University of Bern
Consistency and Provability Workshop

Two Models of Consistency - A New Foundational Landscape

Sergei Artemov

CUNY Graduate Center

April 27, 2026

The Unprovability of Consistency Thesis

Hilbert's consistency program was largely sidelined in the 1930s after Gödel's Second Incompleteness Theorem (G2) was interpreted as a definitive block – the Unprovability of Consistency Thesis (UCT):

> *"There exists no consistency proof of a system that can be formalized in the system itself"* (Encyclopædia Britannica).

In fact, Gödel showed the "consistency formula" unprovable. It was von Neumann who promoted the view that consistency itself was unprovable.

We show that Hilbert's concept of consistency was lost in translation between mathematics and syntax, and offer a proof of PA-consistency formalizable in PA. **This renders UCT false in its naive form**.

We do not alter the well-established G2-based architecture of the universe of logical theories, but rather fill in the void left by UCT-orthodoxy. We add a new research avenue with applications in AI and ATP: the **proof-theoretical foundation for self-verification of consistency**.

A precedent: QM "impossibility proof" and frozen research

> *According to what has become a standard history of quantum mechanics, in 1932 von Neumann persuaded the physics community that hidden variables are impossible ... This state of affairs lasted ... until Bell in 1966 exposed von Neumann's proof as obviously wrong. The realization that von Neumann's proof was fallacious then rehabilitated hidden variables and made serious foundational research possible again*[1]*.*

The von Neumann justification of the Unprovability of Consistency Thesis shares a common logical structure with his quantum impossibility proof: both were mathematically sound arguments based on fallacious assumptions. Just as Bell rehabilitated hidden variables, we are rehabilitating consistency proofs. We identify and correct the assumptions, prove what was previously declared impossible, and, like Bell, reopen the original research programs.

---

[1] Dennis Dieks, "Von Neumann's impossibility proof: Mathematics in the service of rhetorics." Studies in History and Philosophy of Modern Physics (Elsevier) 60:136–148, (2017).

Peano Arithmetic PA

Peano Arithmetic PA is a formal first-order theory containing constant 0, the successor function $'$, and all primitive recursive functions with their defining identities.

In addition, PA has the Induction schema which is an infinite series of formulas

$$[\varphi(0) \wedge \forall x\, (\varphi(x) \to \varphi(x'))] \to \forall x\, \varphi(x)$$

for each formula $\varphi(x)$. The Induction schema is our primary example of a **serial property**: an infinite, primitive recursive sequence of formulas.

PA represents all conventional computations, does not use higher order or set-theoretic principles.

Models of arithmetic

The standard model of PA is the set

$$\begin{array}{ccccccc} \bullet & \longrightarrow & \bullet & \longrightarrow & \bullet & \longrightarrow & \cdots \\ 0 & & 1 & & 2 & & \end{array}$$

with standard operations of addition and multiplication. It is easy to show that there are (countable) models of PA not isomorphic to the standard one. Take a fresh constant c and consider a theory

$$\widetilde{\mathrm{PA}} = \mathrm{PA} \ \cup \{c > 0,\ c > 1,\ c > 2,\ \ldots\}.$$

$\widetilde{\mathrm{PA}}$ is consistent since each of its finite subsystems is (obviously) consistent, hence $\widetilde{\mathrm{PA}}$ has a model $\mathcal{M}$, which is a model for PA with "infinite" numbers, non-isomorphic to the standard model:

$$\begin{array}{ccccccccccccc} \bullet & \longrightarrow & \bullet & \longrightarrow & \bullet & \longrightarrow & \cdots & \quad & \cdots & \longrightarrow & \bullet & \longrightarrow & \cdots \\ 0 & & 1 & & 2 & & & & & & c & & \end{array} \ .$$

Numerals vs. natural numbers

For the purposes of Hilbert's consistency program, standard natural numbers are represented constructively as PA-numerals[2]

$$0, 0', 0'', 0''', \ \ldots \ .$$

The standard numbers as a set are not definable in PA. Gödel numbering codes syntactic objects – formulas, finite sequences, *etc.* – by numerals.

