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

9/23/2021

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

9/23/2021

RAIR
Rensselaer AI and Reasoning Lab

OXFORD
UNIVERSITY PRESS

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

9/23/2021

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

9/23/2021

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

9/23/2021

Note: This is a version designed for those who have had at least one university-level course in formal logic with coverage through $\mathscr{L}_1$.

R A I R
Rensselaer AI and Reasoning Lab

OXFORD UNIVERSITY PRESS

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: $\bar{P} \to \bar{P}$ 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

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

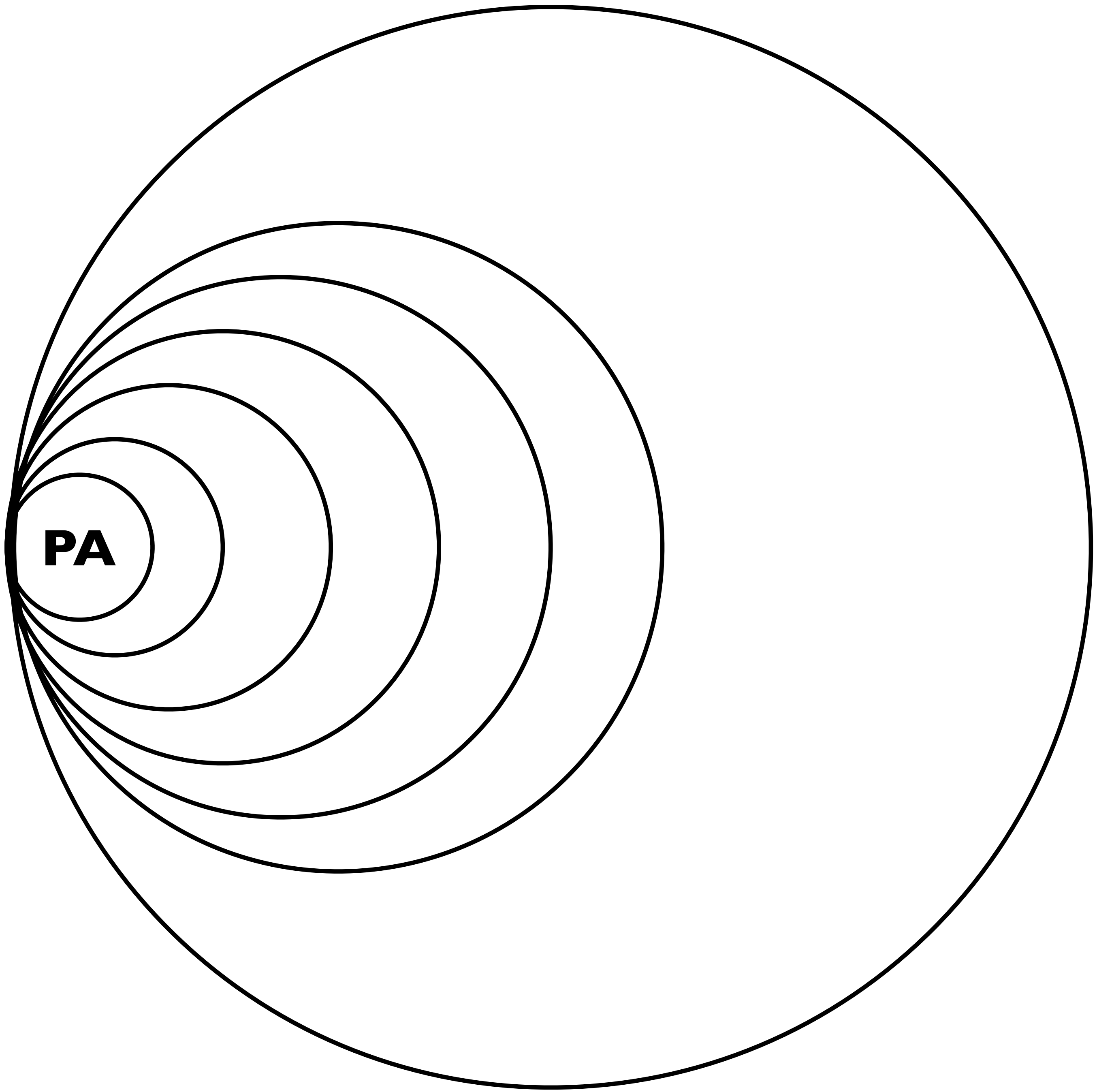
where $\phi(x)$ is open wff with variable x, and perhaps others, free.

Is there buried inconsistency in here?!?

