Gödel's Incompleteness Theorems

Scott Sanderson (@scottbsanderson)

March 8, 2017

Papers We Love - Boston

The Stage

David Hilbert

In 1900, published a list of the 23 most important unsolved problems in mathematics for the 20th century.

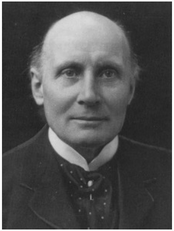
David Hilbert

In 1900, published a list of the 23 most important unsolved problems in mathematics for the 20th century.

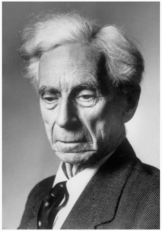
Problem 2 is "prove the consistency of aritmetic".

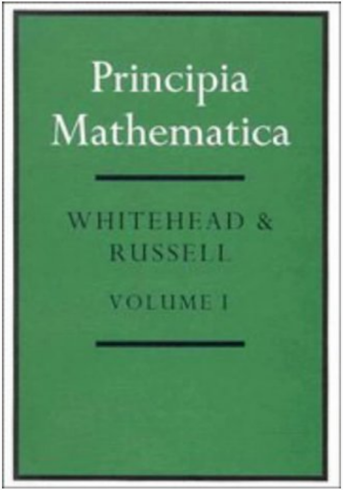
Principia Mathematica and Russell's Logicism

Alfred North Whitehead



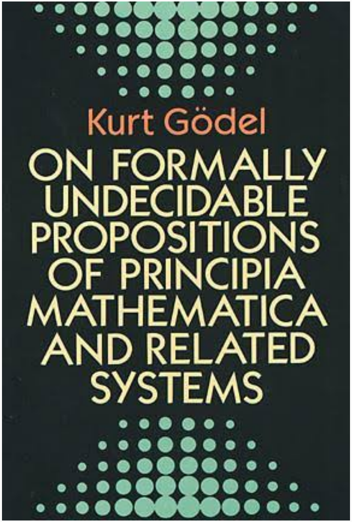
Bertrand Russell





The Paper

The Paper


Kurt Gödel
ON FORMALLY UNDECIDABLE PROPOSITIONS OF PRINCIPIA MATHEMATICA AND RELATED SYSTEMS

Informal Statement of the Incompleteness Theorems

Theorem (Gödel's First Incompleteness Theorem)
Let F be a consistent formal system that can express arithmetic. There are statements well-formed statements in the language of F which are neither provable nor disprovable in F.

Informal Statement of the Incompleteness Theorems

Theorem (Gödel's First Incompleteness Theorem)
Let F be a consistent formal system that can express arithmetic. There are statements well-formed statements in the language of F which are neither provable nor disprovable in F.

Theorem (Gödel's Informal Statement of the Second Incompleteness Theorem)
Let F be a consistent formal system that can express arithmetic. F cannot prove its own consistency.

lolwut?

What do we mean by “formal system”?

lolwut?

What do we mean by “formal system”?

What does it mean for F to be “consistent”?

lolwut?

What do we mean by “formal system”?

What does it mean for F to be “consistent”?

How can a statement be neither provable nor disprovable?

lolwut?

What do we mean by “formal system”?

What does it mean for F to be “consistent”?

How can a statement be neither provable nor disprovable?

What would it mean for F to prove its own consistency?

What is a Formal System?

A **formal system** F is:

- An **alphabet** of symbols (e.g. $0, ', \neg, \exists, \forall, \to$).
- A **grammar**, describing the strings of symbols that are **well-formed** sentences.
- A set of **axioms** containing initial theorems of F.
- A set of **rules of inference** describing how to transform existing theorems into new theorems.

Robinson Arithmetic

Robinson Arithmetic (**Q**) is a simple axiomatization of arithmetic on the natural numbers.

Robinson Arithmetic

Robinson Arithmetic (**Q**) is a simple axiomatization of arithmetic on the natural numbers.