> *The principal difference between the informal arithmetic and the formal arithmetic* PA *is in quantification. Since standard natural numbers cannot be defined in* PA*, "for all natural numbers n ..." is replaced in* PA *by the formal quantifier "*$\forall x \ldots$*" which refers to all elements of a model, possibly non-standard: this makes* $\forall x\, \varphi$ *stronger than "for all natural numbers n,* $\varphi(n)$*."* **G2 lives in this gap!**

[2] cf. Hilbert strokes, Zermelo, von Neumann, Church numerals, *etc.*

On Hilbert's notion of consistency

For Hilbert, the domain of number theory are numerals. A finitary general proposition is "a hypothetical judgment that comes to assert something when **a numeral** is given."[3]

The statement of consistency,

no finite sequence S of formulas is a derivation of a contradiction,

after a straightforward Gödel coding of finite sequences by numerals, becomes a Hilbertian "finitary general proposition":

any numeral n is not a code of a derivation of a contradiction.

So, the Hilbertian consistency definition concerns exclusively finite derivations represented by their **numeral** codes.

[3] D. Hilbert, "Die Grundlagen der Mathematik," 1928.

What is a formalization

i) **Direct formalization**. Let H be an informal arithmetical property. The formalization procedure f does the following: the primitive recursive operations, logical connectives are formalized as is, and *"for any natural number n"* is formalized as "$\forall x$." If f succeeds, it converts H into a PA-object $f(H)$ which is a PA-formula, or a set of PA-formulas.

ii) **Gödelian formalization** is a standard arithmetization (syntactic objects are assigned numerical codes, operations on objects become functions on codes) followed by the direct formalization (i) in PA. Formalization of reasoning about math objects is a standard step-by-step conversion of informal reasoning into formal derivations in PA about Gödelian codes.

The Gödelian formalization allows for claiming provability of non-arithmetical combinatorial properties, e.g., "G2 is provable in PA," "Positive Introspection, Löb Theorem are provable in PA," etc.

All formalizations in our proof are standard Gödelian!

For what purposes do we formalize proofs?

Suppose we are interested in

$$\textit{whether a contentual property } H \textit{ of natural numbers holds.} \quad (1)$$

Suppose also that its formalized version $f(H)$ is a PA-formula.

Verification (assumes soundness).

A. If PA proves $f(H)$, then $f(H)$ holds in all models of PA, including the standard model, which gives an affirmative answer to (1).

B. If PA proves the negation of $f(H)$, then $f(H)$ fails in each model of PA including the standard model, which gives a negative answer to (1).

C. However, if neither **A** nor **B** holds, then we have no answer to (1) and must rely on other tools.

Assumptions Checking (syntactic).

If a contentual proof is formalized in PA, then this proof does not use the principles outside PA.

Avoiding the “vacuous consistency” trap

What is the point of establishing the provability of consistency of a theory T in T itself, since if T is inconsistent, the answer would be vacuously affirmative?

This is a wrong question. We not just claim “provability of consistency,” but offer **an explicit proof of consistency**, which makes a difference.

The Hilbertian reply: “a proof of T-consistency in T” is in fact a pair:
i) a mathematical proof of T-consistency (fails for inconsistent Ts);
ii) a formalization of (i) in T for the assumptions checking.

Our addition: An explicit checkable proof $\mathcal{C}$ of consistency constitutes a foundational progress crucial for verification applications.

> *Before, the trust in T's consistency included a trust in the infinite set of T-derivations. With the consistency proof $\mathcal{C}$,* **such trust reduces to checking a single** $\mathcal{C}$*, which can be done manually, or using an independent, trusted proof checker.*

Formalizing consistency in PA: arithmetization

The mathematical formulation of PA**-consistency**,

$$\textit{no } D \textit{ is a } \mathrm{PA}\textit{-derivation of } (0=1), \qquad (2)$$

uses an informal universal quantifier over finite sequences D of formulas, not in the language of arithmetic, hence, for its arithmetical analysis, Gödel coding is required.

Notations: $x{:}y$ is the **proof predicate** "*x is a* PA*-derivation of y*." In a more common notation it is $\mathrm{Prf}_{\mathrm{PA}}(x,y)$.

$\mathit{Consistent}(x)$ is $\neg x{:}(0=1)$.

