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

Intro to Formal Logic (With AI)
4/13/2026

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

Intro to Formal Logic (With AI)
4/13/2026

RAIR
Rensselaer AI and Reasoning Lab

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

Intro to Formal Logic (With AI)
4/13/2026

RAIR
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

Intro to Formal Logic (With AI)
4/13/2026

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

Inconsistency is — for classical logic/mathematics — like nuclear fission/fusion.

Inconsistency is — for classical logic/mathematics — like nuclear fission/fusion.

Foundation of Math Crumbles!

Inconsistency is — for classical logic/mathematics — like nuclear fission/fusion.

Foundation of Math Crumbles!

NYS 3 …

Inconsistency is — for classical logic/mathematics — like nuclear fission/fusion.

Foundation of Math Crumbles!

NYS 3 …

le Carré …

Inconsistency is — for classical logic/mathematics — like nuclear fission/fusion.

Foundation of Math Crumbles!

NYS 3 …

le Carré …

Famous Mathematician @ Plum Blossom: "Very funny, Selmer. No, there doesn't exist an *x* s.t. if it's a bald eagle then everything is!"

Inconsistency is — for classical logic/mathematics — like nuclear fission/fusion.

Foundation of Math Crumbles!

NYS 3 …

le Carré …

Famous Mathematician @ Plum Blossom: "Very funny, Selmer. No, there doesn't exist an *x* s.t. if it's a bald eagle then everything is!"

Plantinga's proof that God's existence is consistent with moral evil.

Inconsistency is — for classical logic/mathematics — like nuclear fission/fusion.

Foundation of Math Crumbles!

NYS 3 …

le Carré …

Famous Mathematician @ Plum Blossom: "Very funny, Selmer. No, there doesn't exist an *x* s.t. if it's a bald eagle then everything is!"

Plantinga's proof that God's existence is consistent with moral evil.

The RAIR Lab research gets to keep its funding (bc of inconsistency).

Inconsistency is — for classical logic/mathematics — like nuclear fission/fusion.

Foundation of Math Crumbles!

NYS 3 …

le Carré …

Famous Mathematician @ Plum Blossom: "Very funny, Selmer. No, there doesn't exist an *x* s.t. if it's a bald eagle then everything is!"

Plantinga's proof that God's existence is consistent with moral evil.

The RAIR Lab research gets to keep its funding (bc of inconsistency).

G2 — Liar Paradox exploited — entails that these researchers can keep their funding!

Background Context …

Could Artificial Intelligence Ever … Match Personal Intelligence?

Gödelian Essays on Persons vs. Pretenders

by Selmer Bringsjord et al.

- 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!

Could Artificial Intelligence Ever … Match Personal Intelligence?

Gödelian Essays on Persons vs. Pretenders

by Selmer Bringsjord et al.

- 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!

Could Artificial Intelligence Ever … Match Personal Intelligence?

Gödelian Essays on Persons vs. Pretenders

by Selmer Bringsjord et al.

- 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.*

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) ¬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):

$$
\begin{array}{ll}
\text{A1} & \forall x(0 \neq s(x)) \\
\text{A2} & \forall x \forall y(s(x) = s(y) \rightarrow x = y) \\
\text{A3} & \forall x(x \neq 0 \rightarrow \exists y(x = s(y)) \\
\text{A4} & \forall x(x + 0 = x) \\
\text{A5} & \forall x \forall y(x + s(y) = s(x + y)) \\
\text{A6} & \forall x(x \times 0 = 0) \\
\text{A7} & \forall x \forall y(x \times s(y) = (x \times y) + x)
\end{array}
$$

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.

RussellsLetter2Frege_IFLAI1S24

Problem Bank

Edit Problems

In-Class Exploratory

The challenge here is to prove that from Russell's instantiation of Frege's doomed Axiom V a contradiction can be promptly derived. The letter has of course been examined in some detail by S Bringsjord (in the Feb 29 2024 lecture in the 2024 lecture lineup). Put meta-logically, your task in the present problem is to build a proof that confirms this:

$$\{\exists x \forall y((y \in x) \rightarrow (y \notin y))\} \vdash \zeta \wedge \neg \zeta.$$

Make sure you understand that the given here is an instantiation of Frege's Axiom V; i.e. it's an instantiation of

$$\exists x \forall y((y \in x) \rightarrow \phi(y)).$$

(The notation $\phi(y)$, recall from lecture, is the standard way in mathematical logic to say that y is free in ϕ.) **Note**: Your finished proof is allowed to make use the PC-provability oracle (but *only* that oracle).

