# Gödel's
# Second Incompleteness Theorem
# (G2)

**Selmer Bringsjord**

Rensselaer AI & Reasoning (RAIR) Lab
Department of Cognitive Science
Department of Computer Science
Lally School of Management & Technology
Rensselaer Polytechnic Institute (RPI)
Troy, New York 12180 USA

4/7/2025

R A I R
Rensselaer AI and Reasoning Lab

# Gödel's
# Second Incompleteness Theorem
## (G2)

**Selmer Bringsjord**

Rensselaer AI & Reasoning (RAIR) Lab
Department of Cognitive Science
Department of Computer Science
Lally School of Management & Technology
Rensselaer Polytechnic Institute (RPI)
Troy, New York 12180 USA

4/7/2025

# Gödel's
# <u>Second</u> Incompleteness Theorem
# (G2)


**Selmer Bringsjord**

Rensselaer AI & Reasoning (RAIR) Lab
Department of Cognitive Science
Department of Computer Science
Lally School of Management & Technology
Rensselaer Polytechnic Institute (RPI)
Troy, New York 12180 USA


4/7/2025

R A I R
Rensselaer AI and Reasoning Lab

OXFORD UNIVERSITY PRESS

# Gödel's
# Second Incompleteness Theorem
## (G2)

**Selmer Bringsjord**

Rensselaer AI & Reasoning (RAIR) Lab
Department of Cognitive Science
Department of Computer Science
Lally School of Management & Technology
Rensselaer Polytechnic Institute (RPI)
Troy, New York 12180 USA

4/7/2025

R A I R
Rensselaer AI and Reasoning Lab

OXFORD UNIVERSITY PRESS

# Gödel's
# Second Incompleteness Theorem
## (G2)

**Selmer Bringsjord**

Rensselaer AI & Reasoning (RAIR) Lab
Department of Cognitive Science
Department of Computer Science
Lally School of Management & Technology
Rensselaer Polytechnic Institute (RPI)
Troy, New York 12180 USA

4/7/2025

**Note**: This is a version designed for those who have had at least one university-level course in formal logic with coverage through $\mathscr{L}_2$.

Rensselaer AI and Reasoning Lab

# Background Context …

# *Gödel's Great Theorems* (OUP)

by Selmer Bringsjord

- Introduction ("The Wager")
- Brief Preliminaries (e.g. the propositional calculus & FOL)
- The Completeness Theorem
- The First Incompleteness Theorem
- The Second Incompleteness Theorem
- The Speedup Theorem
- The Continuum-Hypothesis Theorem
- The Time-Travel Theorem
- Gödel's "God Theorem"
- Could a Finite Machine Match Gödel's Greatness?

STOP & REVIEW IF NEEDED!

# *Gödel's Great Theorems* (OUP)

by Selmer Bringsjord

- Introduction ("The Wager")
- Brief Preliminaries (e.g. the propositional calculus & FOL)
- The Completeness Theorem
- The First Incompleteness Theorem
- The Second Incompleteness Theorem
- The Speedup Theorem
- The Continuum-Hypothesis Theorem
- The Time-Travel Theorem
- Gödel's "God Theorem"
- Could a Finite Machine Match Gödel's Greatness?

STOP & REVIEW IF NEEDED!

# *Gödel's Great Theorems* (OUP)

by Selmer Bringsjord

- Introduction ("The Wager")
- Brief Preliminaries (e.g. the propositional calculus & FOL)
- The Completeness Theorem
- The First Incompleteness Theorem
- ➡ The Second Incompleteness Theorem
- The Speedup Theorem
- The Continuum-Hypothesis Theorem
- The Time-Travel Theorem
- Gödel's "God Theorem"
- Could a Finite Machine Match Gödel's Greatness?

A corollary of the First Incompleteness Theorem: *We cannot prove (in "classical" mathematics) that mathematics is consistent.*

# *Gödel's Great Theorems* (OUP)

by Selmer Bringsjord

- Introduction ("The Wager")
- Brief Preliminaries (e.g. the propositional calculus & FOL)
- The Completeness Theorem
- The First Incompleteness Theorem
- The Second Incompleteness Theorem
- The Speedup Theorem
- The Continuum-Hypothesis Theorem
- The Time-Travel Theorem
- Gödel's "God Theorem"
- Could a Machine Match Gödel's Genius?

# *Gödel's Great Theorems* (OUP)

by Selmer Bringsjord

- Introduction ("The Wager")
- Brief Preliminaries (e.g. the propositional calculus & FOL)
- The Completeness Theorem
- The First Incompleteness Theorem
- The Second Incompleteness Theorem
- The Speedup Theorem
- The Continuum-Hypothesis Theorem
- The Time-Travel Theorem
- Gödel's "God Theorem"
- Could a Machine Match Gödel's Genius?

By far the greatest of GGT; Selm's analysis based Sherlock Holmes' mystery "Silver Blaze."

# The “Gödelian” Liar (from me)

# The "Gödelian" Liar (from me)

$\bar{P}$: This sentence is unprovable.

# The "Gödelian" Liar (from me)

$\bar{P}$: This sentence is unprovable.