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

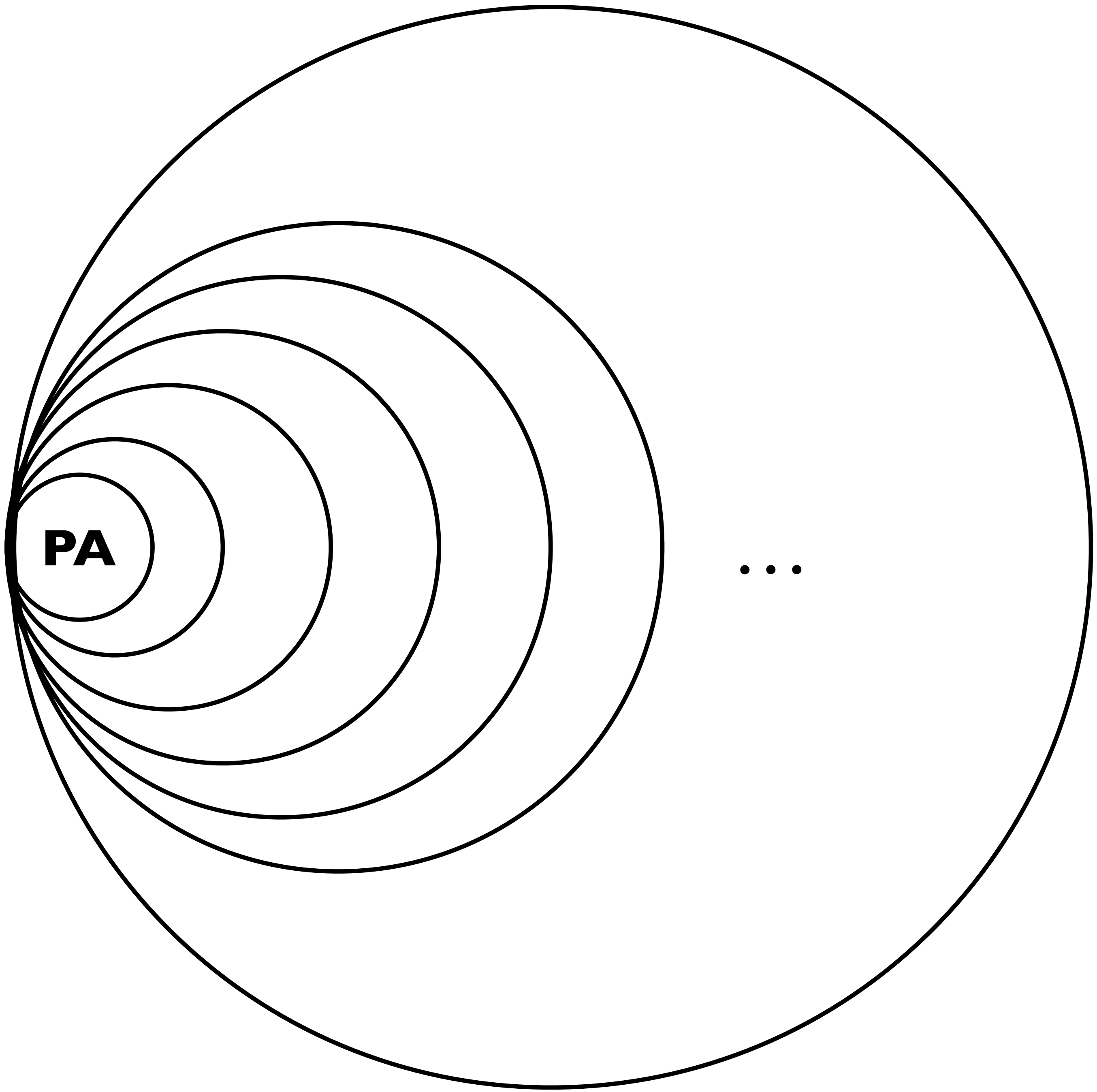
Is there buried inconsistency in here?!?

