The Provability of Consistency

Sergei Artemov

CUNY Graduate Center

May 29, 2019

Paths to proving consistency

We consider the question of *proving consistency of Peano arithmetic* PA *by means formalizable in* PA.

Several paths converge at this point:

1. **Historical**, via Hilbert's Program and Gödel's Incompleteness.
2. **Foundational**, whether tools formalizable in a theory T are sufficient for establishing consistency of T.
3. **Mathematical**, whether the arithmetized consistency $\mathrm{Con}(T)$ is a fair representation of mathematical consistency of T.
4. **Constructive**, BHK semantics, Gödel's S4, the Logic of Proofs, and tracking witnesses in arithmetic reasoning.

Hilbert's consistency program

The goal of Hilbert's consistency program was to give "finitary" proofs that there can be no derivation of a contradiction in mathematical theories. For Hilbert, the domain of contentual number theory are **numerals** such as

$$|, ||, |||, ||||, \ldots$$

A finitary general proposition is "**a hypothetical judgment that comes to assert something when a numeral is given**" (Hilbert, 1928)

For Hilbert, the statement of consistency is of such a general form: *for a given sequence of formulas S, S is not a derivation of a contradiction*. Within this talk, we will call this statement **Hilbert consistency**.

This Hilbert's approach hinted at formalizing the consistency property as **an arithmetical scheme with a numeral parameter**.

Disclaimer

Despite this mentioning of Hilbert's consistency program, in this work, we **do not study Hilbert's finitism** (which has not even been definitively described) but rather focus on the class of proofs

by means formalizable in PA**.**

G2 and Formalization Principle

Formal derivations are finite sequences of formulas. Gödel's arithmetization numerically encodes those derivations and then uses numeric quantifiers to represent universal properties of derivations, including the consistency formula Con(T),

$$\forall x \text{ ``}x \textit{ is not a code of a proof of a contradiction in } T\text{.''}$$

By Gödel's Second Incompleteness Theorem, G2, PA, if consistent, does not prove Con(PA).

To connect G2 to the real question of (un)provability of PA-consistency, one has to rely on Formalization Principle, FP,

any finitary reasoning may be formalized as a derivation in PA.

Gödel and Hilbert vs. von Neumann on FP

In the principal G2 paper, "On formally undecidable propositions . . ." of 1931, speaking of G2, Gödel directly challenges FP:

> *... it is conceivable that there exist finitary proofs that cannot be expressed in the formalism of [our basic system].*

Hilbert has rejected FP is strong words.

Von Neumann, however, was an active promoter of FP and of reading arithmetical consistency formulas like Con(PA) as contentual consistency statements, which we call **von Neumann consistency**.

Von Neumann's viewpoint appeared to prevail in the public opinion, *de facto* in the form of the following Strong Formalization Principle.

Strong Formalization Principle, SFP

Any reasoning by means of PA *may be formalized as a derivation in* PA

SFP is more general than FP, since most authors appear to agree that finitary reasoning tools are formalizable in PA. Therefore, Gödel's and Hilbert's reservations concerning FP automatically translate to similar reservations concerning SFP.

SFP is needed to connect G2 with the popular opinion that methods formalizable in PA cannot prove consistency of PA: **without SFP, such a conclusion is not warranted**.

Mathematical consistency of PA vs. Con(PA)

By construction, Con(PA) holds in the standard model of arithmetic iff PA is consistent. However, since we are interested in **provability** of this formula in PA, we have to analyze validity of Con(PA) in **all models** of PA, most of them nonstandard.

In a given nonstandard model, the quantifier "for all x" spills over to nonstandard/infinite numbers, and hence Con(PA) **states consistency of both standard and nonstandard proof codes**. This is stronger than mathematical consistency of PA which speaks exclusively about sequences S of formulas and such sequences have only standard integer codes.

Mathematically, by G2, PA does not prove Con(PA) hence there are models of PA with inconsistent proofs. However, **all such "bad" proofs turned out to be infinite/nonstandard**, hence G2 does not appear to be about real PA-derivations which are all finite and which Hilbert's consistency program has been all about.