(Now a brief remark on matters covered by in class by Bringsjord when second-order logic = $\mathscr{L}_2$ arrives on the scene: Longer term, and certainly constituting evidence of Frege's capacity for ingenius, intricate deduction, it has recently been realized that while Frege himself relied on Axiom V to obtain what is known as **Hume's Principle** (= HP), this reliance is avoidable. That from just HP we can deduce all of Peano Arithmetic (**PA**) (!) is a result Frege can be credited with showing; the result is known today as Frege's Theorem (= FT). Following the link just given will reward the reader with an understanding of HP, and how how to obtain **PA** from it.)

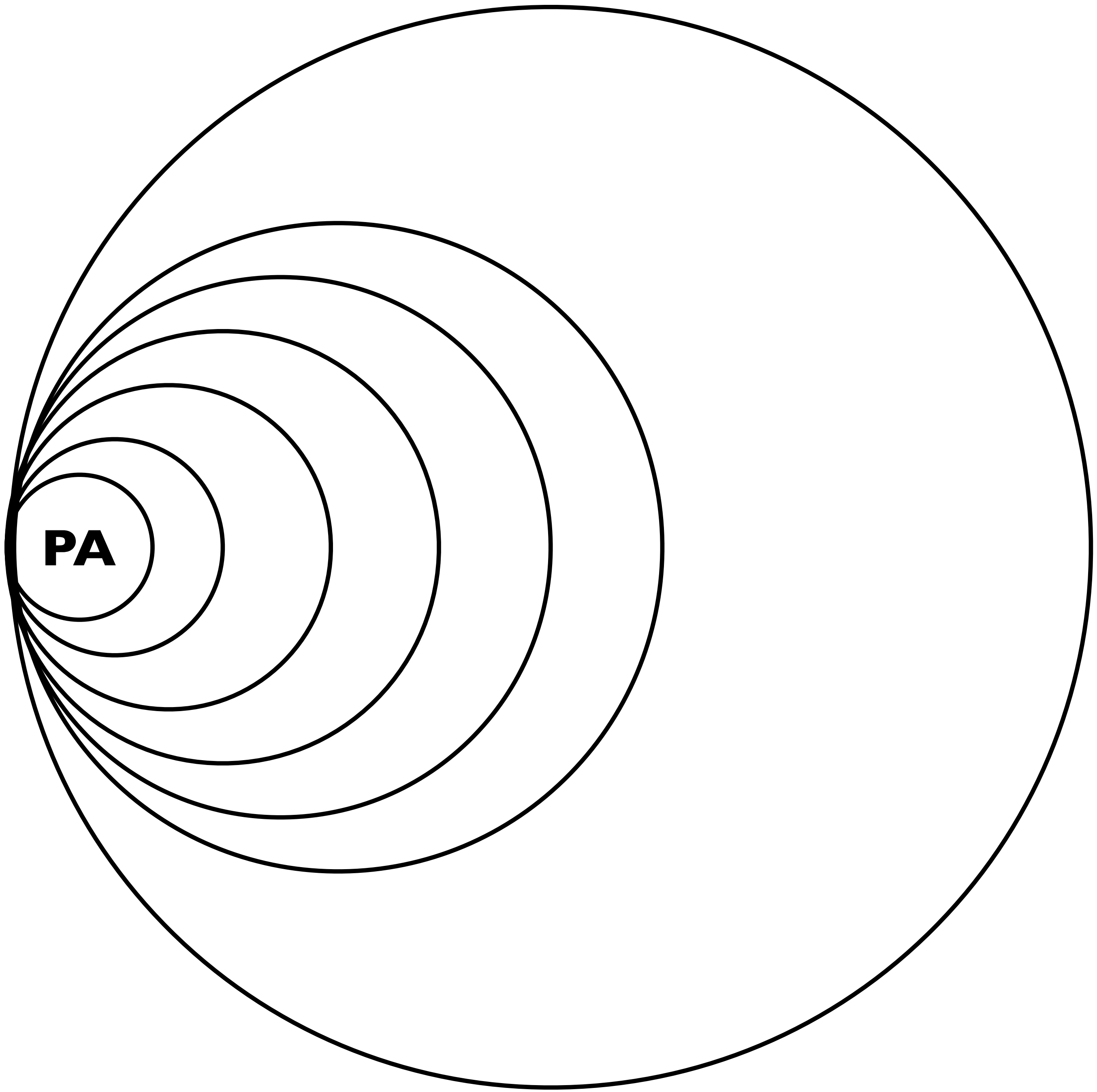
Deadline December 4, 2025 at 5:50 PM EST

Is there buried inconsistency in here?!?

