The purpose of this note is to give a streamlined proof of Gö del's first incompleteness theorem, perhaps more compact than in the textbook. Recall that Peano Arithmetic PA has an *intended model*, which is the structure $\mathfrak{N}$ of all natural numbers with the usual ordering, addition, and multiplication. PA has many other models as well.

I first need to introduce several basic notions regarding the syntax of Peano Arithmetic; they will be useful also outside of the specific context of Gödel's first incompleteness theorem.

Definition 0.1. Given a natural number n, the symbol $\underline{n}$ denotes the term $SS \dots S0$ of the language of PA, where the successor function is applied n times. The terms of this form are called *numerals*.

Make sure that the distinction between natural numbers and numerals denoting them is clear. A number is a number; a numeral is a term in the language of Peano Arithmetic. The distinction may seem miniscule, but in the confusing context of the incompleteness theorem such distinctions make a difference.

Definition 0.2. A *bounded quantifier* in a formula of the language of PA is one of the form $\forall x < y\ \phi$ or $\exists x < y\ \phi$, where y is a term not mentioning x.

Note that if one runs some sort of brute force verification algorithm for sentences, quantifiers present in principle an unsurmoutable obstacle as there are infinitely many numbers to try and plug in into the quantified variable. The bounded quantifiers, on the other hand, do not present such an obstacle.

Definition 0.3. A Δ_0 formula is one in which all quantifiers are bounded. A Σ_1 formula is one of the form $\exists x \phi$ where ϕ is Δ_0. A Π_1 formula is one of the form $\forall x \phi$ where ϕ is Δ_0.

This is just the beginning of a complexity hierarchy in which one counts the number of alternating unbounded quantifiers. Given a formula, we often want to find the lowest stage of the hierarchy which contains a formula provably equivalent to the given one. This often has major practical consequences. In the context of the incompleteness theorem, the following information about the bottom stage of the hierarchy will be useful:

Theorem 0.4. *Let ϕ be a Δ_0 sentence. Then PA proves ϕ if and only if $\mathfrak{N} \models \phi$.*

In particular, for every Δ_0 sentence ϕ, PA proves either it or its negation since one of the two is satisfied in $\mathfrak{N}$.

Proof. The prof proceeds by lengthy and uneventful induction on the complexity of ϕ. I will only indicate two essential claims proved by induction on m:

Claim 0.5. *For all numbers n, m, k, if $n + m = k$ then PA proves $\underline{n} + \underline{m} = \underline{k}$.*

Claim 0.6. *For every number m, PA proves $\forall x < \underline{m}\ x = 0 \lor x = S0 \lor \dots \lor x = \underline{m-1}$.*

□

The incompleteness theorem is a consequence of a key idea: it is possible to arithmetize the syntax of PA inside PA. The arithmetization of syntax is a map $\phi \mapsto \ulcorner \phi \urcorner$ which assigns to every formula ϕ of the language of PA a numeral $\ulcorner \phi \urcorner$ (often called the Gödel number or code of ϕ) such that

1. we can talk about provability. This means that there is a (Σ_1) formula $\text{Provable}(x)$ such that for every sentence ψ, $PA \vdash \psi$ just in case $\mathfrak{N} \models \text{Provable}(\ulcorner \psi \urcorner)$;
2. we can talk about substitution. This means that there is a Δ_0 formula $\text{Subst}(x, y, z)$ such that for every formula ϕ of one free variable and every number n, $\mathfrak{N} \models \ulcorner \phi(\underline{n}) \urcorner$ is the unique y such that $\text{Subst}(\ulcorner \phi \urcorner, \underline{n}, y)$ holds.

One rather inscrutable feature of arithmetization of syntax is that for any formula $\phi(x)$ of one free variable x, one can form a sentence by plugging in the code for ϕ into the variable of ϕ: $\phi(\ulcorner \phi \urcorner)$. It is perhaps not clear why one would want to do that, but the result is the following key lemma:

Lemma 0.7. (Diagonalization lemma) *For every formula ϕ of one free variable there is a sentence ψ such that Peano Arithmetic proves $\psi \leftrightarrow \phi(\ulcorner \psi \urcorner)$.*

Proof. This uses only the substitution demand on the arithmetization of syntax. Let $\theta(x)$ be the formula $\exists y\ \phi(y) \wedge \text{Subst}(x, x, y) \wedge \forall z < y\ \neg\text{Subst}(x, x, z)$. Let $\psi = \theta(\ulcorner \theta \urcorner)$. I claim that ψ works. For this, note first that PA proves that $\ulcorner \psi \urcorner$ is the smallest y such that $\text{Subst}(\ulcorner \theta \urcorner, \ulcorner \theta \urcorner, y)$ since this sentence is Δ_0 and satisfied in $\mathfrak{N}$.

To see the left-to-right implication, in the theory PA, ψ I must prove $\phi(\ulcorner \psi \urcorner)$. To do this, by the preceding paragraph, the first existential quantifier of ψ can be witnessed only by $\ulcorner \psi \urcorner$ and so $\phi(\ulcorner \psi \urcorner)$ must hold.

To see the right-to-left implication, in the theory $PA, \phi(\ulcorner \psi \urcorner)$ I must prove ψ. To see this, look at the first existential quantifier in ψ and note that it is witnessed by $\ulcorner \psi \urcorner$ by the assumption and the first paragraph of this proof. □

Corollary 0.8. (Gödel's first incompleteness theorem) *There is a sentence ψ in the language of PA such that PA does not prove ψ and PA does not prove $\neg\psi$ either.*

Proof. Let ψ be a sentence such that PA proves $\psi \leftrightarrow \neg\text{Provable}(\ulcorner \psi \urcorner)$. I claim that ψ works.

To see this, assume towards contradiction that PA proves ψ. Then it also proves $\neg\text{Provable}(\ulcorner \psi \urcorner)$ and so $\mathfrak{N} \models \neg\text{Provable}(\ulcorner \psi \urcorner)$. By the first item in the arithmetization of syntax description, it should be the case that PA does not prove ψ, which is a contradiction.

Now assume towards contradiction that PA proves $\neg\psi$. In such a case, it also proves $\text{Provable}(\ulcorner \psi \urcorner)$ and so $\mathfrak{N} \models \text{Provable}(\ulcorner \psi \urcorner)$. By the first item in the arithmetization of syntax description, it should be the case that PA proves ψ, which is a contradiction in PA. □

The sentence ψ produced in the previous proof is provably equivalent to a Π_1 sentence $\neg\text{Provable}(\ulcorner\psi\urcorner)$. This is the minimal complexity of an undecidable sentence. The sentence ψ is satisfied in the model $\mathfrak{N}$.

Corollary 0.9. (Tarski's undefinability of truth) *There is no formula $\phi(x)$ in the language of PA such that for every sentence ψ, $\mathfrak{N} \models \psi \leftrightarrow \phi(\ulcorner\psi\urcorner)$.*

Proof. Given a putative candidate for such a formula ϕ, use the diagonalization lemma to find a sentence ψ such that PA proves $\psi \leftrightarrow \phi(\ulcorner\psi\urcorner)$ and ask whether $\mathfrak{N} \models \psi$ or not. There will be a contradiction in both cases. □