistent formal system $\mathfrak{A}$, though all that follows from the consistency of $\mathfrak{A}$ is that not-p is not provable within $\mathfrak{A}$, and one could still recognise not-p through some conceptual considerations with cannot be formally represented in $\mathfrak{A}$.

Also Gödel stated his first incompleteness theorem, given as a critique of Hilbert's programme and the inadequacy of consistency for Hilbert's purpose:

> One can (under the assumption of the consistency of classical mathematics) even give examples of statements (and even of the sort of Goldbach's or Fermat's), which are conceptually correct but unprovable in the formal system of classical mathematics. Therefore, if one adjoins the negation of such a statement to the axioms of classical mathematics, then one obtains a consistent system in which a conceptually false sentence is provable. [45]

8.1 Gödels Refutation of the Original Programme

It is clear that Gödel's second incompleteness theorem shows that Hilbert's programme, in his original form, can never be attained. Hilbert wanted to formalise mathematics in a consistent system, and as he wanted to do *all* mathematics in this system, also the consistency of that system should be proved within the system, which is impossible, according to the second incompleteness theorem. The first incompleteness theorem showed that the programme developed by Hilbert, if it would prove consistency, could never prove completeness, and therefore could never be completed.

Smorynski presented an argument that even the first incompleteness theorem by itself defeats Hilbert's programme.[46], He argues in the following way. Suppose T is a theory formalising ideal mathematics. If there would be a finitary consistency proof of T, then T must be conservative over S, a subtheory of T formalising finitary real mathematics, for Π_1-formulae. Because if T were not conservative over S, the consistency of T could never be proved in S. We may apply the first incompleteness theorem to S, as finitary real mathematics clearly contains enough arithmetic. So there is a sentence G_S which is not derivable in S if S is consistent. T formalises the sentence 'if S is consistent, then G_S', as T formalises ideal mathematics (which is, roughly, all of mathematics that is not finitary), and finitary mathematics as a subtheory, which means it formalises all of mathematics. But it also proves the soundness of S and therefore its consis-

[44] As cited in [Smorynski, 1986], p.51

[45] As cited in [Smorynski, 1986] p.52

[46] He does so in [Smorynski, 1977], according to [Zach, 2006]. The statement is indeed in [Smorynski, 1977], at p. 825, the argument however I cannot find, so the one explained here is from [Zach, 2006].

tency[47], which means T proves G_S. This means there is a true real statement, provable in ideal mathematics, but undecidable in real mathematics. So ideal mathematics is *not* conservative over finitary real mathematics.

Smorynski also notes why this means that its consistency cannot be proved in finitary real mathematics either[48], by showing that consistency implies conservation in this case, which would lead to a contradiction.

Let φ be a Π_1 formula, say $\forall x \psi x$, with $\psi \in \Delta_0$. If $T \vdash \varphi$, then there is a derivation d of φ from T. But derivations are concrete objects, which means that for some real formula P encoding derivations in T, we have $S \vdash P(d, \ulcorner\varphi\urcorner)$. If φ were false, there would be an a such that $\neg\psi a$, and there would be a c such that $S \vdash P(c, \ulcorner\neg\varphi\urcorner$. There would even be the following, stronger, assertion, with c_x depending on x:

$$S \vdash \neg\psi x \to P(c_x, \ulcorner\neg\varphi\urcorner) \tag{129}$$

Now if S would prove the consistency of T, we would have

$$S \vdash \neg(P(d, \ulcorner\varphi\urcorner) \wedge P(c, \ulcorner\neg\varphi\urcorner)) \tag{130}$$

because S proves there is no contradiction in T,because it proves the consistency of T. Because $S \vdash P(d, \ulcorner\varphi\urcorner)$ this means we have $S \vdash \neg P(c, \ulcorner\neg\varphi\urcorner)$, so by the implication stated above we have $S \vdash \psi x$ for free variable x, so we have $S \vdash \varphi$.[49]

So if it is shown by the first incompleteness theorem that T is not conservative over S, we know by the argument given above that S does not prove the consistency of T.

Gödel's original formulation, as cited above, also shows another important consequence of the theorems. It shows that, if a certain formal system is consistent, it is incomplete, and we can find a sentence which is unprovable in the system, but conceptually correct. The system obtained when adding the negation of that sentence to the original system, would be consistent. But also, it would be proving a conceptually false statement!