Suppose that $\bar{P}$ is true. Then we can immediately deduce that $\bar{P}$ is provable, because here is a proof: $\bar{P} \rightarrow \bar{P}$ is an easy theorem, and from it and our supposition we deduce $\bar{P}$ by *modus ponens*. But since what $\bar{P}$ says is that it's unprovable, we have deduced that $\bar{P}$ is false under our initial supposition.

# The "Gödelian" Liar (from me)

$\bar{P}$: This sentence is unprovable.

Suppose that $\bar{P}$ is true. Then we can immediately deduce that $\bar{P}$ is provable, because here is a proof: $\bar{P} \rightarrow \bar{P}$ is an easy theorem, and from it and our supposition we deduce $\bar{P}$ by *modus ponens*. But since what $\bar{P}$ says is that it's unprovable, we have deduced that $\bar{P}$ is false under our initial supposition.

Suppose on the other hand that $\bar{P}$ is false. Then we can immediately deduce that $\bar{P}$ is unprovable: Suppose for *reductio* that $\bar{P}$ *is* provable; then $\bar{P}$ holds as a result of some proof, but what $\bar{P}$ says is that it's unprovable; and so we have contradiction. But since what $\bar{P}$ says is that it's unprovable, and we have just proved that under our supposition, we arrive at the conclusion that $\bar{P}$ is true.

# The "Gödelian" Liar (from me)

$\bar{P}$: This sentence is unprovable.

Suppose that $\bar{P}$ is true. Then we can immediately deduce that $\bar{P}$ is provable, because here is a proof: $\bar{P} \rightarrow \bar{P}$ is an easy theorem, and from it and our supposition we deduce $\bar{P}$ by *modus ponens*. But since what $\bar{P}$ says is that it's unprovable, we have deduced that $\bar{P}$ is false under our initial supposition.

Suppose on the other hand that $\bar{P}$ is false. Then we can immediately deduce that $\bar{P}$ is unprovable: Suppose for *reductio* that $\bar{P}$ *is* provable; then $\bar{P}$ holds as a result of some proof, but what $\bar{P}$ says is that it's unprovable; and so we have contradiction. But since what $\bar{P}$ says is that it's unprovable, and we have just proved that under our supposition, we arrive at the conclusion that $\bar{P}$ is true.

T($\bar{P}$) iff (i.e., if & only if) ¬T($\bar{P}$) = F($\bar{P}$)

# The “Gödelian” Liar (from me)

$\bar{P}$: This sentence is unprovable.

Suppose that $\bar{P}$ is true. Then we can immediately deduce that $\bar{P}$ is provable, because here is a proof: $\bar{P} \rightarrow \bar{P}$ is an easy theorem, and from it and our supposition we deduce $\bar{P}$ by *modus ponens*. But since what $\bar{P}$ says is that it's unprovable, we have deduced that $\bar{P}$ is false under our initial supposition.

Suppose on the other hand that $\bar{P}$ is false. Then we can immediately deduce that $\bar{P}$ is unprovable: Suppose for *reductio* that $\bar{P}$ *is* provable; then $\bar{P}$ holds as a result of some proof, but what $\bar{P}$ says is that it's unprovable; and so we have contradiction. But since what $\bar{P}$ says is that it's unprovable, and we have just proved that under our supposition, we arrive at the conclusion that $\bar{P}$ is true.

$T(\bar{P})$ iff (i.e., if & only if) $\neg T(\bar{P}) = F(\bar{P})$

Contradiction!

# The "Gödelian" Liar (from me)

$\bar{\pi}$ = "I'm unprovable."

$\bar{P}$: This sentence is unprovable.

Suppose that $\bar{P}$ is true. Then we can immediately deduce that $\bar{P}$ is provable, because here is a proof: $\bar{P} \rightarrow \bar{P}$ is an easy theorem, and from it and our supposition we deduce $\bar{P}$ by *modus ponens*. But since what $\bar{P}$ says is that it's unprovable, we have deduced that $\bar{P}$ is false under our initial supposition.

Suppose on the other hand that $\bar{P}$ is false. Then we can immediately deduce that $\bar{P}$ is unprovable: Suppose for *reductio* that $\bar{P}$ *is* provable; then $\bar{P}$ holds as a result of some proof, but what $\bar{P}$ says is that it's unprovable; and so we have contradiction. But since what $\bar{P}$ says is that it's unprovable, and we have just proved that under our supposition, we arrive at the conclusion that $\bar{P}$ is true.

T($\bar{P}$) iff (i.e., if & only if) ¬T($\bar{P}$) = F($\bar{P}$)

Contradiction!

All of this is fishy; but Gödel, as we've seen, transformed it (by e.g. use of his encryption scheme) into utterly precise, impactful, indisputable reasoning …

# **PA** (Peano Arithmetic):

| | |
|---|---|
| A1 | $\forall x(0 \neq s(x))$ |
| A2 | $\forall x \forall y(s(x) = s(y) \rightarrow x = y)$ |
| A3 | $\forall x(x \neq 0 \rightarrow \exists y(x = s(y)))$ |
| A4 | $\forall x(x + 0 = x)$ |
| A5 | $\forall x \forall y(x + s(y) = s(x + y))$ |
| A6 | $\forall x(x \times 0 = 0)$ |
| A7 | $\forall x \forall y(x \times s(y) = (x \times y) + x)$ |