$\mathit{Consistent}(n)$ states that

"*n is not a code of a* PA*-derivation of* $(0=1)$."

Arithmetized PA**-consistency** is the contentual property

$$\textit{for any natural number } n,\ \mathit{Consistent}(n), \qquad (3)$$

still not in the language of PA since quantifiers over standard numbers are not expressible in PA.

The consistency formula $\mathrm{Con}_{\mathrm{PA}}$

Traditional route: turn a blind eye to the domain violation and assume that the mathematical statement of consistency (3) is represented in PA by a formula $\mathrm{Con}_{\mathrm{PA}}$:

$$\mathrm{Con}_{\mathrm{PA}} = \forall x\, \mathit{Consistent}(x) = \forall x\, [\neg x{:}(0\!=\!1)]. \qquad (4)$$

> *Treating* $\mathrm{Con}_{\mathrm{PA}}$ *as* **the definition of** PA-**consistency** *in the context of provability in* PA *has been a mistake.* $\mathrm{Con}_{\mathrm{PA}}$ *is not equivalent in* PA *to the original definition* (2) *and* (3) *which account for standard numerical codes only.*

Gödel's Second Incompleteness Theorem.
If PA *is consistent, then* PA *does not prove* $\mathrm{Con}_{\mathrm{PA}}$*.*

Corollary. *There is a model* $\mathcal{M}$ *of* PA *in which* $\mathrm{Con}_{\mathrm{PA}}$ *is false.*

Gödel's "monster." There is a "proof" of $(0\!=\!1)$ in $\mathcal{M}$. However, such a "proof" cannot be a numeral $n = 0, 1, 2, \ldots$, i.e., it is non-standard in $\mathcal{M}$.

Consequences of the domain violation in $\mathsf{Con}_{\mathsf{PA}}$

As we saw, there are models of PA with inconsistent proofs. However, all such "bad" proofs turned out to be nonstandard, hence G2 does not appear to be about real PA-derivations which are all finite and which the original consistency question has been all about.

> $\mathsf{Con}_{\mathsf{PA}}$ *is infested with junk proof codes, which* PA *is too weak to sort out. This distorts the intrinsic nature of consistency.*

It is easy to show mathematically (done in 2025) that

$\mathsf{Con}_{\mathsf{PA}}$ **is strictly stronger in** PA **than the consistency property** (3).[4]

Hence **deriving the Unprovability of Consistency Thesis from the unprovability of** $\mathsf{Con}_{\mathsf{PA}}$ **in** PA **has been a "strengthening fallacy."**

So, we are back to square one with respect to the Hilbertian question of whether the consistency of PA can be established within PA. Using $\mathsf{Con}_{\mathsf{PA}}$ here fails, so we have to seek another route.

[4] Sergei Artemov. "Consistency formula is strictly stronger in PA than PA-consistency," arXiv:2508.20346, 2025.

By definition, the consistency is a serial property

A **serial property is a primitive recursive series of formulas** $F_1, F_2, F_3, \ldots$.

> *Serial properties (schemas) are common objects in mathematics and logic: tautologies in* PA, *Induction schema in* PA, PA *itself is a schema, Reflection schemas, Comprehension/Separation schemas in* ZF, *etc.*

By definition, the PA-consistency (3) is a serial property $\mathsf{ConS}_{\mathsf{PA}}$:

$$Consistent(n), \ \ n = 0, 1, 2, \ldots \ , \tag{5}$$

which is strictly weaker in PA than $\mathsf{Con}_{\mathsf{PA}}$.

The semantic features, such as truth in a model, naturally extend from formulas to sets/series of formulas. Now we need to formalize the methods that mathematicians use **to prove serial properties**.

Selector proofs

The proofs of serial properties are **selector proofs** that can be traced back to Hilbert, Brouwer, Heyting, Kolmogorov, Kreisel, and have been used in everyday mathematics.

A proof of $\mathcal{F} = \{F_0, F_1, \ldots, F_n, \ldots\}$ in a theory T is a pair of

(i) *selector* $s(x)$: an operation that given n provides a proof of F_n in T;

(ii) *verifier* v: a proof in T that the selector does (i).