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

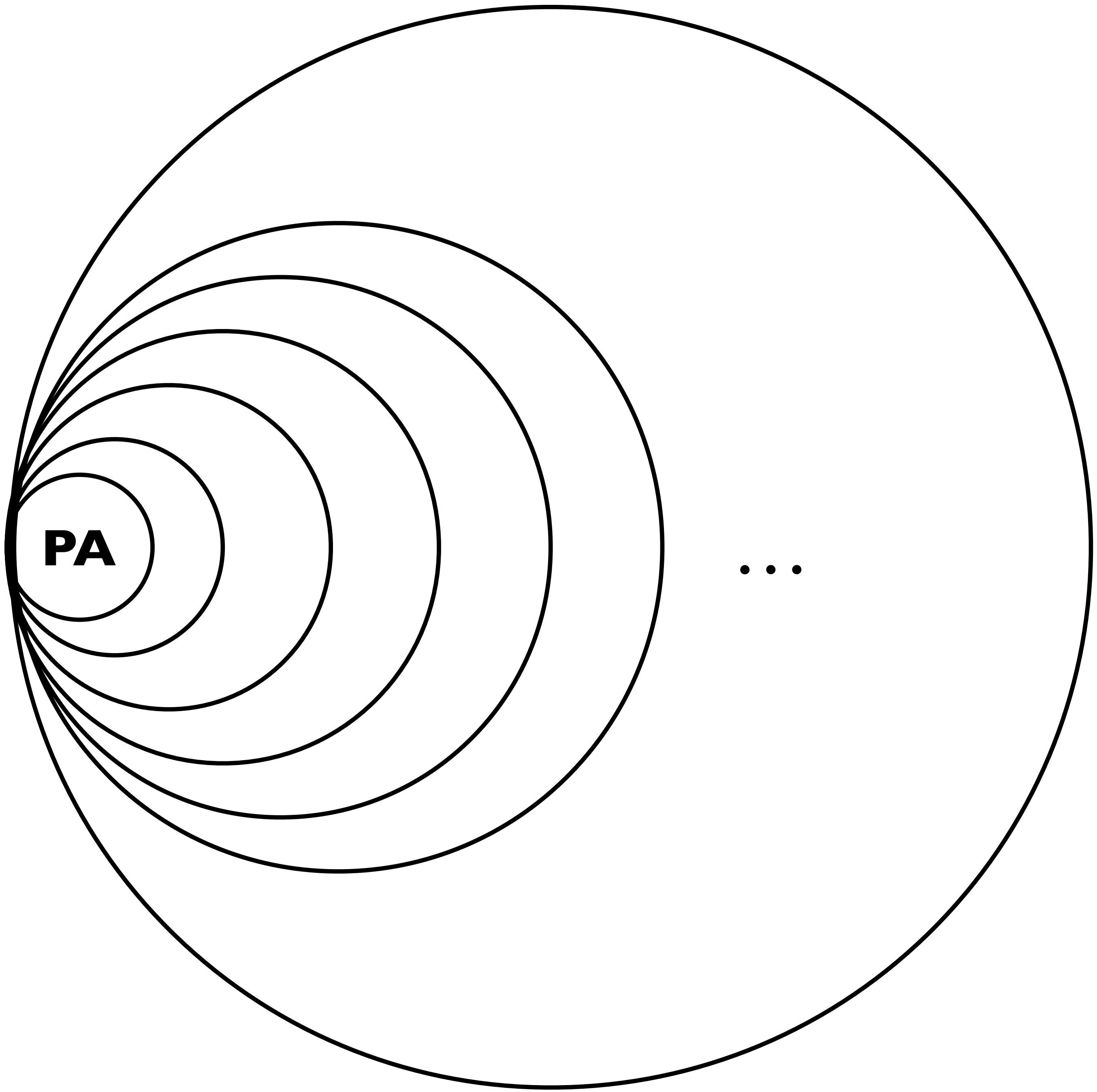
Is there buried inconsistency in here?!?