Defines $+$ and $\times$ in terms of **successor** $'$.

Robinson Arithmetic

Robinson Arithmetic (**Q**) is a simple axiomatization of arithmetic on the natural numbers.

Defines $+$ and $\times$ in terms of **successor** $'$.

0 is the only constant. Other numeric constants written as $0'''...'$.

Symbols

$0, ', \neg, =, +, \times, \wedge, \vee, \forall, \exists, \rightarrow, (,)$

$a, b, c, \ldots, x, y, z, \ldots$ (as many as needed)

Well-Formed Sentences

$$0 + 0 = 0$$

$$0' + 0' = 0''$$

$$\forall x(x + 0 = x)$$

$$\forall x \exists y(y + y = x)$$

Invalid Sentences

$$'0$$

$$= +xy$$

$$\forall\exists$$

Robinson Arithmetic

Axioms

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

Robinson Arithmetic

Axioms

1. $\forall x(\neg(0 = x'))$
2. $\forall x \forall y(x' = y' \to x = y)$

Robinson Arithmetic

Axioms

1. $\forall x(\neg(0 = x'))$
2. $\forall x \forall y(x' = y' \to x = y)$
3. $\forall x(\neg(x = 0) \to \exists y(x = y'))$

Robinson Arithmetic

Axioms

1. $\forall x(\neg(0 = x'))$
2. $\forall x \forall y(x' = y' \to x = y)$
3. $\forall x(\neg(x = 0) \to \exists y(x = y'))$
4. $\forall x(x + 0 = 0)$

Robinson Arithmetic

Axioms

1. $\forall x(\neg(0 = x'))$
2. $\forall x \forall y(x' = y' \rightarrow x = y)$
3. $\forall x(\neg(x = 0) \rightarrow \exists y(x = y'))$
4. $\forall x(x + 0 = 0)$
5. $\forall x((x + y') = (x + y)')$

Robinson Arithmetic

Axioms

1. $\forall x(\neg(0 = x'))$
2. $\forall x \forall y(x' = y' \rightarrow x = y)$
3. $\forall x(\neg(x = 0) \rightarrow \exists y(x = y'))$
4. $\forall x(x + 0 = 0)$
5. $\forall x((x + y') = (x + y)')$
6. $\forall x(x \times 0 = 0)$

Axioms

1. $\forall x(\neg(0 = x'))$
2. $\forall x \forall y(x' = y' \to x = y)$
3. $\forall x(\neg(x = 0) \to \exists y(x = y'))$
4. $\forall x(x + 0 = 0)$
5. $\forall x((x + y') = (x + y)')$
6. $\forall x(x \times 0 = 0)$
7. $\forall x \forall y(x \times y' = (x \times y) + x)$

Rules of Inference

$$p, (p \to q) \vdash q$$

$$\neg\neg p \to p$$

$$p \land q \vdash p$$

$$(p \to q) \vdash (\neg q \to \neg p)$$

$$\ldots$$

Consistency

A formal system F is **inconsistent** if, for some proposition p:

$$F \vdash p$$

and

$$F \vdash \neg p$$

Consistency

A formal system F is **inconsistent** if, for some proposition p:

$$F \vdash p$$

and

$$F \vdash \neg p$$

A formal system is **consistent** if it is not inconsistent.

The First Theorem

Theorem (First Incompleteness Theorem)
Let F be a consistent formal system that contains $\mathbf{Q}$. Then we can construct a sentence G_F such that:

1. $F \nvdash G_F$.
2. $F \nvdash \neg G_F$.

Liar Paradox

Consider the sentence **“This sentence is false.”**.

Liar Paradox

Consider the sentence **"This sentence is false."**.

If it's true, then by its own assertion it must be false. ☹

Liar Paradox

Consider the sentence **"This sentence is false."**.

If it's true, then by its own assertion it must be false. ☹

If it's false, then, by its own assertion, it must be true. ☹

Liar Paradox

Consider the sentence **“This sentence is false.”**.