We call such pairs (i) and (ii) *selector proofs*. In our notations,

$$v : \forall x\, [s(x){:}F(x)].$$

In this work, selectors are explicit primitive recursive operations but this can be naturally extended to other provably total functions.

Direct precursors of selector proofs

Hilbert ϵ-substitution is basically a selector proof method.

In fact, Hilbert provides rather explicit guidelines for consistency proofs. Here is a quote from Richard Zach's exposition of Hilbert program[5]:

> *"What is required for a consistency proof is an operation which, given a formal derivation, transforms such a derivation into one of a special form, plus proofs that the operation in fact succeeds in every case and that proofs of the special kind cannot be proofs of an inconsistency."*

Selector proofs are natural generalization of ϵ-substitution when we allow using arbitrary primitive recursive selectors rather than confining them to specific substitution chains.

BHK semantic for $\forall$:

a proof of $\forall x\, A(x)$ *is a function converting* c *into a proof of* $A(c)$,

introduces selector explicitly, and Kreisel work adds the verification step.

[5] R. Zach. Hilbert's program then and now. In *Philosophy of Logic*, pp. 411–447, North-Holland, 2007. Also the SEP article "Hilbert's Program."

Selector Proofs: formalization

A proof of a serial property $\mathcal{F} = \{F_0, F_1, \ldots, F_n, \ldots\}$ *in a theory* T is a pair of

(i) *selector* $s(x)$: an operation that given n provides a proof of F_n in T;

(ii) *verifier* v: a proof in T that the selector does (i),

$$v : \forall x\,[s(x){:}F(x)].$$

Note that the condition (ii) yields a Gödelian-style formalization of the verifier as a conventional finite derivation v in T:

$$T \vdash \forall x\,[s(x){:}F(x)].$$

So, **each selector proof in** PA **reduces to a finite derivation in** PA.

Conservativity of selector proofs in PA

Proposition. *If a serial property* $\mathcal{F} = \{F_0, F_1, \ldots, F_n, \ldots\}$ *is selector provable in* PA*, then each of the* F_n*'s is conventionally provable.*

Proof. We use the *explicit reflection principle*

$$t{:}F \rightarrow F. \tag{6}$$

> *Gödel suggested that the explicit reflection is* PA*-provable at a public lecture in 1938, but this work remained unpublished until 1995. During that gap, the provability of explicit reflection was independently recognized by Artemov and T. Strassen in 1992.*

Suppose

$$\mathsf{PA} \vdash \forall x\,[s(x){:}F(x)],$$

then for each n,

$$\mathsf{PA} \vdash s(n){:}F(n),$$

and, by explicit reflection,

$$\mathsf{PA} \vdash F(n).$$

Conservativity of selector proofs in PA

This property is fundamental: **selector proofs are finite** PA-**derivations that represent proofs of (possibly infinite) groups of** PA-**theorems**.

In other words, selector proofs are just an overlooked by formal logic mathematical way to prove theorems of PA in groups.

$$\mathsf{PA} + \textit{selector proofs} = \mathsf{PA}^{6}.$$

[6] After Lois learns that Clark is Superman, Clark/Superman still remains the same since adding knowledge doesn't change the underlying entity.

Example 1

Remarkably, selector proofs have been tacitly adopted in mathematics.

Complete Induction principle, $\mathcal{CI}$: *for any formula* ψ*,*

if for all x $[\forall y < x\ \psi(y)$ *implies* $\psi(x)]$, *then* $\forall x\, \psi(x)$.

Complete Induction for PA is provable by means of PA. Here is a **textbook proof of** $\mathcal{CI}$:

> *Apply the usual* PA*-induction to* $\forall y < x\ \psi(y)$ *to get the* $\mathcal{CI}$ *statement* $\mathcal{CI}(\psi)$ *for* ψ*.*

This is a selector proof which, given ψ selects a derivation of $\mathcal{CI}(\psi)$ in a way that provably works for any input ψ.

Example 2

Multiplying polynomials. *The product of polynomials is a polynomial.*[7]

Here is its standard mathematical proof.

> *Given a pair of polynomials f, g, using the well-known formula, calculate coefficients of the product polynomial $p_{f \cdot g}$, and* **prove in arithmetic** *that*
>
> $$f \cdot g = p_{f \cdot g}. \tag{7}$$

This is a selector proof: for each f, g, it finds a proof of (7) in PA.