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



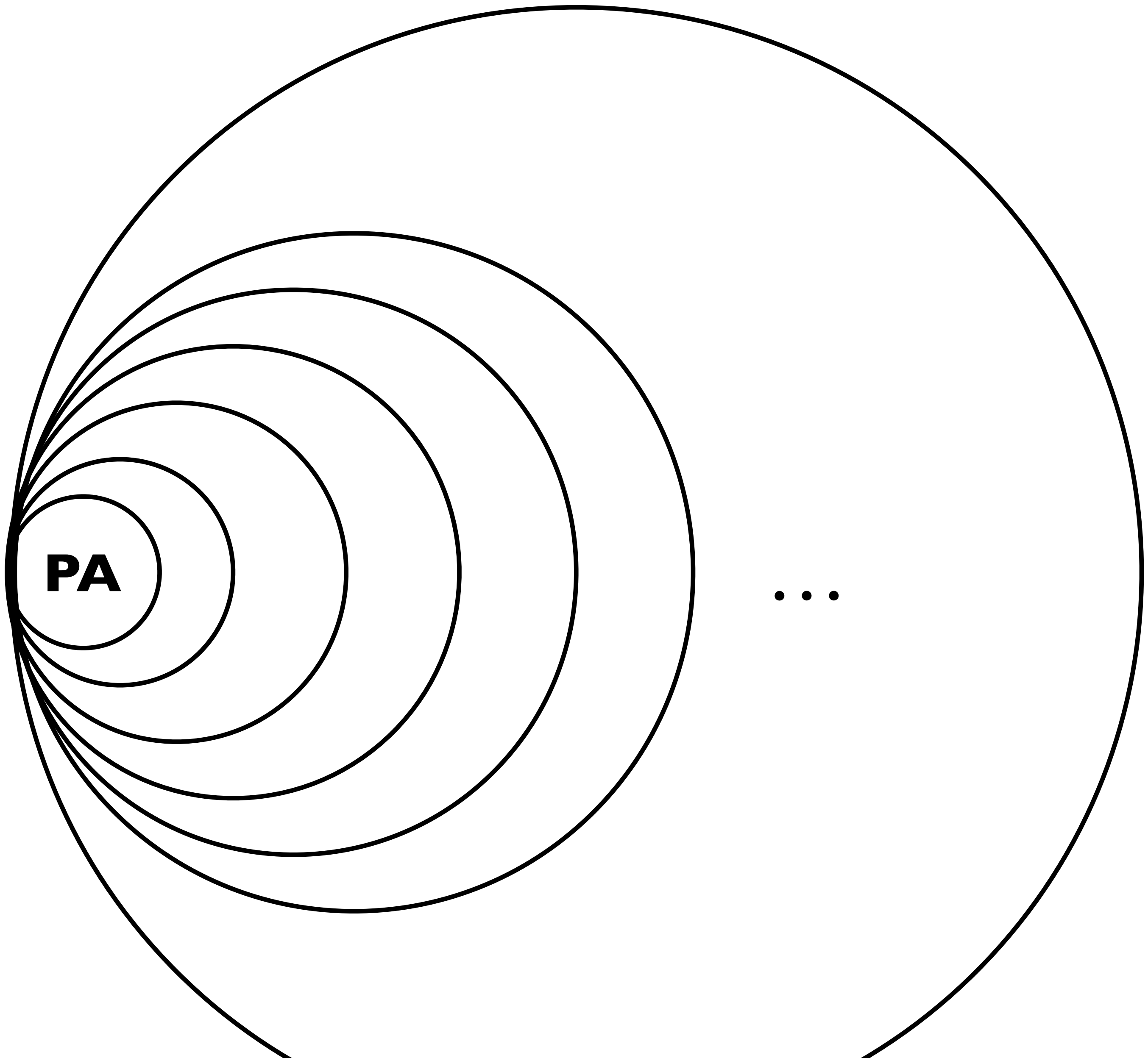
Is there buried inconsistency in here?!?

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



Is there buried inconsistency in here?!?

