On the Second Incompleteness Theorem

Pavel Pudlák

Mathematical Institute, Czech Academy of Sciences, Prague [1]

Tribute to Kurt Gödel 2020, Brno

[1] supported by EPAC, grant 19-27871X of the Czech Grant Agency

Overview

sentences stronger than consistency statements

$Prov_{PA}(x)$ – a formalization of *"sentence x is provable in PA"*[2]

[2] I will talk about PA (Peano Arithmetic), but everything holds true also for other theories.

sentences stronger than consistency statements

$Prov_{PA}(x)$ – a formalization of *"sentence x is provable in PA"*[2]

$Con(PA)$ – a formalization of *"PA is consistent"*

$$Con(PA) \equiv \neg Prov_{PA}(\ulcorner 0 = 1 \urcorner)$$

[2] I will talk about PA (Peano Arithmetic), but everything holds true also for other theories.

1. Iterated consistency statements

the consistency of $PA + Con(PA)$, formally

$$Con(PA + Con(PA))$$

Proposition

$Con(PA + Con(PA))$ is strictly stronger than $Con(PA)$.

Proof.

Suppose it is not. Then

$$PA \vdash Con(PA) \to Con(PA + Con(PA))$$

This is equivalent to

$$PA + Con(PA) \vdash Con(PA + Con(PA))$$

which contradicts to the 2. incompleteness theorem for $PA + Con(PA)$. $\square$

We can go on and get stronger and stronger sentences

$$Con(PA + Con(PA + Con(PA)))$$

$$Con(PA + Con(PA + Con(PA + Con(PA))))$$

etc.

We can go on and get stronger and stronger sentences

$$Con(PA + Con(PA + Con(PA)))$$

$$Con(PA + Con(PA + Con(PA + Con(PA))))$$

etc.

Lemma

$$Con(PA + Con(PA)) \equiv \neg Prov_{PA}(\ulcorner \neg Con(PA) \urcorner)$$

2. Reflection principles

reflection principle for sentence ϕ: *if* ϕ *is provable, then* ϕ *is true;* formally

$$Prov_{PA}(\ulcorner \phi \urcorner) \to \phi$$

2. Reflection principles

reflection principle for sentence ϕ: *if ϕ is provable, then ϕ is true;* formally

$$Prov_{PA}(\ulcorner \phi \urcorner) \rightarrow \phi$$

Proposition

- *For ϕ equal to $0 = 1$, the reflection principle is equivalent to $Con(PA)$.*
- *For some ϕ, the reflection principle does not follow from consistency.*

Proof.

Take $\phi := \neg Con(PA)$. Then the reflection principle for ϕ is

$$Prov_{PA}(\ulcorner \neg Con(PA) \urcorner) \to \neg Con(PA)$$

Equivalently,

$$Con(PA) \to \neg Prov_{PA}(\ulcorner \neg Con(PA) \urcorner)$$

By Lemma, this is equivalent to

$$Con(PA) \to Con(PA + Con(PA))$$

Proof.

Take $\phi := \neg Con(PA)$. Then the reflection principle for ϕ is

$$Prov_{PA}(\ulcorner \neg Con(PA) \urcorner) \to \neg Con(PA)$$

Equivalently,

$$Con(PA) \to \neg Prov_{PA}(\ulcorner \neg Con(PA) \urcorner)$$

By Lemma, this is equivalent to

$$Con(PA) \to Con(PA + Con(PA))$$

Argunig by contradiction, suppose that the reflection principle is provable from $Con(PA)$. Formally,

$$PA \ \vdash \ Con(PA) \to (Con(PA) \to Con(PA + Con(PA))),$$

which is equivalent to

$$PA + Con(PA) \ \vdash \ Con(PA + Con(PA))).$$

But this contradicts to the 2. incompleteness theorem for $PA + Con(PA)$.

□

Uniform reflection principles

The uniform Σ_k reflection principle:

For every Σ_k sentence ϕ, if ϕ is provable in PA, then ϕ is true.