And, every sentence that is the universal closure of an instance of

$$([\phi(0) \wedge \forall x(\phi(x) \rightarrow \phi(s(x)))] \rightarrow \forall x \phi(x))$$

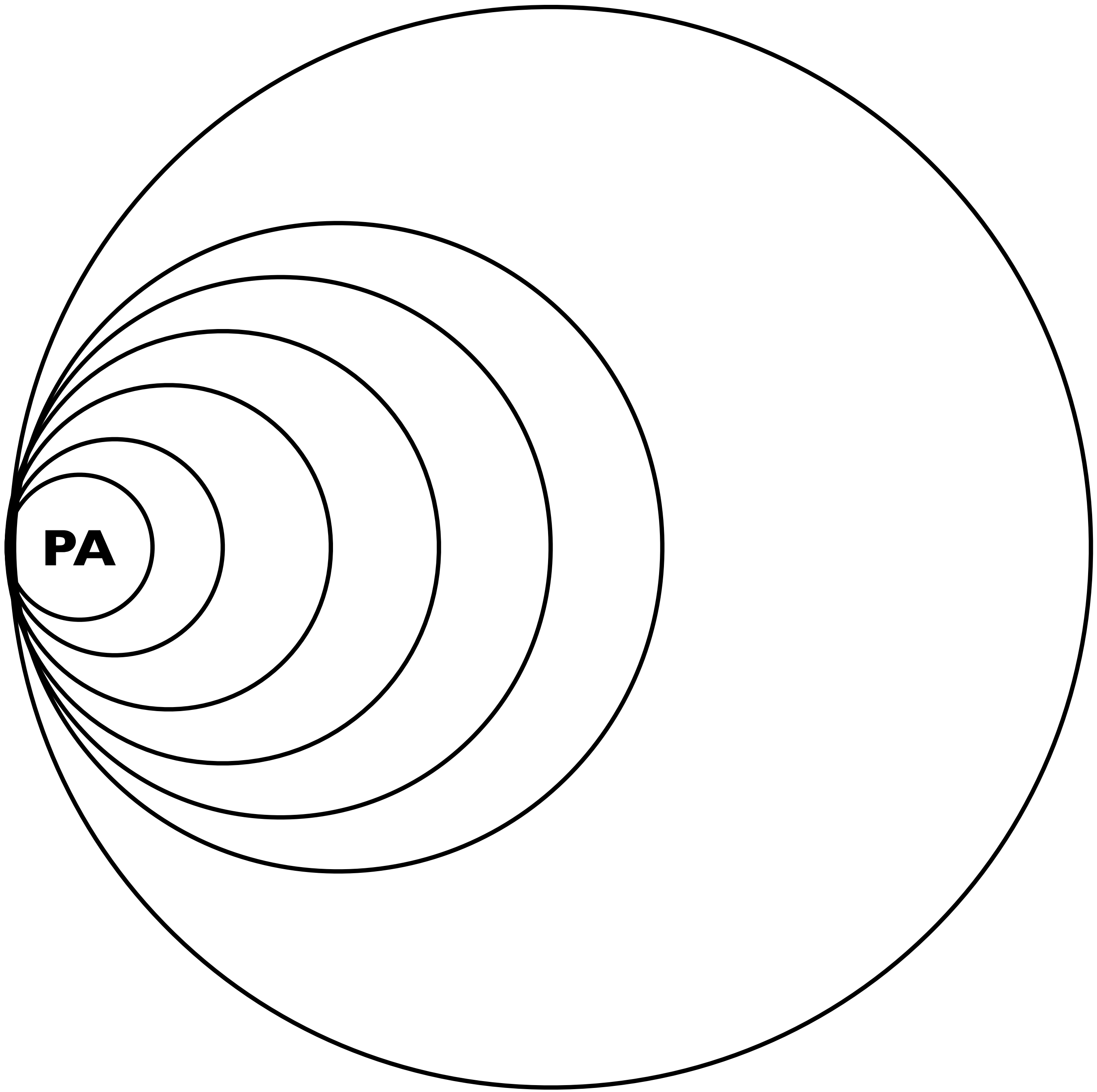
where $\phi(x)$ is open wff with variable $x$, and perhaps others, free.

# Is there buried inconsistency in here?!?

Courtesy of Gödel: Given certain limitative assumptions about "proof power," we can't prove that there isn't!