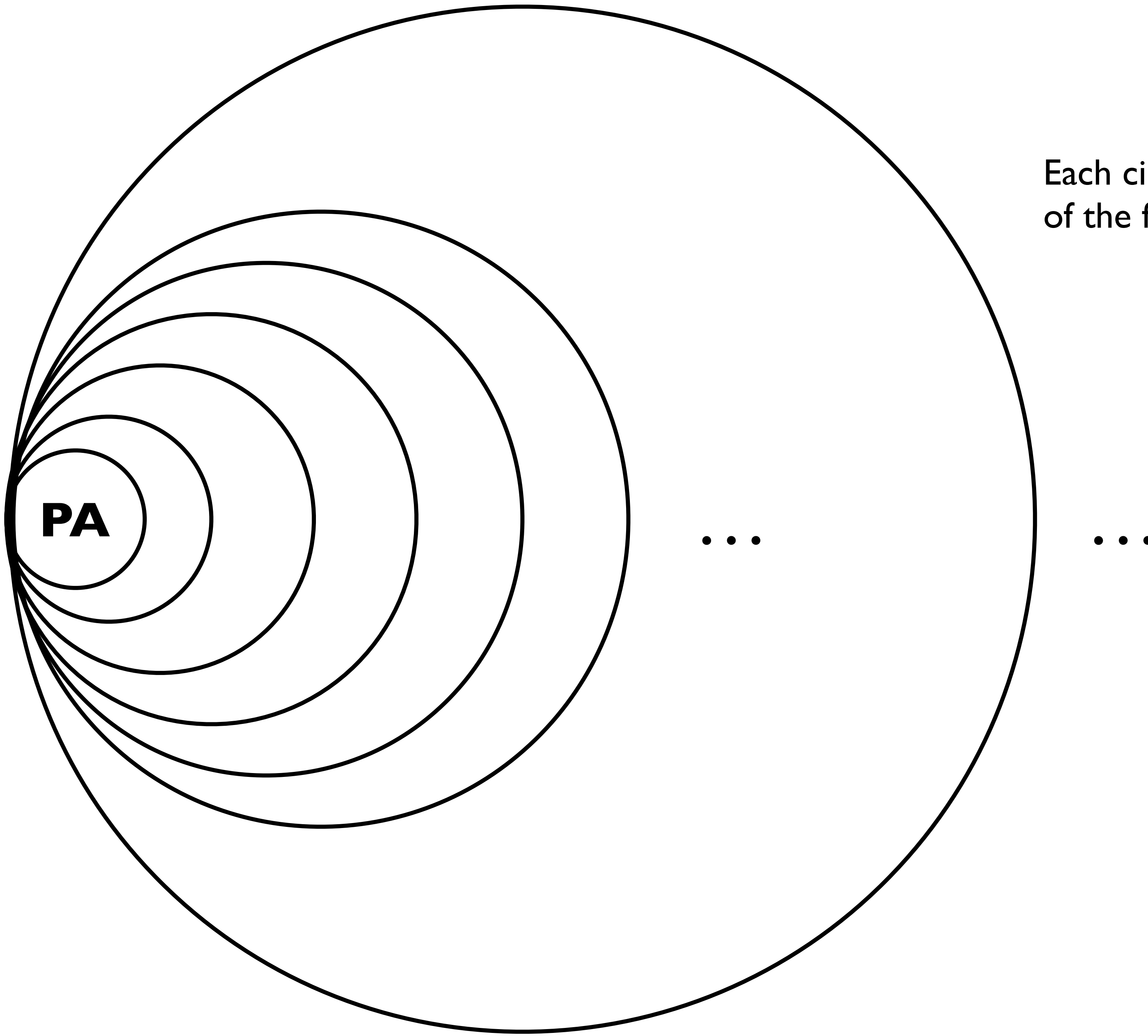
Courtesy of Gödel: Given certain limitative assumptions about "proof power," we can't prove that there isn't!



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.’’

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."

"Don't take Selmer's $ away because he's using full elementary number theory in his AI system because you're worried about inconsistency! — he has no alternative!"

$\Phi \supseteq$ PA?????? —Selmer check this.

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

pocket theorem

We cannot prove from Φ, a set of formulae about basic arithmetic (possibly augmented with other math based on elementary number theory), that no contradiction can be deduced from Φ.

$\Phi \supseteq$ PA?????? —Selmer check this.

Gödel's
Second Incompleteness Theorem

pocket theorem

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

(i) Con $\mathbf{\Phi}$;
(ii) membership in $\mathbf{\Phi}$ is algorithmically decidable; and
(iii) Cap $\mathbf{\Phi}$.

Then $\mathbf{\Phi} \nvdash \neg\pi(\hat{n}^{0=1})$.

We cannot prove from Φ, a set of formulae about basic arithmetic (possibly augmented with other math based on elementary number theory), that no contradiction can be deduced from Φ.

$\Phi \supseteq$ PA?????? —Selmer check this.

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

pocket theorem

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

(i) Con Φ;
(ii) membership in Φ is algorithmically decidable; and
(iii) Cap Φ.

Then $\Phi \nvdash \neg\pi(\hat{n}^{0=1})$.