Formally it is an *arithmetical sentence*

$$\forall x \in \Sigma_k(Prov_{PA}(x) \to True_{\Sigma_k}(x)).$$

Uniform reflection principles

The uniform Σ_k reflection principle:

For every Σ_k sentence ϕ, if ϕ is provable in PA, then ϕ is true.

Formally it is an *arithmetical sentence*

$$\forall x \in \Sigma_k(Prov_{PA}(x) \to True_{\Sigma_k}(x)).$$

Note: We cannot define $True(x)$ for all arithmetical sentences.

Uniform reflection principles

The uniform Σ_k reflection principle:

For every Σ_k sentence ϕ, if ϕ is provable in PA, then ϕ is true.

Formally it is an *arithmetical sentence*

$$\forall x \in \Sigma_k(Prov_{PA}(x) \to True_{\Sigma_k}(x)).$$

Note: We cannot define $True(x)$ for all arithmetical sentences.

Proposition

Already the Σ_1-uniform reflection principle implies all iterated consistency statements.

Proof.

- easy exercise. □

Essentially all independent combinatorial sentences that we know are equivalent to Σ_1-reflection principles.

In particular, the Paris-Harrington Theorem is equivalent to the Σ_1-reflection principle for PA.

Soundness

In *metatheory* we can state *soundness* of PA. Formally it is the sentence

$$\forall x \in \mathit{ArithSent}\ (\mathit{Prov}_{\mathit{PA}}(x) \to \mathit{True}_{\mathit{ArithSent}}(x)),$$

where ArithSent is the set of arithmetical sentences. *This is not an arithmetical sentence.*

Soundness

In *metatheory* we can state *soundness* of PA. Formally it is the sentence

$$\forall x \in \mathit{ArithSent}\ (\mathit{Prov}_{PA}(x) \to \mathit{True}_{\mathit{ArithSent}}(x)),$$

where ArithSent is the set of arithmetical sentences. *This is not an arithmetical sentence.*

Proposition

ZFC proves the soundness of PA.

Proof.

ZFC proves that $\mathbb{N}$ is a model of PA. $\square$

The Lucas-Penrose falacy

J. R. Lucas:

"...given any machine which is consistent and capable of doing simple arithmetic, there is a formula which it is incapable of producing as being true ...which we can see to be true. It follows ...that minds are essentially different from machines."[3]

[3] Minds, machines and Gödel, Philosophy, 1961.

The Lucas-Penrose falacy

J. R. Lucas:

"...given any machine which is consistent and capable of doing simple arithmetic, there is a formula which it is incapable of producing as being true ... which we can see to be true. It follows ... that minds are essentially different from machines."[3]

A serious scientist should ask himself (herself):

Why *"we can see to be true"?*

[3] Minds, machines and Gödel, Philosophy, 1961.

The Lucas-Penrose falacy

J. R. Lucas:

"...given any machine which is consistent and capable of doing simple arithmetic, there is a formula which it is incapable of producing as being true ...which we can see to be true. It follows ...that minds are essentially different from machines."[3]

A serious scientist should ask himself (herself):

Why *"we can see to be true"?*

If you asked them they would probably answer: *because we are different from machines.*

[3] Minds, machines and Gödel, Philosophy, 1961.

What is wrong in these arguments?

What is wrong in these arguments?

The 2nd incompleteness theorem *does* apply to human mind. All mathematical assumptions a typical mathematician uses can be encapsulated into

$$\mathit{ZFC} + \exists \text{ inaccessible cardinal}$$

Because this theory proves the arithmetical soundness of ZFC.

What is wrong in these arguments?

The 2nd incompleteness theorem *does* apply to human mind. All mathematical assumptions a typical mathematician uses can be encapsulated into

$$ZFC + \exists \text{ inaccessible cardinal}$$

Because this theory proves the arithmetical soundness of ZFC.

Answer: **Simple logical errors** such as starting with an assumption and then using a different one, introducing another assumption in the course of the proof, etc.