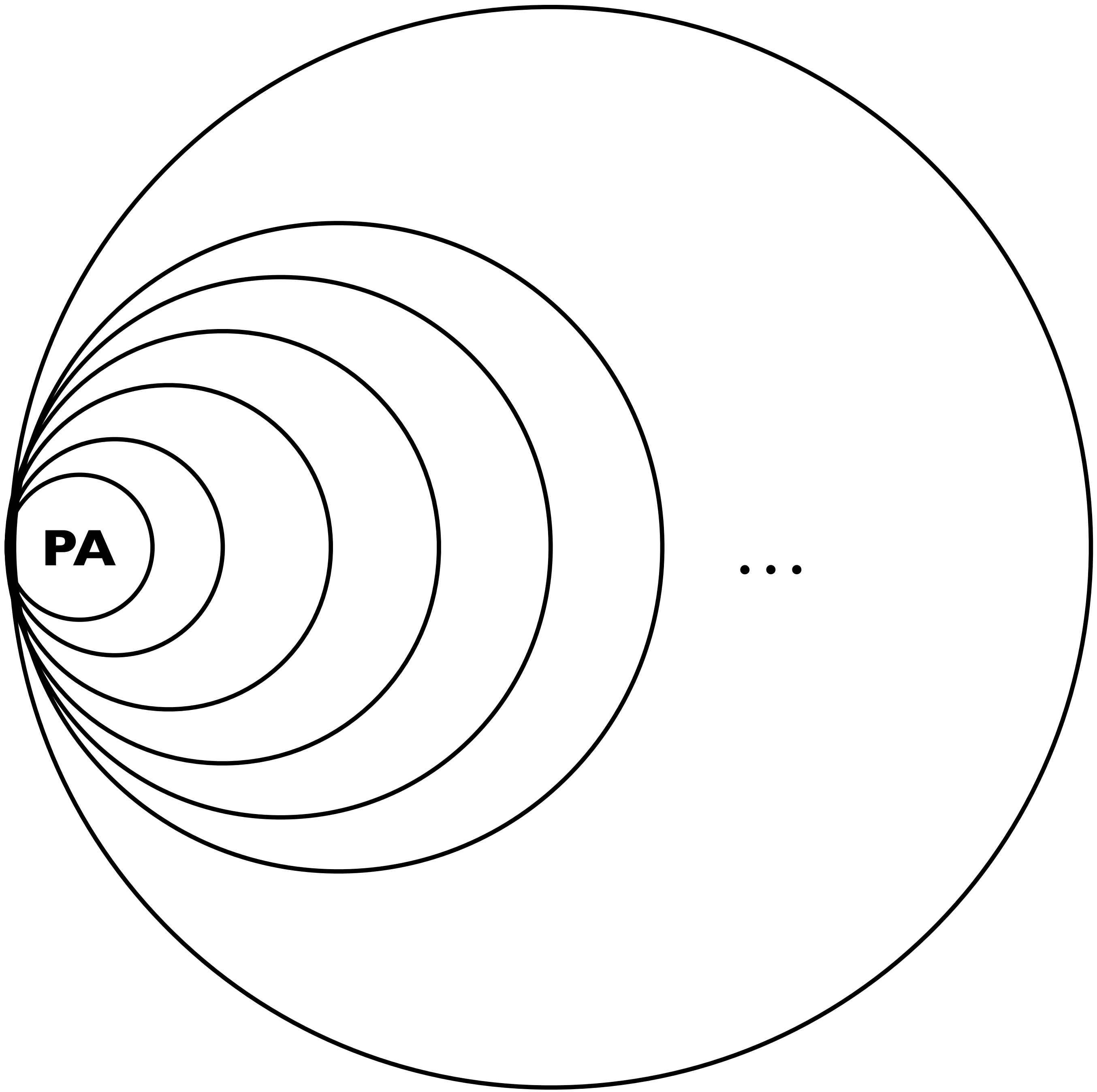
# Is there buried inconsistency in here?!?

Courtesy of Gödel: Given certain limitative assumptions about “proof power,” we can’t prove that there isn’t!



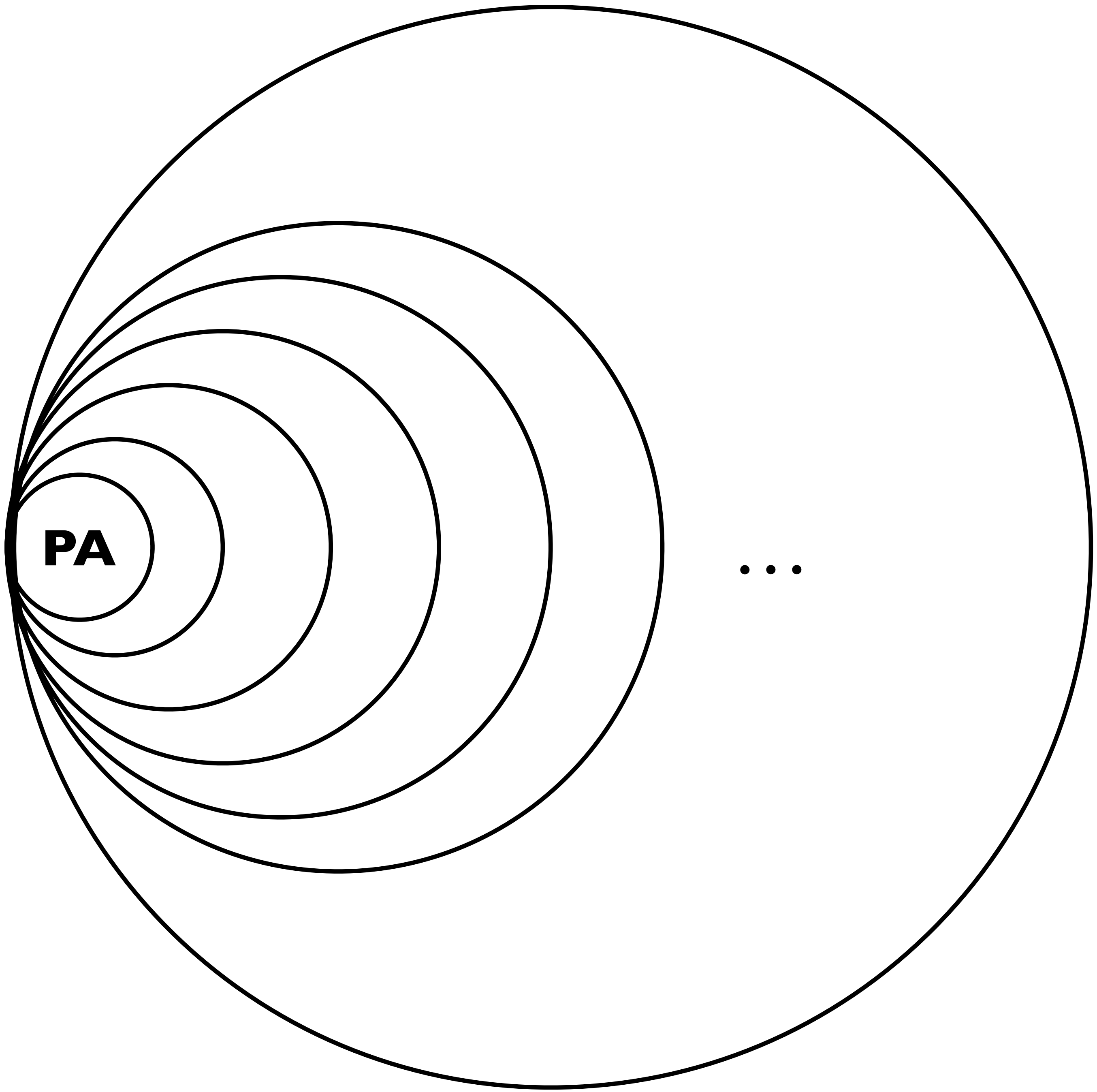
# Is there buried inconsistency in here?!?