This shows that maybe Brouwer was right: consistency is not enough to ensure truth. Indeed, consistency is only a weak soundness condition.

We can indeed say that through finitistic reasoning, indeed we cannot prove the consistency of PA. However, if we widen our scope, we see that the consistency of PA can be proved in ZFC quite easily.[50] An often heard objection against this, is a consistency proof of PA in ZFC is only meaningful if ZFC *itself* is consistent.

This argument of course makes sense, but it is striking to see that those kind of objections are never heard in the field of 'ordinary' mathematics. Returning to a famous theorem we've seen earlier, when Andrew Wiles proved Fermat's theorem, there was no objection like: 'No, he only proved that *if* ZFC is consistent, then there is no solution in positive integers to $x^n + y^n = z^n$ for $n > 2$'. Now one could say that there is always an implicit 'if ZFC is consistent' in front of every mathematical theorem, and only in the case of consistency theorems it is important to explicitly state them. However, there is no way to make sense

[47] Zach does not explain why this is true.

[48] [Smorynski, 1977], p.824

[49] Smorynski notes that this proof is a bit vague and refers to other parts of [Smorynski, 1977] for a detailed proof.

[50] [Franzén, 2005], p. 106

of those theorems, if one seriously doubts the consistency of ZFC.[51] So we may conclude that the consistency of PA can indeed be proved in ZFC, if we view it as just an ordinary proof in mathematics. If one equally doubts the consistency of ZFC, then of course the proof will not be convincing, but this doubt only makes sense given a skeptical attitude towards all other mathematics as well.

So we see that if our scope is widened, it is still possible to give proofs of consistency like the ones Hilbert sought for, and this was also Gödel's own opinion[52]: he actually himself gave one way of extending the notion of finitistic proof, so that acceptable consistency proofs could be carried out. This was done in his *Dialectica Interpretation*, published in 1958. We will return to this and other ways of adjusting the programme in the next section.

[51] From [Franzén, 2005] p.112

[52] From [Franzén, 2005], p.39

9 Adjustment of the Programme to Gödel's Theorems

The question that arises now, is whether or not this refutation of the ambitious programme proposed by Hilbert, is in fact the end of it. Many people have worked on it afterwards, trying to work their way around the incompleteness theorems. It was Hilbert himself who made the first attempts.

It was only half a year later, in the beginning of 1931, that Hilbert was confronted with Gödel's first theorem.

He tried to find his way around the theorem. His solution was to extend the following rule:

$$\begin{array}{l}\text{if, for every numeral } n, \text{ the numeral formula } \varphi(n) \text{ can be checked} \\ \text{ to be a correct one, then conclude } \forall x \varphi x.\end{array} \tag{131}$$

His extension was called the ω-rule, and the simplest form reads as follows:

$$\begin{array}{l}\text{from } \varphi(0), \varphi(1), \dots \\ \text{infer } \forall x \varphi x\end{array} \tag{132}$$

This rule allows the derivation of all true arithmetic sentences, under a variety of restrictions. However this method was not considered finitistic, Hilbert proved the Law of the Excluded Middle in the last of his full publications, by deriving all true arithmetic sentences via the ω-rule.

Gödel did not understand how Hilbert could still write a paper like this, after he had shown his results.[53]. However, he did not commit himself to the question whether or not he had destroyed Hilbert's programme.

For the first volume of *Grundlagen der Mathematik*, published in 1934, Hilbert wrote the preface. In this preface he said the following:

> ...the occasionally held opinion, that from the results of Gödel follows the non-executability of my Proof Theory, is shown to be erroneous. This result shows indeed only that for more advanced consistency proofs one must use the finite standpoint in a deeper way than is necessary for the consideration of elementary formalisms.[54]

This remark, however, is rather odd. Because if we would accept the distinction, made by Hilbert, between actual mathematics and finite metamathematics, then there are only two possible conclusions. By this time, Gödel had already proved his second incompleteness theorem. Either, the same refutation would still hold, according to the second incompleteness theorem, or the formal codification of actual mathematics cannot adequately represent finitary mathematics, meaning that the codification is not strong enough for the second incompleteness theorem to apply to it.