[7] A polynomial is a term $a_n x^n + a_{n-1} x^{n-1} + \ldots + a_1 x + a_0$, where each a_i is a numeral and x a variable. In $f \cdot g$, "$\cdot$" stands for the usual PA multiplication.

Example 3

Any tautology containing propositional variables, e.g., the double negation law, DNL in arithmetic: *for any formula X,*

$$X \leftrightarrow \neg\neg X. \tag{8}$$

The standard proof of DNL in PA is

For a given X, build the usual derivation $D(X)$ of (8) *in* PA.

This is a selector proof which builds an individual PA-derivation for each instance of DNL in a way that provably works for any input X. This proof can be easily formalized in PA.

This example suggests that selector proofs are ubiquitous in meta-mathematics Gödel-formalizable in PA.

Example 4: a proof of self-consistency

Let Con_T be the standard Gödelian consistency formula for a theory $T \supseteq \mathrm{PA}$. Consider theories:

$$\mathrm{PA}_0 = \mathrm{PA}, \qquad \mathrm{PA}_{i+1} = \mathrm{PA}_i + \mathrm{Con}_{\mathrm{PA}_i}, \qquad \mathrm{PA}^{\omega} = \bigcup \mathrm{PA}_i.$$

Consider a folklore consistency proof

> *Let D be a derivation in PA^{ω}. Find n such that D is a derivation in PA_n. $\mathrm{Con}_{\mathrm{PA}_n}$ – one of the postulates of PA^{ω} – implies that D does not end with $(0=1)$.*

This is a mathematical proof of consistency of PA^{ω} in PA^{ω}.

On the other hand, by G2, PA^{ω} cannot prove $\mathrm{Con}_{\mathrm{PA}^{\omega}}$, but this fact does not harm the given consistency proof, which does not derive $\mathrm{Con}_{\mathrm{PA}^{\omega}}$.

The main result

We present a selector proof of PA-consistency in its serial form. Together with the explicit reflection in PA, this offers a well-principled mathematical proof of

$$\textit{for all } n,\ Consistent(n)$$

using only axioms of PA. This proof is Gödel-formalizable in PA.

Note that, by G2,

$$\mathsf{PA} \nvdash \forall x\, Consistent(x).$$

A peer-reviewed journal publication:

Sergei Artemov. "Serial properties, selector proofs and the provability of consistency," Journal of Logic and Computation (Oxford), Volume 35, Issue 3, exae034, 2025.

The sketch of the consistency proof

A natural approach to establish that no PA-derivation D proves $(0 = 1)$ would be to argue that all formulas in D are true, whereas $(0 = 1)$ is false, hence $(0 = 1)$ does not occur in D. This proof is not formalizable in PA since the notion of "true" is not represented by an arithmetical formula.

The Proof Theory of the 1950s built **partial truth predicates** which are definable in PA for formulas of bounded complexity. We do not need a universal one-size-fits-all "truth formula"; it suffices to provide a truth definition that works for all formulas in a given derivation D. It took over 70 years to collect the infinite family of such proofs of D-consistency into one universal arithmetical sentence, a selector proof in PA.

> *This proof can be viewed as the completion of the Mostowski Reflexivity Theorem, MRT, that* PA *proves the consistency of its finite fragments. MRT de facto builds an explicit selector for proving consistency, but relaxes it to an implicit provability claim "the consistency is provable." This falls short of claiming "consistency" since the implicit reflection is not provable.*

Compact and Non-Compact Proofs in Mathematics

The usual way to prove

$$\textit{for all } n,\ F(n) \tag{9}$$

in mathematics is: given an arbitrary n, provide an argument $\mathcal{A}(n)$ concluding $F(n)$. Imagine that $\mathcal{A}(n)$ **requires more and more new axioms with the growth of** n. In mathematics, this will still be a legitimate proof of (9).

> *Such "non-compactness" appears, e.g., in MRT. The proof of MRT needs an unlimited access of $\mathcal{A}(n)$'s to the* PA *induction.*[8]

To study this phenomenon, we need to make individual reasonings $\mathcal{A}(n)$ explicit and consider proofs (9) in the enhanced format

$$\textit{for all } n,\ A(n){:}F(n).$$

In other words, we have to shape a proof of (9) as a selector proof.