If it’s true, then by its own assertion it must be false. ☹

If it’s false, then, by its own assertion, it must be true. ☹

Many answers have been proposed for how to understand the logical content of this sentence.

Proof Paradox

Consider the sentence **“This sentence is not provable”**.

Proof Paradox

Consider the sentence **“This sentence is not provable”**.

If it’s provable, then we have a proof that the sentence isn’t provable. ☹

Proof Paradox

Consider the sentence **“This sentence is not provable”**.

If it’s provable, then we have a proof that the sentence isn’t provable. ☹

If it’s not provable, then we don’t have such a proof. ☺

Proof Paradox

Consider the sentence **“This sentence is not provable”**.

> If it’s provable, then we have a proof that the sentence isn’t provable. ☹
>
> If it’s not provable, then we don’t have such a proof. ☺

If our deductive system is consistent, then the sentence is true, which means we can’t prove that it’s true

Proof Idea for First Incompleteness Theorem

Proof Idea

Construct a sentence in $\mathbf{Q}$ that asserts its own unprovability.

Definition (Weak Representability)
A set S of natural numbers is **weakly representable** in F if there exists a formula $A(x)$ in the language of F such that:

$$n \in S \iff F \vdash A(\underline{n})$$

Definition (Strong Representability)
A set S of natural numbers is **strongly representable** in F if there exists a formula $A(x)$ in the language of F such that:

$$n \in S \implies F \vdash A(\underline{n})$$

$$n \notin S \implies F \vdash \neg A(\underline{n})$$

Representability (in plain English)

Weak representability means that we can write down a formula A such that F proves $A(\underline{x})$ whenever $x \in S$.

Representability (in plain English)

Weak representability means that we can write down a formula A such that F proves $A(\underline{x})$ whenever $x \in S$.

Strong representability means that can write down a formula A such that F proves $A(\underline{x})$ whenever $x \in S$ **and** F proves $\neg A(\underline{x})$ whenever $x \notin S$.

Representability (in plain English)

Weak representability means that we can write down a formula A such that F proves $A(\underline{x})$ whenever $x \in S$.

Strong representability means that can write down a formula A such that F proves $A(\underline{x})$ whenever $x \in S$ **and** F proves $\neg A(\underline{x})$ whenever $x \notin S$.

We can define representability for relations analogously using formulae with more than one free variable.

Examples of Representability

$$
\begin{aligned}
Even(x) &\equiv \exists y(x = y \times 0'') \\
Odd(x) &\equiv \exists y(x = y' \wedge Even(y)) \\
LessThan(x, y) &\equiv \exists n(x + n' = y) \\
Divides(x, y) &\equiv \exists n(n \times y = x) \\
Prime(x) &\equiv \forall y((Divides(x, y) \rightarrow (y = 0' \vee y = x)))
\end{aligned}
$$

Gödel Numbering

Problem

How to make assertions about provability of **Q**-sentences in **Q**?

Gödel Numbering

Problem

How to make assertions about provability of **Q**-sentences in **Q**?

Idea

Encode assertions about proofs as assertions about numbers.

Gödel Numbering

Problem

How to make assertions about provability of **Q**-sentences in **Q**?

Idea

Encode assertions about proofs as assertions about numbers.

Requires that we can produce a correspondence between numbers and sentences.

Gödel Numbering

Assign a positive integer to each symbol in the alphabet of the language of F.