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



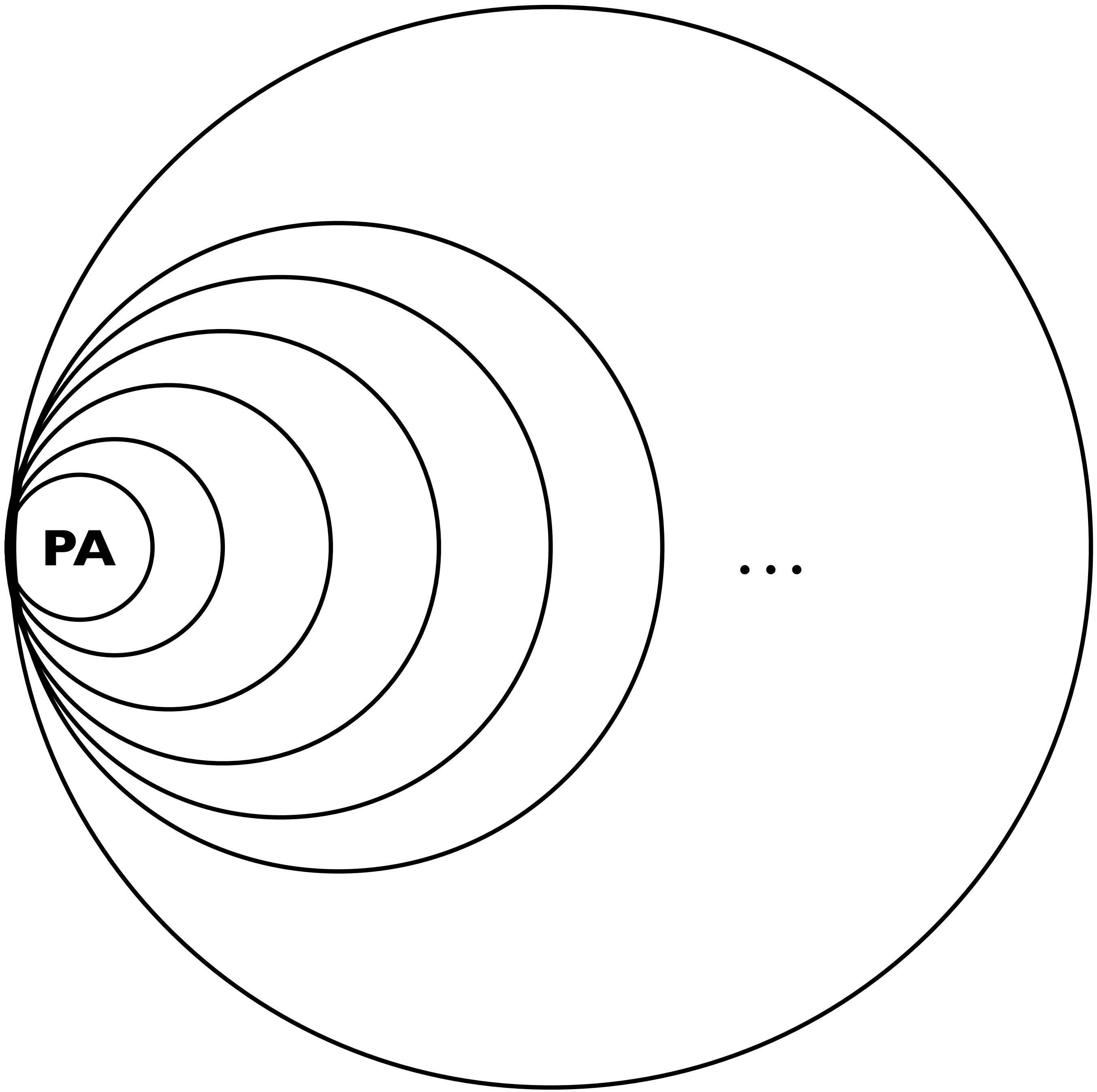
Is there buried inconsistency in here?!?

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



Is there buried inconsistency in here?!?