Courtesy of Gödel: Given certain limitative assumptions about "proof power," we can't prove that there isn't!



# Is there buried inconsistency in here?!?

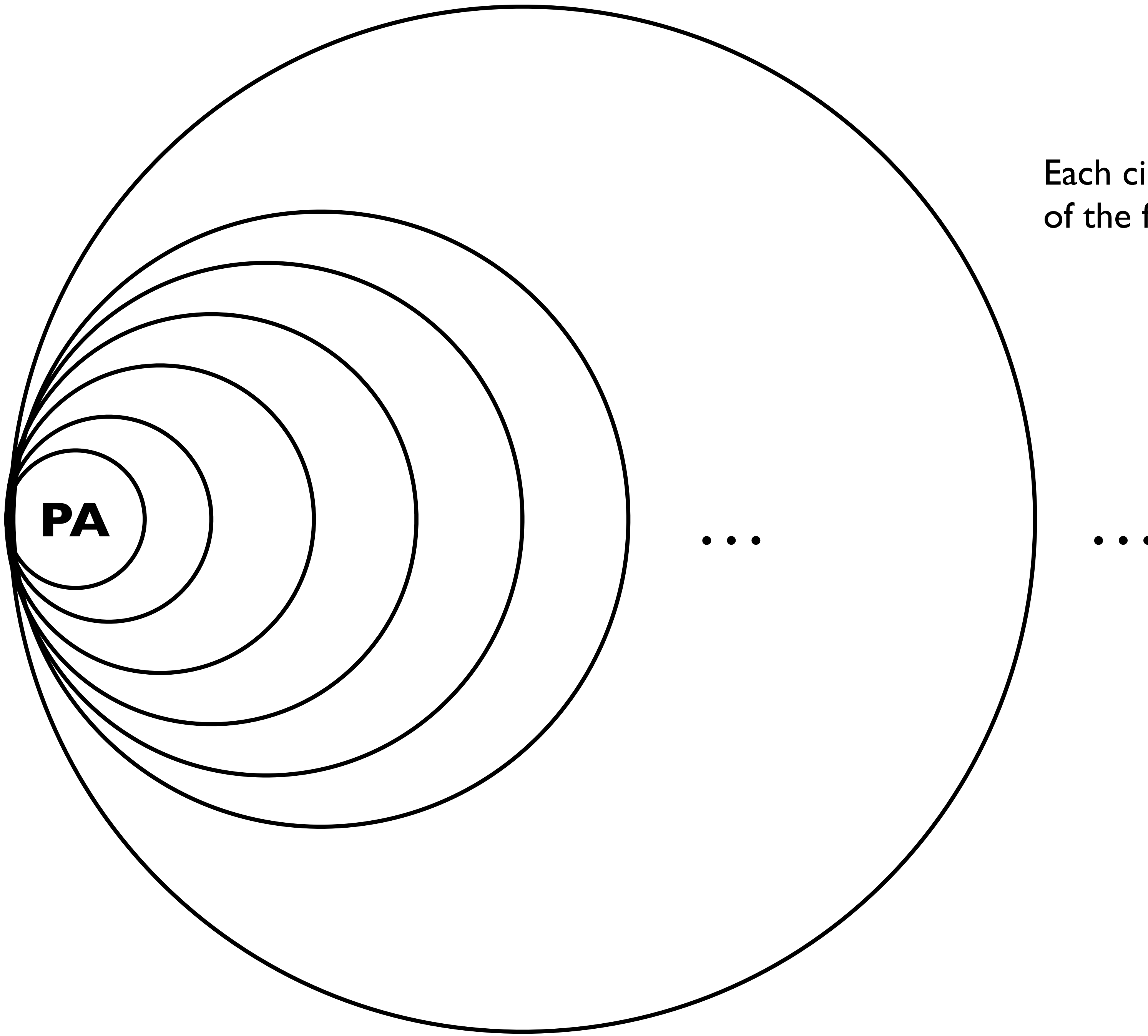
Courtesy of Gödel: Given certain limitative assumptions about "proof power," we can't prove that there isn't!