$C(0) = 1$	$C(=) = 5$	$C(\neg) = 9$	$C(a) = 13$
$C(') = 2$	$C(() = 6$	$C(\forall) = 10$	$C(b) = 14$
$C(+) = 3$	$C()) = 7$	$C(\wedge) = 11$	$\ldots$
$C(\times) = 4$	$C(\rightarrow) = 8$	$C(\vee) = 12$	$C(x_i) = 12 + i$

Gödel Numbering

Encode sequences of symbols in the exponents of the powers of prime numbers.

$$\begin{aligned} \ulcorner 0 + 0 = 0 \urcorner &= 2^{C(0)} 3^{C(+)} 5^{C0} 7^{C(=)} 11^{C(0)} \\ &= 2^1 3^3 5^1 7^5 11^1 \\ &= 4537890 \end{aligned}$$

Gödel Numbering

Encode sequences of symbols in the exponents of the powers of prime numbers.

$$\begin{aligned} \ulcorner 0 + 0 = 0 \urcorner &= 2^{C(0)} 3^{C(+)} 5^{C0} 7^{C(=)} 11^{C(0)} \\ &= 2^1 3^3 5^1 7^5 11^1 \\ &= 4537890 \end{aligned}$$

We can use similar techniques to encode sequences of sequences.

Encoding Proof

Theorem (Valid Proofs are Strongly Representable in Q)
There is a formula of F, $Proof_F(x, y)$ that strongly represents the relation:

x is the Gödel number of a proof of a formula with Gödel number y

Encoding Provability

Corollary (Provability is Weakly Representable in Q)
There is a formula of F, $\mathit{Provable}_F(x)$, that weakly represents the relation:

x is the Gödel number of a provable sentence.

Encoding Provability

Corollary (Provability is Weakly Representable in Q)
There is a formula of F, $Provable_F(x)$, that weakly represents the relation:

x is the Gödel number of a provable sentence.

$$Provable_F(x) \equiv \exists y(Proof_F(x, y))$$

Encoding Self-Reference

Theorem (Diagonalization Lemma)
Let $A(x)$ be a formula in the language of F with one free variable. We can construct a sentence D_A such that:

$$F \vdash (D_A \leftrightarrow A(\ulcorner D_A \urcorner))$$

We refer to D_A as the **diagonalization** of A.

Proof of the First Incompleteness Theorem

Let G_F be the diagonalization of $\neg \mathit{Provable}_F(x)$.

Proof of the First Incompleteness Theorem

Let G_F be the diagonalization of $\neg Provable_F(x)$.

By the Diagonalization Lemma:

$$F \vdash (G_F \leftrightarrow \neg Provable_F(\ulcorner G_F \urcorner)) \tag{1}$$

Proof of the First Incompleteness Theorem

By the Diagonalization Lemma:

$$F \vdash (G_F \leftrightarrow \neg Provable_F(\ulcorner G_F \urcorner)) \tag{1}$$

Suppose $F \vdash G_F$:

By weak representability of $Provable_F$:

$$F \vdash Provable_F(\ulcorner G_F \urcorner)$$

Proof of the First Incompleteness Theorem

By the Diagonalization Lemma:

$$F \vdash (G_F \leftrightarrow \neg Provable_F(\ulcorner G_F \urcorner)) \tag{1}$$

Suppose $F \vdash G_F$**:**

By weak representability of $Provable_F$:

$$F \vdash Provable_F(\ulcorner G_F \urcorner)$$

but by (1),

$$F \vdash \neg Provable_F(\ulcorner G_F \urcorner)$$

Proof of the First Incompleteness Theorem

By the Diagonalization Lemma:

$$F \vdash (G_F \leftrightarrow \neg Provable_F(\ulcorner G_F \urcorner)) \tag{1}$$

Suppose $F \vdash G_F$:

By weak representability of $Provable_F$:

$$F \vdash Provable_F(\ulcorner G_F \urcorner)$$

but by (1),

$$F \vdash \neg Provable_F(\ulcorner G_F \urcorner)$$

So, if F is consistent, $F \nvdash G_F$.