Most frequent error: *failure to distinguish between consistency and soundness.*

Example

"Even if we adjoin to a formal system the infinite set of axioms consisting of Gödelian formulae, the resulting system is still incomplete, and conatins a formula which cannot be proved-in-he-system, although a rational being can, standing outside the system, see that it is true."[4]

[4] Lucas, the same article.

Example

"Even if we adjoin to a formal system the infinite set of axioms consisting of Gödelian formulae, the resulting system is still incomplete, and conatins a formula which cannot be proved-in-he-system, although a rational being can, standing outside the system, see that it is true."[4]

Let S be the system, S extended with Gödelian formulae is

$$T := S + Con(S) + Con(S + Con(S)) + Con(S + Con(S + Con(S))) + \dots$$

The "rational being" not only assumes that *S is consistent*, but in fact that *S is sound*. We know that already a weak form of soundness (Σ_1-reflection principle for S) implies the consistency of T.

[4]Lucas, the same article.

What about Gödel?

[5] K. Gödel, Some basic theorems on the foundations of mathematics and their implications.

What about Gödel?

- Gödel thought that it is possible (maybe even believed) that human mind is superior to machines,
- but also he was aware of the fact that the 2nd incompleteness theorem cannot be used to prove it.

[5] K. Gödel, Some basic theorems on the foundations of mathematics and their implications.

What about Gödel?

- Gödel thought that it is possible (maybe even believed) that human mind is superior to machines,
- but also he was aware of the fact that the 2nd incompleteness theorem cannot be used to prove it.

"Either... the human mind ... infinitely surpasses the powers of any finite machine, or else there exist absolutely unsolvable diophantine problems."[5]

[5] K. Gödel, Some basic theorems on the foundations of mathematics and their implications.

What about Gödel?

- Gödel thought that it is possible (maybe even believed) that human mind is superior to machines,
- but also he was aware of the fact that the 2nd incompleteness theorem cannot be used to prove it.

"Either... the human mind ... infinitely surpasses the powers of any finite machine, or else there exist absolutely unsolvable diophantine problems."[5]

How can Lucas and Penrose believe that Gödel overlooked their simple arguments that, as they think, eliminate the second possibility?

[5] K. Gödel, Some basic theorems on the foundations of mathematics and their implications.

What about Gödel?

- Gödel thought that it is possible (maybe even believed) that human mind is superior to machines,
- but also he was aware of the fact that the 2nd incompleteness theorem cannot be used to prove it.

"Either... the human mind ... infinitely surpasses the powers of any finite machine, or else there exist absolutely unsolvable diophantine problems."[5]

How can Lucas and Penrose believe that Gödel overlooked their simple arguments that, as they think, eliminate the second possibility?

More about this in my book *Logical Foundations of Mathematics and Computational Complexity*, Chapter 7.

[5] K. Gödel, Some basic theorems on the foundations of mathematics and their implications.

Proofs without selfreference

A proof of the 1st incompleteness theorem based on Kolmogorov's complexity[6]

Let U be a universal Turing machine, such that

1. For every binary string x, $U(x)$ is a binary string, or undefined if the machine does not stop.
2. For every other machine M of this kind, there exists a binary string p such that for all x, $U(px) = M(x)$.

Definition

The Kolmogorov complexity of a binary string y, denoted by $C(x)$, is the least n such that there exists a string x, $|x| = n$ such that $U(x) = y$.

Lemma

For every n there exists y with $|y| = n$ and $C(y) \geq n$.

Proof - simple countig.

[6]Probably due to G. J. Chaitin

Theorem

For every consistent recursively axiomatized consistent theory T, there exists a constant k_T such that T does not prove $C(a) > k_T$ for any concrete string a.

Theorem

For every consistent recursively axiomatized consistent theory T, there exists a constant k_T such that T does not prove $C(a) > k_T$ for any concrete string a.

Proof.