Each circle is a larger part of the formal sciences.

# Is there buried inconsistency in here?!?

Courtesy of Gödel: Given certain limitative assumptions about "proof power," we can't prove that there isn't!




Each circle is a larger part of the formal sciences.

# G2 as Slogan …

# G2 as Slogan ...

"We can't use math to ascertain whether mathematics is consistent."

# G2 as Slogan …

# G2 as Slogan …

"If we are restricted to certain kinds of formal reasoning, and feel we must have all of **PA** (math, engineering, etc.), we can't ascertain whether mathematics is consistent."

# Gödel's Second Incompleteness Theorem

# Gödel's Second Incompleteness Theorem

Suppose $\mathbf{\Phi} \supset \mathbf{PA}$ that is

(i) **Con** $\mathbf{\Phi}$;
(ii) Turing-decidable (i.e. membership in $\mathbf{\Phi}$ is Turing-decidable); and
(iii) sufficiently expressive to capture all of the operations of a Turing machine (i.e. **Repr** $\mathbf{\Phi}$).

Then $\mathbf{\Phi} \nvdash \text{consis}_{\mathbf{\Phi}}$.

# Gödel's <u>Second</u> Incompleteness Theorem

Remember Church's Theorem!

Suppose $\Phi \supset \mathbf{PA}$ that is

(i) **Con** $\Phi$;
(ii) Turing-decidable (i.e. membership in $\Phi$ is Turing-decidable); and
(iii) sufficiently expressive to capture all of the operations of a Turing machine (i.e. **Repr** $\Phi$).

Then $\Phi \nvdash consis_{\Phi}$.

# Gödel's <u>Second</u> Incompleteness Theorem

Remember Church's Theorem!

Suppose $\Phi \supset \mathbf{PA}$ that is

(i) **Con** $\Phi$;
(ii) Turing-decidable (i.e. membership in $\Phi$ is Turing-decidable); and
(iii) sufficiently expressive to capture all of the operations of a Turing machine (i.e. **Repr** $\Phi$).

Then $\Phi \nvdash \text{consis}_{\Phi}$.

To prove G2, we shall once again allow ourselves …

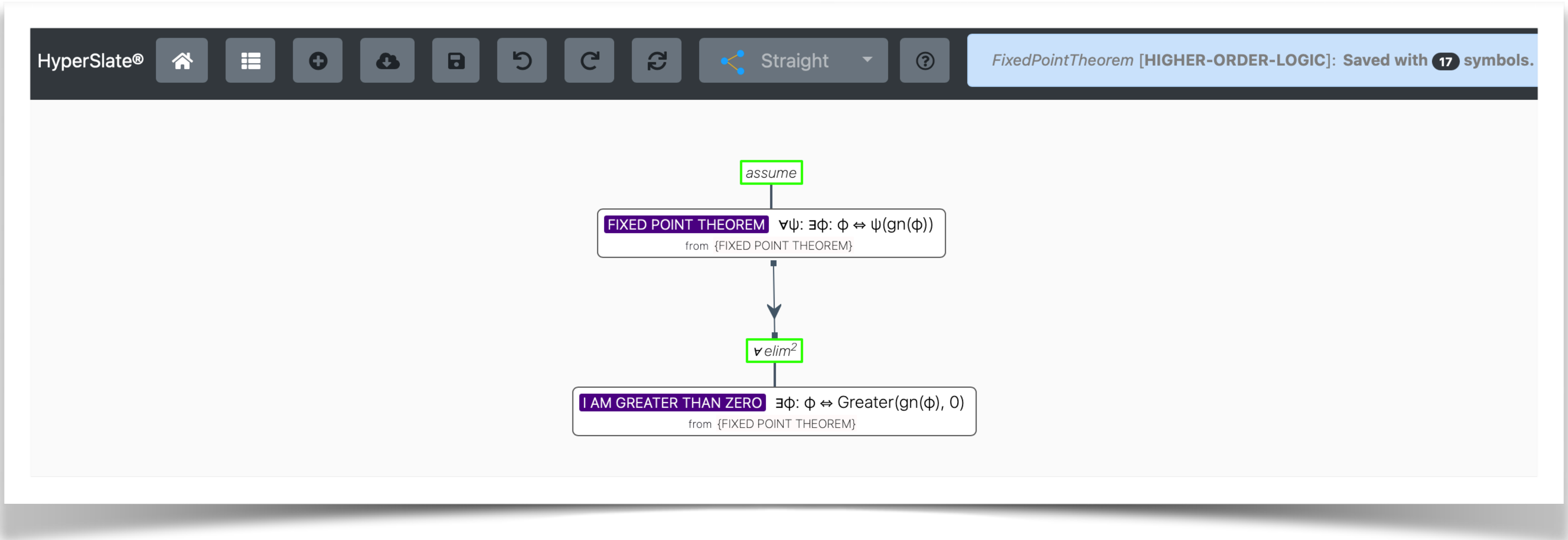
# The Fixed Point Theorem (FPT)

Assume that $\Phi$ is a set of arithmetic sentences such that **Repr** $\Phi$. There for every arithmetic formula $\psi(x)$ with one free variable $x$, there is an arithmetic sentence $\phi$ s.t.

$$\Phi \vdash \phi \leftrightarrow \psi(\hat{n}^{\phi}).$$

We can intuitively understand $\phi$ to be saying:
"I have the property ascribed to me by the formula $\psi$."