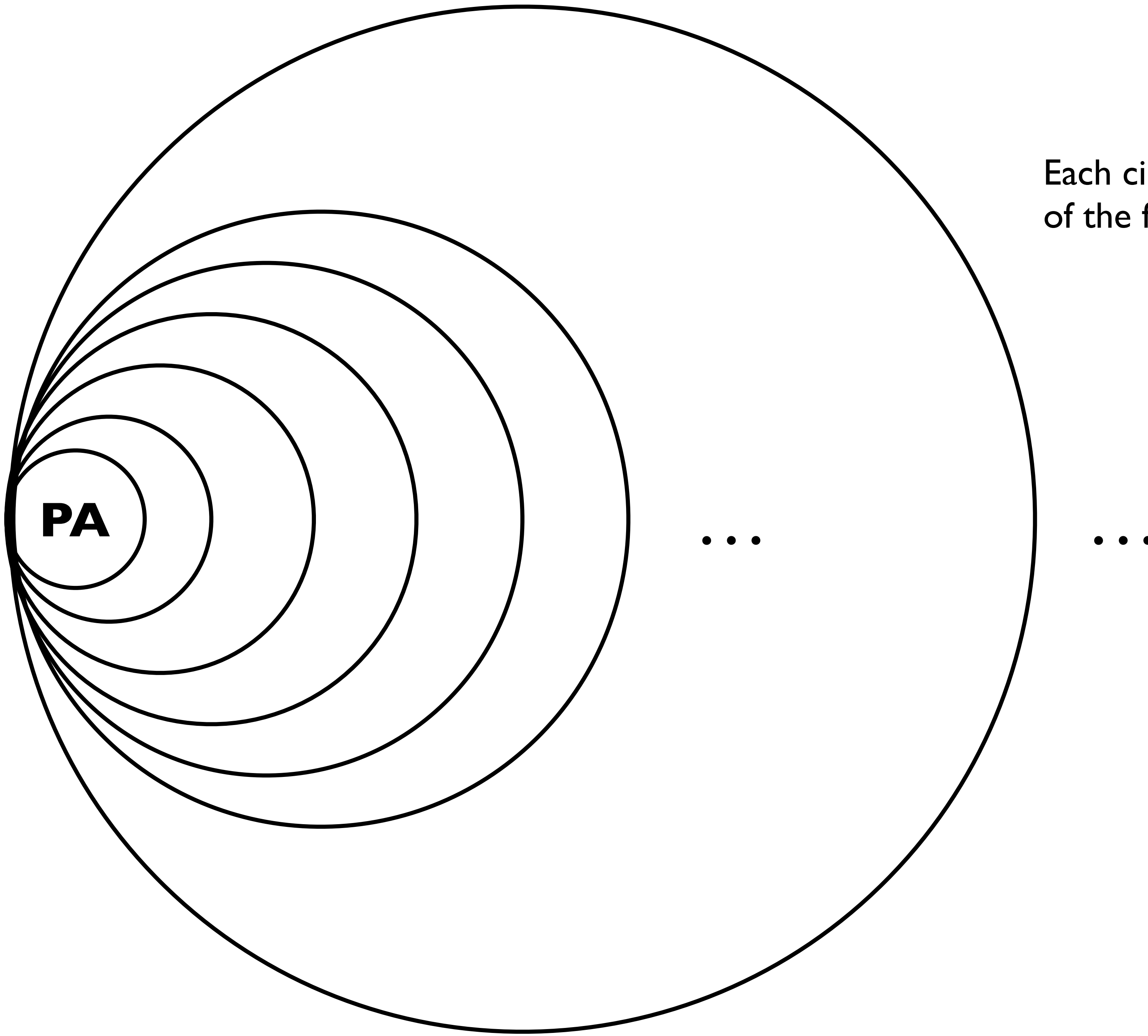
Courtesy of Gödel: Given certain limitative assumptions about "proof power," we can't prove that there isn't!



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 <u>Second</u> Incompleteness Theorem

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

(i) Con $\mathbf{\Phi}$;
(ii) 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 $\mathbf{\Phi} \supset \mathbf{PA}$ that is

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

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

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

Remember Church's Theorem!

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

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

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

To prove G2, we shall once again allow ourselves …

The Fixed Point Theorem (FPT)

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

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

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

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. We set the sentence consis_{Φ} to $\neg\text{Prov}_{\Phi}(n^{0=1})$. Suppose for *reductio* that $\Phi \vdash \text{consis}_{\Phi}$. Since we know that if Φ is consistent no such formula as expresses "I'm not provable from Φ," i.e. formula $\bar{\pi}$, can be provable from Φ (Why?), we have: $\Phi \nvdash \bar{\pi}$. In addition, by Repr, we have this object-level conditional:

(1) $\text{consis}_{\Phi} \rightarrow \neg\text{Prov}_{\Phi}(n^{\bar{\pi}})$,

and in fact the proof of this can be done in Φ itself, so we have:

(**) = (2) $\Phi \vdash \text{consis}_{\Phi} \rightarrow \neg\text{Prov}_{\Phi}(n^{\bar{\pi}})$.

But now from our assumption for indirect proof and (2) we can deduce that $\Phi \vdash \neg\text{Prov}_{\Phi}(n^{\bar{\pi}})$. An instantiation of the Fixed Point Theorem yields:

(FPT*) = (3) $\Phi \vdash \bar{\pi} \leftrightarrow \neg\text{Prov}_{\Phi}(n^{\bar{\pi}})$

and right to left on this biconditional entails that we can prove from Φ that $\bar{\pi}$ — contradiction! **QED**

Med nok penger, kan logikk løse alle problemer.