These remarks by Hilbert have later been read in a different way: that there is a hierarchy of formal systems, which is never ending, of actual mathematics. Also, there is a hierarchy of deepenings of metamathematics, corresponding to

[53] He wrote to Olga Taussky-Todd: 'How can he write such a paper after what I have done?', [Smorynski, 1986] p. 53

[54] As cited in [Smorynski, 1986], p.54

the former hierarchy. In such a deepening of metamathematics one could prove the consistency of the corresponding formal system of actual mathematics. This led to a more successful development. However, the question remains why these consistency proofs are given, as Gödel had also shown that consistency did not imply truth.

There have been different ways in which Hilbert's programme has been adapted to evade Gödel's theorems. Some of the more common of these methods will be discussed below.

9.1 The Generalised Hilbert's Programme

As shortly noted when discussing finitism (section 2.2), Hilbert's programme can be seen as an attempt to offer an intuitionistic justification of classical mathematics.

There have been many proof-theoretical approaches, using a generalisation of the finitary standpoint to accomplish an analysis of systems of classical mathematics. Roughly we can divide those in an approach by ordinal analysis and a functional approach.

9.1.1 Ordinal Analysis

Gentzen has given a consistency proof of PA that uses transfinite induction up to ε_0. This was actually one of the first proof-theoretic results after Gödel's theorems had been presented. Gentzen uses a system of notations for ordinals less than ε_0, and he proves that the reduction procedure for derivations in PA terminates. This proof is based on induction of these ordinal notations.

This system and proof have been the basis for a theory of *ordinal analysis.* To give an ordinal analysis of a theory T, one has to produce an ordinal notational system for ordinals less then some ordinal α, such that for every $\beta < \alpha$, the formalisation of the transfinite induction principle for β is provable in T. Then one can prove, in practice, the consistency of T using transfinite induction up to α and finitary methods.

An example: the consistency of PA can be shown by induction up to ε_0, and it can also be shown that for all $\beta < \varepsilon_0$, PA proves the formalisation of the transfinite induction principle for β. So if β is the ordinal number of some recursive well-order $<$, and φ represents this well-order, then we have PA $\vdash \forall x(\forall y(\varphi(y,x) \to \psi(x)) \to \forall x \psi(x)$ for all ψ. This constitutes an ordinal analysis of PA. We say ε_0 is the *proof theoretic ordinal* of PA.

More, stronger results have been attained using this method. However, it is still in question how much of the complex ordinal systems are acceptable as being finitarily. Also it is unclear what the philosophical meaning of these results is. As Feferman puts it[55]:

> As the systems of ordinal notation used for consistency proofs of stronger and stronger theories become more and more complicated, the significance to noncognoscenti of what is thereby accomplished decreases in inverse proportion. Thus, on the one hand, to say that one has obtained a constructive consistency proof of a theory T - without saying anything more - is too general to be informative;

[55] As cited in Zach [2006] p.27

> and, on the other hand, to say that the proof has been carried out by transfinite induction on a certain complicated recursive ordering for some very large ordinals tells us nothing about what constructive principles are involved in the proof of this well-ordering.

9.1.2 Functional Interpretation

The functional interpretation is modelled after Gödels *Dialectica interpretation.* I will present a short overview of how this interpretation works, for a detailed explanation see Feferman and Avigad [1995], from which the overview below has been abstracted.

Gödel's interpretation is based on intuitionistic arithmetic, using the axioms of Heyting Arithmetic, or HA for short. These are the following axioms:

1. From φ, $\varphi \to \psi$ conclude ψ
2. From $\varphi \to \psi, \psi \to \theta$ concude $\varphi \to \theta$
3. $\varphi \vee \varphi \to \varphi$, $\varphi \to \varphi \wedge \varphi$
4. $\varphi \to \varphi \vee \psi, \varphi \wedge \psi \to \varphi$
5. $\varphi \vee \psi \to \psi \vee \varphi, \varphi \wedge \psi \to \psi \wedge \varphi$
6. From $\varphi \to \psi$ conclude $\theta \vee \varphi \to \theta \vee \psi$
7. From $\varphi \to (\psi \to \theta)$ conclude $\varphi \wedge \psi \to \theta$ and conversely
8. $\bot \to \varphi$
9. From $\varphi \to \psi$ conclude $\varphi \to \forall x\psi$, assuming x is not free in φ
10. $\forall x\varphi \to \varphi[t/x]$ assuming t is free for x in φ
11. $\varphi[t/x] \to \exists x\varphi$ assuming t is free for x in φ
12. From $\varphi \to \psi$ conclude $\exists x\varphi \to \psi$, assuming x is not free in ψ

Here $\varphi[t/x]$ denotes the earlier definition of substitution. We define negation by

$$\neg A = A \to \bot \tag{133}$$

here $\bot$ is an identically false statement, which may be identied with $0 = 1$. The equality axioms are given by:

1. $x = x$
2. $x = y \to (\varphi[x/z] \to \varphi[y/z])$, where φ is atomic.