Proof of the First Incompleteness Theorem

Suppose $F \vdash \neg G_F$:

No n is the Gödel number of a proof of G_F, so by strong representability of $Proof_F$, for every numeral $\underline{n}$:

$$F \vdash \neg Proof_F(\underline{n}, \ulcorner G_F \urcorner)$$

Proof of the First Incompleteness Theorem

Suppose $F \vdash \neg G_F$**:**

No n is the Gödel number of a proof of G_F, so by strong representability of $Proof_F$, for every numeral $\underline{n}$:

$$F \vdash \neg Proof_F(\underline{n}, \ulcorner G_F \urcorner)$$

We'd **like** to conclude that:

$$F \vdash \neg \exists n (Proof(n, \ulcorner G_F \urcorner))$$
$$\therefore F \vdash \neg Provable(\ulcorner G_F \urcorner)$$

Proof of the First Incompleteness Theorem

Suppose $F \vdash \neg G_F$**:**

No n is the Gödel number of a proof of G_F, so by strong representability of $Proof_F$, for every numeral $\underline{n}$:

$$F \vdash \neg Proof_F(\underline{n}, \ulcorner G_F \urcorner)$$

We'd **like** to conclude that:

$$F \vdash \neg \exists n (Proof(n, \ulcorner G_F \urcorner))$$

$$\therefore F \vdash \neg Provable(\ulcorner G_F \urcorner)$$

Proof of the First Incompleteness Theorem

Suppose $F \vdash \neg G_F$:

No n is the Gödel number of a proof of G_F, so by strong representability of $Proof_F$, for every numeral $\underline{n}$:

$$F \vdash \neg Proof_F(\underline{n}, \ulcorner G_F \urcorner)$$

We'd **like** to conclude that:

$$F \vdash \neg \exists n (Proof(n, \ulcorner G_F \urcorner))$$

$$\therefore F \vdash \neg Provable(\ulcorner G_F \urcorner)$$

but this actually requires a slightly stronger assumption ω-consistency.

Consistency vs. ω-Consistency

A formal theory is consistent if it is not the case that for some A:

$$F \vdash A$$

$$F \vdash \neg A$$

Consistency vs. ω-Consistency

A formal theory is consistent if it is not the case that for some A:

$$F \vdash A$$

$$F \vdash \neg A$$

A formal theory is ω-**consistent** if is is not the case that for some $A(x)$:

$F \vdash A(\underline{n})$ for every numeral $\underline{n}$.

$$F \vdash \exists n(\neg A(n))$$

Consistency vs. ω-Consistency

A formal theory is consistent if it is not the case that for some A:

$$F \vdash A$$

$$F \vdash \neg A$$

A formal theory is ω-**consistent** if is is not the case that for some $A(x)$:

$F \vdash A(\underline{n})$ for every numeral $\underline{n}$.

$$F \vdash \exists n(\neg A(n))$$

ω-consistency is a stronger condition, and is what Gödel used in his original proof.

Final Statement of the First Incompleteness Theorem

Theorem (Gödel's First Incompleteness Theorem)
Let F be a formal system that contains $\mathbf{Q}$*. Then we can construct a sentence G_F such that:*

1. *If F is consistent, $F \nvdash G_F$.*
2. *If F is ω-consistent, $F \nvdash \neg G_F$.*

Review of First Incompleteness Theorem

Summary:

- We can encode statements **about statements** in arithmetic by **encoding statements as numbers**.

Review of First Incompleteness Theorem

Summary:

- We can encode statements **about statements** in arithmetic by **encoding statements as numbers**.
- We can express "x is a proof of y" (suitably encoded) in arithmetic.

Review of First Incompleteness Theorem

Summary:

- We can encode statements **about statements** in arithmetic by **encoding statements as numbers**.
- We can express "x is a proof of y" (suitably encoded) in arithmetic.
- We can express "x is provable" as "$\exists y$ s.t. y is a proof of x".

Review of First Incompleteness Theorem

Summary:

- We can encode statements **about statements** in arithmetic by **encoding statements as numbers**.
- We can express "x is a proof of y" (suitably encoded) in arithmetic.
- We can express "x is provable" as "$\exists y$ s.t. y is a proof of x".
- Diagonalization transforms a predicate into a statement that is provable in iff its Gödel numbers satisfies the predicate.