Let k be sufficeintly larger than the length of the description of T. Suppose T proves $K(a) > k$ for some string a. Let a be such a string with the shortest T-proof of $K(a) > k$. Then a can produced by an algorith as follows:

systematically generate all T-proofs;
stop and output a if a proof of $K(a) > k$ is found.

The Kolmogorov complexity of this algorithm is essentially the length of the desription of T plus $\log k$. $\square$

Theorem

For every consistent recursively axiomatized consistent theory T, there exists a constant k_T such that T does not prove $C(a) > k_T$ for any concrete string a.

Proof.

Let k be sufficeintly larger than the length of the description of T. Suppose T proves $K(a) > k$ for some string a. Let a be such a string with the shortest T-proof of $K(a) > k$. Then a can produced by an algorith as follows:

`systematically generate all` T`-proofs;`
`stop and output` a `if a proof of` $K(a) > k$ `is found.`

The Kolmogorov complexity of this algorithm is essentially the length of the desription of T plus $\log k$. □

Berry's Paradox

If $T \subseteq S$, then $k_T \leq k_S$.

$k_T \leq K(T)$+constant, but it may be much smaller.

A proof of the 2nd incompleteness theorem based on Kolmogorov's complexity[7]

Definition

A string a of length such that $K(a) \geq n$ is called *Kolmogorov random*. Denote by R_n be the number of Kolmogorov random strings of length n.

Lemma

Let T be consistent recursively axiomatized, $T \supseteq Q$ and let $n > k_T$. If T proves

$$\exists \text{ at least } M \text{ Kolmogorov random strings},$$

then $M < R_n$.

[7] S.Kritchman, R.Raz, *The Surprise Examination Paradox and the Second Incompleteness Theorem* (2010)

Proof.

1. For every a K. nonrandom, T can prove that it is K. nonrandom. Hence T proves that there are at least $2^n - R_n$ nonrandom strings. Hence $M \leq R_n$.
2. Suppose $M = R_n$. Since T proves for $2^n - R_n$ strings that they are K. nonrandom and proves that there are at least M (which is $= R_n$) K. random, it proves that x is K.-nonrandom for every K. nonrandom string x. This contradicts $n > k_T$.

□

Proof of the 2nd Incompleteness Theorem.

By formalizing the lemma in T, we can show that T proves

- If $Con(T)$, then there are more K. random strings than T can prove.

So if T proved $Con(T)$, it would be inconsistent. $\square$

Proof of the 2nd Incompleteness Theorem.

By formalizing the lemma in T, we can show that T proves

- If $Con(T)$, then there are more K. random strings than T can prove.

So if T proved $Con(T)$, it would be inconsistent. $\square$

Theorem

Let T be consistent and $n > k_T$. Then the sentence

$$\exists \text{ exactly } R_n \text{ Kolmogorov random strings}$$

is not provable in T.

How many Kolmogorov random srtings of length n are there?

How many Kolmogorov random srtings of length n are there?

- By the counting argument, at least one.

How many Kolmogorov random srtings of length n are there?

- By the counting argument, at least one.
- There are at least 2.

Proof.

Suppose there is only one. Run in paralele $U(x)$ on all strings x, $|x| < n$. After you get all $|y| \leq n$ as $y = U(x)$ except for one, print the remining one. This is a program shorter than n. $\square$

How many Kolmogorov random srtings of length n are there?

- By the counting argument, at least one.
- There are at least 2.

Proof.

Suppose there is only one. Run in paralele $U(x)$ on all strings x, $|x| < n$. After you get all $|y| \leq n$ as $y = U(x)$ except for one, print the remining one. This is a program shorter than n. □

- Similarly, there are at least 3.
- etc.

Proposition

The number R_n of Kolmogorov random strings of length n satisfies

$$K(R_n) \approx n.$$

A finite version of the 2nd incompleteness theorem

Definitions and notation

$Con_T \ \equiv_{df}$ there is no proof of contradiction in T

$Con_T(n) \ \equiv_{df}$ there is no proof of contradiction in T of length $\leq n$ (where n is represented by a term of length $O(\log n)$.)