Now to obtain classical logic, we have to add the law of the excluded middle; $\varphi \vee \neg\varphi$. To reduce classical predicate logic to intuitionistic predicate logic, one uses the *double-negation* translation, developed by Gödel and Gentzen. This is defined as follows.

1. $\varphi^N = \neg\neg\varphi$ for φ atomic

2. $(\varphi \wedge \psi)^N = \varphi^N \wedge \psi^N$
3. $(\varphi \vee \psi)^N = \neg(\neg\varphi^N \wedge \neg\psi^N)$
4. $(\varphi \to \psi)^N = \varphi^N \to \psi^N$
5. $(\forall x \varphi(x))^N = \forall x \varphi(x)^N$
6. $(\exists x \varphi(x))^N = \neg\forall x \neg\varphi(x)^N$

From an intuitionistic point of view, we have the equivalences $(\varphi \wedge \psi)^N \leftrightarrow \neg\neg(\varphi \vee \psi)^N$ and $(\exists x \varphi)^N \leftrightarrow \neg\neg\exists x \varphi^N$. From a classical point of view, every formula is equivalent to its N-interpretation.

Now we have that if a set of axioms S proves a formula φ using classical logic, then S^N proves φ^N using intuitionistic logic. This leads us to the following result: if PA proves a formula φ, then HA proves φ^N.

Now the Dialectica interpretation reduces HA to a theory T which axiomatises a class of functionals that Gödel called the *primitive recursive functionals of finite type*. T is a quantifier-free theory. Gödel assigned to every term of T a *type symbol*. The set of type symbols is generated inductively by the following rules:

1. 0 is a type
2. If σ and τ are types then so is $\sigma \to \tau$.

The idea is that objects of type 0 are natural numbers, and objects of type $\sigma \to \tau$ are considered to be functions form objects of type σ to objects of type τ.

Now to each type σ we can assign a natural number lev(σ) as its *type level*:

1. $\text{lev}(0) = 0$
2. $\text{lev}(\sigma \to \tau) = \text{max}(\text{lev}(\sigma) + 1, \text{lev}(\tau))$

Now by this convention, every type is assigned a finite level, so the language of T is said to be *of finite type*.

Gödel continues by defining the set of terms of T and assigning types to these terms, and defines terms for the constant function, successor function etc. Now using these one can formulate the axioms of T, which consist of the usual defining equations for functions like the successor function, a rule allowing substitution in the usual way, equality axioms, the axioms of classical propositional axioms and a scheme of induction.

Now to each formula φ in the language of arithmetic, we associate its *Dialectica* interpretation φ^D, which is a formula of the form

$$\varphi^D = \exists x \forall y \varphi_D \tag{134}$$

where φ_D is a quantifier-free formula in the language of T. The free variables of φ_D consist of those free in φ, together with the sequences of variables (possibly empty) x and y. The Dialectica interpretations are defined inductively as follows, where $\varphi^D = \exists x \forall y \varphi_D$ and $\psi^D = \exists u \forall v \psi_D$.

1. For φ an atomic formula, x and y are both empty and $\varphi^D = \varphi_D = \varphi$

2. $(\varphi \wedge \psi)^D = \exists x, u \forall y, v(\varphi_D \wedge \psi_D)$

3. $(\varphi \vee \psi)^D = \exists z, x, u \forall y, v((z = 0 \wedge \varphi_D) \vee (z = 1 \wedge \psi_D))$

4. $(\forall z \varphi(z))^D = \exists X \forall z, y \varphi_D(X(z), y, z)$

5. $(\exists z \varphi(z))^D = \exists z, x \forall y \varphi_D(x, y, z)$

6. $(\varphi \to \psi)^D = \exists U, Y \forall x, v(\varphi_D(x, Y(x, v)) \to \psi_D(U(x), v))$