We cannot prove from Φ, a set of formulae about basic arithmetic (possibly augmented with other math based on elementary number theory), that no contradiction can be deduced from Φ.

$\Phi \supseteq$ PA?????? —Selmer check this.

Gödel's
Second Incompleteness Theorem

pocket theorem

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

(i) Con Φ;
(ii) membership in Φ is algorithmically decidable; and
(iii) Cap Φ.

Then $\Phi \nvdash \neg\pi(\hat{n}^{0=1})$.

We cannot prove from Φ, a set of formulae about basic arithmetic (possibly augmented with other math based on elementary number theory), that no contradiction can be deduced from Φ.

To prove G2, we shall once again allow ourselves …

The Fixed Point Theorem (FPT)

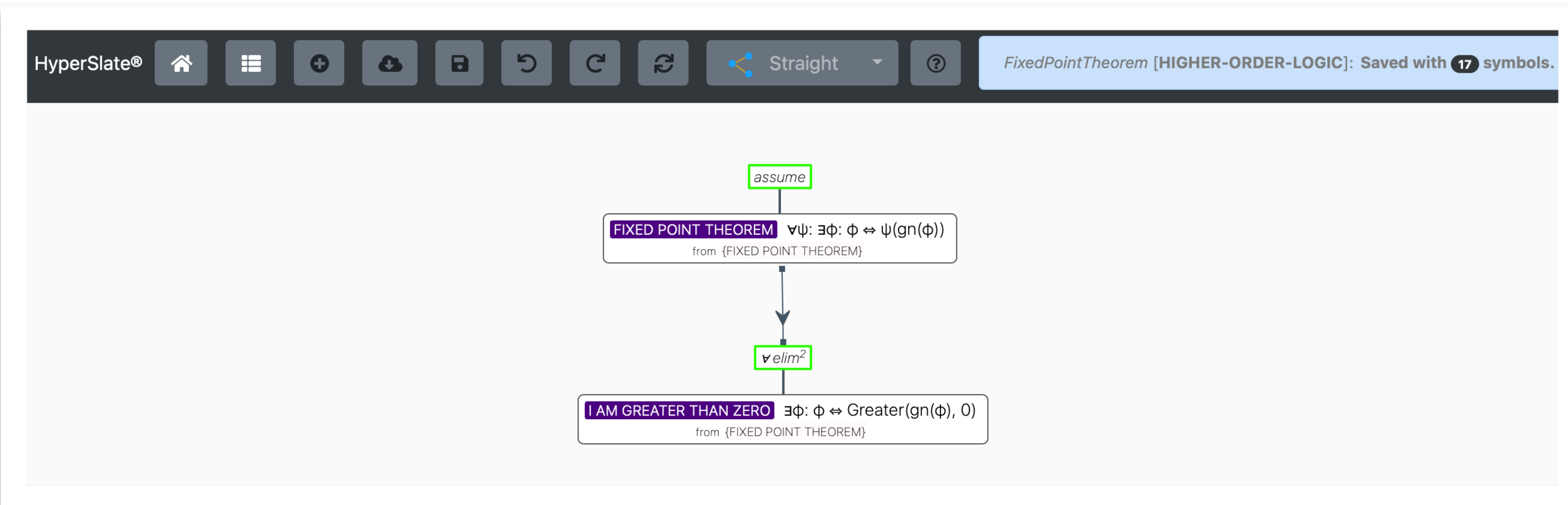
"The Self-Ascription Theorem" (GSAT)

Assume that Φ is a set of arithmetic sentences such that **Cap** Φ. Then for every arithmetic formula $\psi(x)$ with one free variable x, there is an arithmetic sentence ϕ s.t.

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

We can intuitively understand ϕ to be saying:
"I have the property ascribed to me by the formula ψ."

FPT/SATL 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 **PA** system is consistent.)

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

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

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

"I'm not provable from Φ!"

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

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

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

Next, since the provable meta-logical conditional that

if Φ can prove $\mathcal{G}$, then Φ can prove that 0=1

is capturable, and by hypothesis the capturability of Φ is in force, we have:

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

From our supposition and (G2.3) we can deduce by *modus tollens* that Φ proves $\neg\pi(\hat{n}^{\mathcal{G}})$, and from this and (G2.1) by biconditional elimination we have that $\Phi \vdash \mathcal{G}$, i.e., that Φ proves the Gödel sentence, which is impossible, as (G2.2) tells us — contradiction! **QED**

Med nok penger, kan logikk løse alle problemer.