Review of First Incompleteness Theorem

Summary:

- We can encode statements **about statements** in arithmetic by **encoding statements as numbers**.
- We can express "x is a proof of y" (suitably encoded) in arithmetic.
- We can express "x is provable" as "$\exists y$ s.t. y is a proof of x".
- Diagonalization transforms a predicate into a statement that is provable in iff its Gödel numbers satisfies the predicate.
- Applying diagonalization to "x is not provable" yields a statement that's provable only if it's not provable.

The Second Theorem

Theorem (Gödel's Second Incompleteness Theorem)
Let F be a consistent formal system that contains Peano Arithmetic, and define:

$$Cons_F \equiv \neg Provable_F(0 = 0')$$

then

$$F \nvdash Cons_F$$

Theorem (Gödel's Second Incompleteness Theorem)
Let F be a consistent formal system that contains Peano Arithmetic, and define:

$$\mathit{Cons}_F \equiv \neg \mathit{Provable}_F(0 = 0')$$

then

$$F \nvdash \mathit{Cons}_F$$

In other words, if F is consistent, then F **can't prove that it doesn't prove contradictions**.

Proof Sketch

Main Idea:

Let G_F be the Gödel sentence constructed in the proof of the first theorem. Prove

$$F \vdash (Cons_F \leftrightarrow G_F)$$

By the first theorem, if F is consistent, then $F \nvdash G_F$.

Therefore, $F \nvdash Cons_F$.

Proof Sketch

How do we show that F proves the equivalence of $Cons_F$ and G_F?

Proof Sketch

How do we show that F proves the equivalence of $Cons_F$ and G_F?

Essentially, the idea is to translate the argument from theorem 1 **into the language of** F.

Proof Sketch

How do we show that F proves the equivalence of $Cons_F$ and G_F?

Essentially, the idea is to translate the argument from theorem 1 **into the language of** F.

We show that $F \vdash (\neg G_F \to \bot \to \neg Cons_F)$, which implies that $F \vdash Cons_F \to G_F$.

Challenges

Derivation requires **Löb's Derivability Conditions** (or equivalent) to be shown for the provability predicate $Provable_F$ (abbreviated here as $Prov_F$):

$$F \vdash A \implies Prov_F(\ulcorner A \urcorner) \tag{1}$$

$$F \vdash Prov_F(\ulcorner A \urcorner) \to Prov_F(\ulcorner Prov_F(\ulcorner A \urcorner) \urcorner) \tag{2}$$

$$F \vdash Prov_F(\ulcorner A \urcorner) \land Prov_F(\ulcorner A \to B \urcorner) \to Prov_F(\ulcorner B \urcorner) \tag{3}$$

Epilogue

Implications for Hilbert's Program

David Hilbert

Second Theorem rules out proving consistency of infinitistic mathematics using finitistic means.

Implications for Hilbert's Program

David Hilbert

Second Theorem rules out proving consistency of infinitistic mathematics using finitistic means.

Still possible to prove arithmetic consistent via external means.

Implications for Hilbert's Program

David Hilbert

Second Theorem rules out proving consistency of infinitistic mathematics using finitistic means.

Still possible to prove arithmetic consistent via external means.

Proven consistent by Gentzen in 1936 using transfinite induction.

Implications for Hilbert's Program

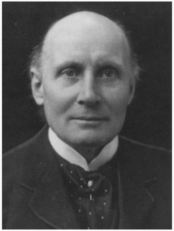
David Hilbert

It is likely that all mathematicians ultimately would have accepted Hilbert's approach had he been able to carry it out successfully. The first steps were inspiring and promising. But then Gödel dealt it a terrific blow (1931), from which it has not yet recovered.

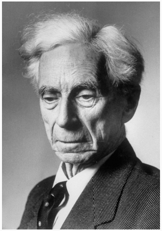
— Hermann Weyl (1946)

Implications for Logicism

Alfred North Whitehead



Bertrand Russell



- Logicism went into decline after Gödel's proofs were announced.

Thanks!

Questions?