And from 6 we obtain

7. $(\neg\varphi)^D = \exists Y \forall x \neg \varphi_D(x, Y(x))$

The capital letters are introduced while 'Skolemising' the existential variables, item 6 will be used to show this method:

$$\begin{aligned}
(\exists x \forall y \varphi_D(x, y) \to \exists u \forall v \psi_D(u, v)) &\leftrightarrow \\
\forall x(\forall y \varphi_D(x, y) \to \exists u \forall v \psi_D(u, v)) &\leftrightarrow \\
\forall x \exists u(\forall y \varphi_D(x, y) \to \forall v \psi_D(u, v)) &\leftrightarrow \\
\forall x \exists u \forall v(\forall y \varphi_D(x, y) \to \psi_D(u, v)) &\leftrightarrow \\
\forall x \exists u \forall v \exists y(\varphi_D(x, y) \to \psi_D(u, v)) &\leftrightarrow \\
\forall x \exists u, Y \forall v(\varphi_D(x, Y(v)) \to \psi_D(u, v)) &\leftrightarrow \\
\exists U, Y \forall x, v(\varphi_D(x, Y(v)) \to \psi_D(U(x), v))&
\end{aligned} \tag{135}$$

Not all of these equivalences are valid in intuitionistic reasoning. However, the Dialectica interpretation can still be used as a tool in constructive metamathematics.

Now Gödel's main result is as follows:

Theorem 52. Suppose φ is a formula in the language of arithmetic, and HA proves φ. Then there is a sequence of terms t such that T proves $\varphi_D(t, y)$.

Combining this with the earlier noted result, we obtain the following:

Suppose φ is a formula in the language of arithmetic, such that PA proves φ. Then there is a sequence of terms t such that T proves $(\varphi^N)_D(t, y)$.

We won't present the proof here. We see that Gödel proved that intuitionistic first-order arithmetic can be reduced to a quantifier-free theory. From this one can derive, more generally, a method to reduce an infinitary theory T to a quantifier-free theory F. This quantifier-free theory F is not strictly finitary, it contains functions of finite type. Usually one first proves the reduction of a classical theory to a variant of intuitionistic arithmetic, and then a reduction of the latter to a quantifier-free functional theory.

The advantage of this approach over the ordinal analysis, is that this method yields an analysis of the computational content of F: the terms of F represent natural classes of functions such as primitive recursive functions. If a function is recursive, and can be proved to be total in T, it is represented by a term of F, thereby belonging to one of those classes.

This also leads to constructive information: the functional interpretation and Gödel's functionals of finite type show much clearer what kind of constructive methods are used.[56]

[56] From Feferman and Avigad [1995] p. 2

9.2 The Relativised Hilbert's Programme

Interpreting Hilbert's programme as an attempt to justify ideal mathematics by restricted methods, one can reduce its goal to justifying only fragments of higher mathematics. In this interpretation, Hilbert's aim was to show that ideal mathematics does not go beyond real mathematics, so that real mathematics is a foundation for ideal mathematics. A finitary consistency proof of ideal mathematics would have accomplished this.

This means that one could give finitary consistency proofs of fragments of higher mathematics, thereby justifying those fragments. Two examples of these kinds of proofs are the work of Feferman and the reverse mathematics project of Friedman and Simpson, both of which we will discuss shortly. What all the relativised Hilbert's programmes have in common, is that they focus on reducing systems of classical mathematics to more restricted systems, in a proof-theoretic, finitistic way.

9.2.1 Feferman's Approach

A reduction in the sense Feferman defined to be a foundational reduction, is a proof that 'a body of mathematics which is justified by a foundational framework $\mathcal{F}_1$ (e.g. finitary, constructive, predicative, infinitary, set-theoretic) can already be justified, in a certain sense, in a weaker, or stricter foundational frame work $\mathcal{F}_2$.'[57]. Generally, this is not possible for a whole body of mathematics, but partial foundational reductions can be achieved. The following definition shows how theories can be reduced, showing conservativity for fragment of mathematics, denoted by φ.