# FPT in HyperSlate®!



Ok; so let's do it … and let's see if you can see why Gödel declared G2 to be a direct "corollary" of G1, and didn't bother to prove it in his original paper …

**Proof**: Suppose that the antecedent (i)–(iii) of **G2** holds. Suppose for *reductio* that

$$\Phi \supseteq \mathbf{PA} \vdash \neg\pi(\hat{n}^{0=1})$$

(Supposition: We can prove from **PA** that this system is consistent.)

We have three ingredients which together suffice for getting to our goal. First, from Gödel's Self-Ascription Theorem (a.k.a. …) we can again directly obtain:

**Proof**: Suppose that the antecedent (i)–(iii) of **G2** holds. Suppose for *reductio* that

$$\Phi \supseteq \mathbf{PA} \vdash \neg\pi(\hat{n}^{0=1})$$

(Supposition: We can prove from **PA** that this system is consistent.)

We have three ingredients which together suffice for getting to our goal. First, from Gödel's Self-Ascription Theorem (a.k.a. …) we can again directly obtain:

(G2.1) $\Phi \supseteq \mathbf{PA} \vdash \mathscr{G} \leftrightarrow \neg\pi(\hat{n}^{\mathscr{G}})$ .

**Proof**: Suppose that the antecedent (i)–(iii) of **G2** holds. Suppose for *reductio* that

$$\Phi \supseteq \mathbf{PA} \vdash \neg\pi(\hat{n}^{0=1})$$

(Supposition: We can prove from **PA** that this system is consistent.)

We have three ingredients which together suffice for getting to our goal. First, from Gödel's Self-Ascription Theorem (a.k.a. …) we can again directly obtain:

(G2.1) $\Phi \supseteq \mathbf{PA} \vdash \mathscr{G} \leftrightarrow \neg\pi(\hat{n}^{\mathscr{G}})$ .

(It's provable from anything that includes **PA** that there's a Gödel sentence $\mathscr{G}$ that says "I'm unprovable.")

**Proof**: Suppose that the antecedent (i)–(iii) of **G2** holds. Suppose for *reductio* that

$$\Phi \supseteq \mathbf{PA} \vdash \neg\pi(\hat{n}^{0=1})$$

(Supposition: We can prove from **PA** that this system is consistent.)

We have three ingredients which together suffice for getting to our goal. First, from Gödel's Self-Ascription Theorem (a.k.a. …) we can again directly obtain:

(G2.1) $\Phi \supseteq \mathbf{PA} \vdash \mathscr{G} \leftrightarrow \neg\pi(\hat{n}^{\mathscr{G}})$ .

(It's provable from anything that includes **PA** that there's a Gödel sentence $\mathscr{G}$ that says "I'm unprovable.")

"I'm not provable from $\Phi$!"

**Proof**: Suppose that the antecedent (i)–(iii) of **G2** holds. Suppose for *reductio* that

$$\Phi \supseteq \mathbf{PA} \vdash \neg\pi(\hat{n}^{0=1})$$

(Supposition: We can prove from **PA** that this system is consistent.)

We have three ingredients which together suffice for getting to our goal. First, from Gödel's Self-Ascription Theorem (a.k.a. …) we can again directly obtain:

(G2.1) $\Phi \supseteq \mathbf{PA} \vdash \mathscr{G} \leftrightarrow \neg\pi(\hat{n}^{\mathscr{G}})$ .

(It's provable from anything that includes **PA** that there's a Gödel sentence $\mathscr{G}$ that says "I'm unprovable.")

*"I'm not provable from* $\Phi$*!"*

Next, we *already* proved (when? … from one half of **G1**!) that:

**Proof**: Suppose that the antecedent (i)–(iii) of **G2** holds. Suppose for *reductio* that

$$\Phi \supseteq \textbf{PA} \vdash \neg\pi(\hat{n}^{0=1})$$

(Supposition: We can prove from **PA** that this system is consistent.)

We have three ingredients which together suffice for getting to our goal. First, from Gödel's Self-Ascription Theorem (a.k.a. …) we can again directly obtain:

(G2.1) $\Phi \supseteq \textbf{PA} \vdash \mathscr{G} \leftrightarrow \neg\pi(\hat{n}^{\mathscr{G}})$ .

> "I'm not provable from $\Phi$!"

(It's provable from anything that includes **PA** that there's a Gödel sentence $\mathscr{G}$ that says "I'm unprovable.")

Next, we *already* proved (when? … from one half of **G1**!) that:

(G2.2) $\Phi \supseteq \textbf{PA} \nvdash \mathscr{G}$.