---

[8] Another example: a hypothetical twin-primes proof.

Compact and Non-Compact Proofs - formally.

Definition *A selector proof in* PA

$$\mathrm{PA} \vdash \forall x\, [s(x){:}F(x)] \qquad (10)$$

is **compact**, *if all $s(n)$'s are derivations in some finitely axiomatized subsystem W of* PA*:*

$$\mathrm{PA} \vdash \forall x\, [s(x){:}_W F(x)].^9 \qquad (11)$$

A selector proof (10) *is* **non-compact** *otherwise.*

A simple but fundamental observation: *Since* PA *proves the uniform reflexivity of its finite fragments, compact provability* (11) *reduces to conventional provability*

$$\mathrm{PA} \vdash \forall x\, F(x).$$

This answers the question: what consistency proofs are banned by G2?

> *G2 bans compact proofs of* PA*-consistency in* PA*. The possibility of non-compact proofs of consistency was left open.*

Naturally, **our proof of** PA-**consistency is non-compact**, as well as proofs from Examples 1 and 4.

---

[9] Here "$:_W$" is the Gödelian proof predicate in W.

New reading of G2

These findings suggest new foundational reading of Gödel's Second Incompleteness Theorem concerning the standard notion of consistency "no PA-derivation ends with $(0=1)$":

> *The consistency of* PA *is not provable within a finite fragment of* PA.

This is complemented with the positive message:

> *The consistency of* PA *is provable within the whole* PA. *By using selector proofs,* PA *formally certifies the consistency of every actual derivation it can ever perform.*

G2 does not establish the absolute unprovability of consistency; it only demonstrates that consistency is a non-compact property that cannot be compressed into a single Gödelian formula.

How far we can go with proving consistency in PA

Kurahashi and Sinclaire, 2019:

PA *cannot selector-prove consistency of any theory*

$$T \supseteq \mathsf{PA} + \mathsf{Con}_{\mathsf{PA}}.$$

This observation allows us to conceptually insulate provable Hilbertian consistency from unprovable Gödelian consistency. The former studies the Provability of Consistency, and the latter serves as a pivotal yardstick for measuring the strength of different theories without interfering with the question of the provability of self-consistency.

Freund/Pakhomov and Gadsby, 2024/25:

However, PA *selector-proves consistency of* T *for some proper extensions* T *of* PA *in particular*

$$T = \mathsf{PA} + \textit{slow consistency of } \mathsf{PA}.$$

The failure of UCT in its Encyclopædia Britannica form

The Unprovability of Consistency Thesis in its naive Encyclopædia Britannica form fails and should be corrected since there exists a consistency proof of PA that can be formalized in PA.

This perspective suggests that the Hilbert Program wasn't "killed" by Gödel. It implies that a system can be consistent and self-aware of that consistency.

The paradigm shift

The ghost exorcised. The perceived inability of a system to verify itself was a byproduct of a restrictive formalization, not a logical law.

The serial reality. Consistency is **non-compact**. By returning to Hilbert's serial definition, internal awareness of integrity becomes a formal reality.

A new chapter for Proof Theory. This re-centers Proof Theory as a practical discipline rather than a study of limitations. It removes the perceived "logical barrier" to developing self-verifying AI and automated theorem provers.

A balanced foundation. We no longer have to choose between Gödel and Hilbert. The Gödelian model measures relative strength; the Hilbertian model provides internal trust. They are complementary, not contradictory.

Instead of an epilogue

In the 19th century, Lord Kelvin's estimate of the Earth's age (20–100 million years) was widely accepted as a physical certainty. It appeared to directly contradict Darwin's evolutionary timeline, which required vast spans of geological time. Kelvin's estimate was so authoritative it was even enshrined in the *Encyclopædia Britannica*.

However, Kelvin's physics was incomplete. The discovery of radioactivity overturned his view, confirming an Earth approximately 4.6 billion years old and validating Darwin's vision.

Just as Kelvin's physics lacked the necessary variables, the 20th-century formalization of consistency was incomplete. By expanding our framework today, we align formal logic with mathematical reality.