Definition 53. Suppose we are given a theory T_1, justified by a foundational framework $\mathcal{F}_1$, and a theory T_2, justified by a weaker foundational framework $\mathcal{F}_2$. A *proof-theoretic reduction, conservative for* φ[58] of T_1 to T_2, is a partial recursive function f such that:

1. Whenever x is (the code of) a proof in T_1 of a formula (with code) y in φ, then $f(x)$ is (the code of) a proof of y in T_2

2. T_2 proves the formalisation of (1).

If such an f exists, we write $T_1 \leq T_2[\varphi]$

The second requirement for f ensures that the function f itself is justified in $\mathcal{F}_2$.

Definition 54. Suppose we are given a theory T_1, justified by a foundational framework $\mathcal{F}_1$, and a theory T_2, justified by a weaker foundational framework $\mathcal{F}_2$. A *partial foundational reduction* of $\mathcal{F}_1$ to $\mathcal{F}_2$ is a proof-theoretic reduction that establishes $T_1 \leq T_2[\varphi]$.

Now it has been shown, in the partial foundational reductions that have been carried out in practice, that f is a primitive recursive function. Also the formalisation of (1), so the second requirement, can be carried out in primitive

[57] Cited from Zach [2006] p.28

[58] Definition from Zach [2006] p.28

recursive arithmetic PRA[59]. We know PRA is justified by the finitary framework directly, this means that partial foundational reductions are all finitarily justified.

A proof-theoretic reduction of a theory T_1 to a theory T_2 yields a consistency proof of T_1 in T_2: of T_2 is not consistent, then a consistency proof of T_1 can be given in T_2. If T_2 is consistent, then there is no proof of $0 = 1$ in T_2, which means there cannot be a proof of $0 = 1$ in T_1 as T_1 is conservative over T_2, which means that T_1 is consistent.

This means that this method of proof-theoretic reduction provides a solution for Hilbert's programme: it yields relativised consistency proofs for specific theories.

Feferman notes that this shows that adopting a viewpoint such as finitism or constructivism does not necessarily mean one has to give up mathematics as we know it. For example, subsystems of analysis have already been reduced to finitary systems. On the other hand, he adds, one should reflect on the reasons for adopting such a viewpoint seriously, because of the sacrifices that are required nonetheless. We will return to this consideration later.

9.2.2 Reverse Mathematics

The programme of reverse mathematics is another continuation of Hilbert's programme, and has been developed by Friedman and Simpson. Gödel had shown that not all of classical mathematics could be reduced to the finitary. The programme is the search for the answer to the following question: how much of classical mathematics can be reduced to the finitary?

As noted before, there are subsystems of analysis which have been reduced (in the Feferman sense) to finitary systems. Reverse mathematics now investigates which theorems of classical mathematics are provable in those subsystems. A typical example is that the Hahn-Banach theorem is provable in a theory denoted WKL_0, which is proof-theoretically reducible to PRA for φ the set of sentences of the form $\forall x \exists y A(x, y)$.

Reverse mathematics is primarily concerned with subsystems of analysis, which are infinitary systems. Gödel showed, however, that not even all true statements in first-order arithmetic are provable in PA. So not even first-order arithmetic can be given a finitary foundation. The question now rises what sort of statements are not provable in PA: are they mathematically interesting? The most famous result to answer this question with, is a proof given by Parris and Harrington: there is a version of the finite Ramsey theorem[60] which is not provable in PA. Clearly this theorem is mathematically relevant. On the other hand, this example is constructed specifically so that it would be independent of PA, and the most of the 'ordinary' statements in mathematics can be proved even in weaker systems than PA, whose consistency can be proved by finitistic (primitive recursive) means.

[59] PRA is a system of arithmetic, weaker than PA. The consensus is that all reasoning in PRA is finitistic. PRA is often used for consistency proofs, e.g. by Gentzen, whose proofs we have seen in section 9.1.

[60] This is a theorem from combinatorics, about the colouring of edges of graphs.

9.3 The Instrumentalist Hilbert's Programme

Detlefsen has given an instrumentalist interpretation of Hilbert's programme, designed to evade the difficulties posed by the incompleteness theorems.

First, Detlefsen gives an analysis and defense of the use of instrumentalism in mathematics. In this analysis Detlefsen states that even though e.g. full set theory is accepted as a formalisation of infinitary mathematics, only parts of it are useful from an instrumentalist view. There are two sorts of ideal proofs which are not useful: first, the ideal proofs of real theorems that are more complex than any real proof of the same theorem and second, the ideal proofs which are so long or complex that they cannot be comprehended by humans. Also Detlefsen states that the proof of the conservativity of ideal theory over real mathematics is only required for the instrumentally useful part. The part of ideal mathematics that is instrumentally useful is called the *Hilbertian residue*. The point now is, that the Hilbertian residue may not contain enough basic arithmetic in order for the incompleteness theorems to apply to it, in particular the second incompleteness theorem. This means that the consistency of the Hilbertian residue could still be proved.