**Proof**: Suppose that the antecedent (i)–(iii) of **G2** holds. Suppose for *reductio* that

$$\Phi \supseteq \mathbf{PA} \vdash \neg\pi(\hat{n}^{0=1})$$

(Supposition: We can prove from **PA** that this system is consistent.)

We have three ingredients which together suffice for getting to our goal. First, from Gödel's Self-Ascription Theorem (a.k.a. …) we can again directly obtain:

(G2.1) $\Phi \supseteq \mathbf{PA} \vdash \mathscr{G} \leftrightarrow \neg\pi(\hat{n}^{\mathscr{G}})\,.$

"I'm not provable from **Φ**!"

(It's provable from anything that includes **PA** that there's a Gödel sentence $\mathscr{G}$ that says "I'm unprovable.")

Next, we *already* proved (when? … from one half of **G1**!) that:

(G2.2) $\Phi \supseteq \mathbf{PA} \nvdash \mathscr{G}$.

(Gödel's sentence $\mathscr{G}$ can't be proved from anything that includes **PA**.)

**Proof**: Suppose that the antecedent (i)–(iii) of **G2** holds. Suppose for *reductio* that

$$\Phi \supseteq \mathbf{PA} \vdash \neg\pi(\hat{n}^{0=1})$$

(Supposition: We can prove from **PA** that this system is consistent.)

We have three ingredients which together suffice for getting to our goal. First, from Gödel's Self-Ascription Theorem (a.k.a. …) we can again directly obtain:

(G2.1) $\Phi \supseteq \mathbf{PA} \vdash \mathscr{G} \leftrightarrow \neg\pi(\hat{n}^{\mathscr{G}})$ .

(It's provable from anything that includes **PA** that there's a Gödel sentence $\mathscr{G}$ that says "I'm unprovable.")

"I'm not provable from $\Phi$!"

Next, we *already* proved (when? … from one half of **G1**!) that:

(G2.2) $\Phi \supseteq \mathbf{PA} \nvdash \mathscr{G}$.

(Gödel's sentence $\mathscr{G}$ can't be proved from anything that includes **PA**.)

Thirdly, we can logicize the provable meta-logical conditional proposition that if $\Phi$ an prove $\mathscr{G}$, then $\Phi$ can prove that 0=1. I.e., we have:

**Proof**: Suppose that the antecedent (i)–(iii) of **G2** holds. Suppose for *reductio* that

$$\Phi \supseteq \mathbf{PA} \vdash \neg\pi(\hat{n}^{0=1})$$

(Supposition: We can prove from **PA** that this system is consistent.)

We have three ingredients which together suffice for getting to our goal. First, from Gödel's Self-Ascription Theorem (a.k.a. …) we can again directly obtain:

(G2.1) $\Phi \supseteq \mathbf{PA} \vdash \mathscr{G} \leftrightarrow \neg\pi(\hat{n}^{\mathscr{G}})$ .

"I'm not provable from $\Phi$!"

(It's provable from anything that includes **PA** that there's a Gödel sentence $\mathscr{G}$ that says "I'm unprovable.")

Next, we *already* proved (when? … from one half of **G1**!) that:

(G2.2) $\Phi \supseteq \mathbf{PA} \nvdash \mathscr{G}$.

(Gödel's sentence $\mathscr{G}$ can't be proved from anything that includes **PA**.)

Thirdly, we can logicize the provable meta-logical conditional proposition that if $\Phi$ an prove $\mathscr{G}$, then $\Phi$ can prove that 0=1. I.e., we have:

(G2.3) $\Phi \supseteq \mathbf{PA} \vdash \pi(\hat{n}^{\mathscr{G}}) \rightarrow \pi(\hat{n}^{0=1})$

**Proof**: Suppose that the antecedent (i)–(iii) of **G2** holds. Suppose for *reductio* that

$\Phi \supseteq$ **PA** $\vdash \neg\pi(\hat{n}^{0=1})$
(Supposition: We can prove from **PA** that this system is consistent.)

We have three ingredients which together suffice for getting to our goal. First, from Gödel's Self-Ascription Theorem (a.k.a. …) we can again directly obtain:

(G2.1) $\Phi \supseteq$ **PA** $\vdash \mathscr{G} \leftrightarrow \neg\pi(\hat{n}^{\mathscr{G}})$ .

"I'm not provable from $\Phi$!"

(It's provable from anything that includes **PA** that there's a Gödel sentence $\mathscr{G}$ that says "I'm unprovable.")

Next, we *already* proved (when? … from one half of **G1**!) that:

(G2.2) $\Phi \supseteq$ **PA** $\nvdash \mathscr{G}$.

(Gödel's sentence $\mathscr{G}$ can't be proved from anything that includes **PA**.)

Thirdly, we can logicize the provable meta-logical conditional proposition that if $\Phi$ an prove $\mathscr{G}$, then $\Phi$ can prove that 0=1. I.e., we have:

(G2.3) $\Phi \supseteq$ **PA** $\vdash \pi(\hat{n}^{\mathscr{G}}) \rightarrow \pi(\hat{n}^{0=1})$

Contradiction! Done! (Can you find it?) **QED**

*Med nok penger, kan logikk løse alle problemer.*