A finite version of the 2nd incompleteness theorem

Definitions and notation

$Con_T \;\equiv_{df}$ there is no proof of contradiction in T

$Con_T(n) \;\equiv_{df}$ there is no proof of contradiction in T of length $\leq n$
(where n is represented by a term of length $O(\log n)$.)

$||\phi||_T$ is the length of the shortest proof of ϕ in T.

- $Con_T(n) \;\equiv\; ||0 = 1||_T > n.$
- $Con_T \;\equiv\; \forall n \; Con_T(n).$

Theorem (Friedman 1979, Pudlák 1984)

Let T be a consistent and sufficiently strong finitely axiomatized theory. Then for some $\epsilon > 0$,

$$||Con_T(n)||_T > n^{\epsilon}.$$

Theorem (Friedman 1979, Pudlák 1984)

Let T be a consistent and sufficiently strong finitely axiomatized theory. Then for some $\epsilon > 0$,

$$||Con_T(n)||_T > n^\epsilon.$$

Remark

- If $T \vdash \forall x \phi(x)$, then $||\phi(n)||_T = O(\log n)$. Hence $T \nvdash \forall x\ Con_T(x)$ which is just Con_T.
- Not only it is consistent with T that there exists a proof of contradiction, but one can show that *it can be "small"*.

Proof-idea

First recall Gödel's proof of the 2nd incompleteness theorem.

1. define $\gamma \ \equiv \ \neg \, Prov_T(\ulcorner\gamma\urcorner)$,
2. prove that if T is consistent, then T does not prove γ,
3. formalize 2. in T and get

$$T \vdash Con_T \to \neg \, Prov_T(\ulcorner\gamma\urcorner)$$

4. by definition of γ this implies

$$T \vdash Con_T \to \gamma$$

and since γ is not provable, also Con_T is not provable.

1. define $\delta(n) \equiv$ *"$\delta(n)$ does not have a proof of length $\leq n$"*; formally
$$\delta(n) \equiv ||\delta(n)||_T > n,$$
2. prove that if T is consistent, then $||\delta(n)||_T > n$,
3. formalize this proof in T and show that
$$Con_T(n^{O(1)}) \rightarrow ||\delta(n)||_T > n$$
has a short T-proof,
4. which is
$$Con_T(n^{O(1)}) \rightarrow \delta(n),$$
5. since $\delta(n)$ does not have a short T-proof, also $Con_T(n^{O(1)})$ cannot have a short proof.

Conjecture (Friedman, FALSE!)

$||Con_T(n)||_T$ *grows exponentially.*

[8] P. Hrubeš constructed a Π_1 sentence ϕ such that $T \nvdash \phi$, yet $||Con_{T+\phi}(n)||_T$ is polynomially bounded.

Conjecture (Friedman, FALSE!)

$||Con_T(n)||_T$ *grows exponentially.*

Conjecture (Mycielski)

$||Con_{T+Con_T}(n)||_T$ *grows exponentially.*

[8] P. Hrubeš constructed a Π_1 sentence ϕ such that $T \nvdash \phi$, yet $||Con_{T+\phi}(n)||_T$ is polynomially bounded.

Conjecture (Friedman, FALSE!)

$||Con_T(n)||_T$ *grows exponentially.*

Conjecture (Mycielski)

$||Con_{T+Con_T}(n)||_T$ *grows exponentially.*

Conjecture

$||Con_S(n)||_T$ *grows exponentially for every* S *that is sufficiently stronger than* T*.*[8]

Conjecture implies $\mathbf{P} \neq \mathbf{NP}$ (in fact even $\mathbf{NEXP} \neq \mathbf{coNEXP}$).

[8] P. Hrubeš constructed a Π_1 sentence ϕ such that $T \nvdash \phi$, yet $||Con_{T+\phi}(n)||_T$ is polynomially bounded.

Thank you!

UTE FOR ADVANCED STUDY ARCHIVES