Another argument Detlefsen[61] uses is that ideal mathematics need not be conservative over the real part. Hilbertian instrumentalism only requires, according to Detlefsen, that the ideal theory does not prove anything which is in conflict with the real theory. If this is true, then, according to Zach [2006], the first incompleteness theorem indeed does not pose a challenge to the instrumentalist interpretation of Hilbert's programme, as this interpretation evades the arguments of Smorynski explained above in section 8.1.[62]

Also Detlefsen presents arguments against the application of the second incompleteness theorem. He gave a version of the ω-rule, and if this version were finitarily acceptable, this would yield a finitarily acceptable method of proof, not capable of being formalised in ideal mathematics. This implies that real mathematics is not a subtheory of ideal mathematics, which means we cannot claim via the second incompleteness theorem that there cannot be a real consistency proof of ideal mathematics (because the fact that real mathematics cannot prove its own consistency is no longer an objection). Detlefsen's version of the ω-rule however, has been criticised seriously.[63] Another argument Detlefsen presents is about the formalisation of consistency. Gödel's second theorem in its generalised form is about the sentence Con_T, stating that T is consistent. However, the fact that Con_T is not provable does not mean there could not be other formalisations of consistency that *are* provable in T. These other formalisations could be found using another definition of the (generalised form of the) predicate $\mathrm{Prf}(x, y)$ that we introduced before. These might, Detlefsen argues, be provable in the corresponding theories. However, these conceptions

[61] The following argument is the main subject of [Detlefsen, 1990], a more detailed explanation can be found there.

[62] I think Zach might have misread Detlefsen [1990], as Detlefsen only argues that ideal mathematics need not be conservative over real mathematics, and indeed the first incompleteness theorem shows that indeed it could not be conservative over real mathematics. Smorynski however, in Smorynski [1977], shows that ideal mathematics cannot be consistent if it is not conservative over real mathematics, which means he is still choosing consistency as a requirement for ideal mathematics, instead of conservativity over real mathematics. so Detlefsen might be right that the first incompleteness theorem does not pose a problem for the instrumentalist interpretation, Smorynski did not state that it would, in Smorynski [1977].

[63] See [Zach, 2006] p. 24

of provability would most likely not have been accepted by Hilbert. An example is the use of Rosser provability (which we have touched upon when discussing Gödel sentences). This states that a derivation of A is only a proof of A if no derivation of $\neg A$ exists with smaller Gödel number than the derivation of A:

$$\mathrm{Prf}_R(\ulcorner A \urcorner) \equiv \exists x(\mathrm{Prf}(x, \ulcorner A \urcorner) \wedge \forall y < x \neg \mathrm{Prf}(y, \ulcorner \neg A \urcorner)) \tag{136}$$

Now we have, for example, that $\neg\mathrm{Prf}_R(\ulcorner 0 = 1 \urcorner)$ is provable in PA! However, the provability of a formula is, in this definition, more than just deriving it from the axioms: one has to check that every derivation with smaller Gödel number does not end in $\neg A$. Also other notions of provability have been studied, for an overview, see [Zach, 2006] p. 25.

10 Conclusion

We have seen an overview of Hilbert's programme, Gödel's theorems and their proofs, and the implications of the incompleteness theorems for the programme. Then we have looked at several ways at which the programme has been extended or adapted to evoid the consequences of the incompleteness theorems.

We have seen that, indeed, Hilbert's original formulation of the programme has been refuted. However there are several ways to re-interpret Hilbert's programme, so that it is still possible to accomplish its goals. As the field of logic and foundations has changed very much since Hilbert started working on his programme, it is not always clear what exactly Hilbert means with his definitions and goals. So, some of the re-interpretations are meant to explain in contemporary language what Hilbert could have meant, others really differ from the original programme in their philosophical meaning.

Also, some of the methods used in the extended programmes might not have been accepted by Hilbert, though others might have, and all of them are in line with his original ideas of which methods would be allowed.

One might say that the extended programmes have goals that have been lowered, compared to Hilbert's original programme, which means their outcomes may not be of much value. Because, yes, indeed, when you lower your goal, it is easier to reach it. You could keep lowering it, until you have made sure it is possible to reach your goal.

However, the adaptations made to Hilbert's programme might be of a different kind: substantial results have been achieved and the original formulation has mostly just been adapted to fit in the present conception of the foundations of mathematics, even though this meant it had to be weakened. So we may say that even though Hilbert's original formulation of the programme indeed has been defeated by Gödel's theorems, this has not been the end of it: many important results have been achieved, showing that the programme could be continued in other forms.

Hilbert thought consitency implied truth and therefore existence. This was the main reason for him to focus his programme on consistency proofs. The question remains, now, what reasons we have to proceed giving consistency proofs and reducing systems to weaker systems, now that we know that consistency does not neccesarily imply existence, or truth, as shown by Gödel's theorems.

References

L. E. J. Brouwer. *Over de Grondslagen der Wiskunde.* Maas & van Suchtelen, Amsterdam - Leipzig, 1907.

A. Cantini. Paradoxes and contemporary logic. In E. N. Zalta, editor, *The Stanford Encyclopedia of Philosophy.* Winter 2012 edition, 2012.

D. Dalen. *L.E.J. Brouwer: Topologist, Intuitionist, Philosopher.* SpringerLink : Bücher. Springer London, 2013. ISBN 9781447146162. URL `http://books.google.nl/books?id=gFbqtEMohBYC`.

R. Dedekind. The nature and meaning of numbers. In *Essays on the Theory of Numbers.* Open Court Pub. Co., 1901. Authorised translation by Wooster Woodruff Beman.

M. Detlefsen. On an alleged refutation of hilbert's program using gödel's first incompleteness theorem. *Journal of Philosophical Logic*, 19:343–377, 1990.

S. Feferman and J. Avigad. Gödel's functional ("dialectica") interpretation. In S. Buss, editor, *Handbook of Proof Theory*, chapter 6. Elsevier Science B.V., 1995.

T. Franzén. *Gödel's Theorem: an Incomplete Guide to its Use and Abuse.* A K Peters, Ltd., 2005.

D. Hilbert. *Grundlagen der Geometrie.* Teubner, Leipzig, 1899.

D. Hilbert. Axiomatisches denken. *Mathematische Annalen*, 78(1):405 – 415, December 1917.

T. Hofweber. Logic and ontology. In E. N. Zalta, editor, *The Stanford Encyclopedia of Philosophy.* Spring 2013 edition, 2013.

K. Ireland and M. Rosen. *A Classical Introduction to Modern Number Theory.* Springer-Verlag, New York, 2nd ed. edition, 1990.

J. J. C. Kuiper. *Ideas and Explorations : Brouwers Road to Intuitionism.* PhD thesis, Universiteit Utrecht, 2004.

U. Majer and M. Hallett, editors. *David Hilbert's Lectures on the Foundations of Geometry, 1891-1902.* Springer, New York, 2004.

H. Mendell. Aristotle and mathematics. In E. N. Zalta, editor, *The Stanford Encyclopedia of Philosophy.* Winter 2008 edition, 2008.

I. Moerdijk and J. van Oosten. Sets, models and proofs, 2000. Revised in 2012.

G. Moser and R. Zach. The epsilon calculus and herbrand complexity. *Studia Logia*, 99:1–23, 2005.

J. J. O'Connor and E. F. Robertson. Georg Ferdinand Ludwig Philipp Cantor. In *The MacTutor History of Mathematics archive.* 1988.

C. Smorynski. The incompleteness theorems. In J. Barwise, editor, *Handbook of Mathematical Logic*, pages 821–865. North-Holland Publishing Company, 1977.

C. Smorynski. Hilbert's programme. In *Logic's Lost Genius: the Life of Gerhard Gentzen*, chapter 4. Addison-Wesley, 1986.

A. Troelstra and D. van Dalen. *Constructivism in Mathematics, an Introduction*, volume 1. Elsevier, Amsterdam, 1988.

J. van Oosten. Gödels incompleteness theorems, 2009.

R. Zach. Hilbert's program then and now. In *Philosophy of Logic*, pages 411 – 447. Elsevier, 2006.

E. N. Zalta. Gottlob frege. In E. N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Spring 2013 edition, 2013.