Con(PA) is unprovable because of a technicality?

Arithmetization and consequent factoring the informal universal quantifier

"any finite sequence S"

into the language of PA, thereby making it an internalized quantifier,

"any number x,"

appear to distort the foundational picture and make consistency unprovable for a seemingly nonessential reason: the language of PA is too weak to sort out fake codes.

In this respect, a better arithmetical presentation of consistency of PA is offered by **a scheme with a numeral parameter** n, ConS(PA):

"*a* PA-*proof with code n does not contain* $0=1$*,"*

a Hilbertian "hypothetical judgment when a numeral n is given" – rather than as a Π_1-formula Con(PA).

Arithmetical schemes are necessary

We argue that arithmetical schemes not reducible to formulas should be included into proof theoretical considerations. For example, the intuition "**any principle of** PA **is provable by means formalizable in** PA" is not supported by the existing toolkit of arithmetical formalizations.

Consider the **Induction Principle**: for each formula φ,

$$\varphi(0) \wedge \forall x(\varphi(x) \rightarrow \varphi(x+1)) \rightarrow \forall x \varphi(x).$$

There is no single formula *IND* which logically implies all *Ind*(φ)'s and is provable in PA. O/w, *IND* and all *Ind*(φ)'s were derivable in a finite fragment of PA which is impossible (PA is not finitely axiomatizable).

So, the arithmetical representation of Induction Principle is a **scheme**

$$\{Ind(\varphi) \mid \varphi \textit{ is an arithmetical formula}\}.$$

The same holds for Reflection Principle, Explicit Reflection Principle, Σ_1-Completeness, etc.: they are all represented by **schemes** rather than by single formulas and widely used in proof theory.

What counts as a proof of a scheme?

Naïvely, a scheme is provable iff each of its instances is provable. However, this does not automatically extend to provability by means formalizable in PA. Otherwise, any true Π_1-sentence $\forall x S(x)$ would be, counterintuitively, PA-provable as a scheme:

$$\{S(n) \mid n = 0, 1, 2, \ldots\}.$$

However, the generic justification of this is not formalizable in PA.

For the consistency proof, we apply an intuitively safe two-stage approach for proving a scheme $S(n)$ by means formalizable in PA:

i) find a mathematical proof of $S(n)$ as Hilbert's "hypothetical judgment when a numeral n is given";

ii) step-by-step formalize (i) in PA.

Counterexample to SFP with schemes:

It is assumed that each arithmetical formula ψ expresses a contentual property of natural numbers 'ψ.'

Induction Principle $\mathit{Ind}(\varphi)$ is the scheme

$$\varphi(0) \wedge \forall x(\varphi(x) \rightarrow \varphi(x+1)) \rightarrow \forall x \varphi(x).$$

Obviously, $\mathit{Ind}(\varphi)$ is provable by means of PA. Indeed, given φ, assume '$\varphi(0)$' and '$\forall x(\varphi(x) \rightarrow \varphi(x+1))$'. By induction, conclude '$\forall x \varphi(x)$'.

A straightforward formalization of this proof in PA produces an obvious primitive recursive term $p(x)$ such that

$$\mathrm{PA} \vdash \forall x \text{ "}p(x) \textit{ is a proof of } \mathit{Ind}(x)\text{"}.$$

Both conditions (i) and (ii) are met. Therefore, Induction Principle, as a scheme, is provable by means formalizable in PA, but, as it was shown earlier, cannot be proved in PA as a single formula.

How to prove Hilbert consistency of PA by means of PA.

Consider consistency in its original Hilbert form:

"no sequence of formulas S is a derivation of a contradiction."

Our strategy: find a way to reason about real PA-derivations S as combinatorial objects and avoid arithmetization.

Once we have decided to avoid arithmetization, finitary mathematical proofs of Hilbert consistency readily suggest themselves. We are presenting one below.

Partial truth definitions in PA

In metamathematics of the first-order arithmetic, there is a well-known construction called *partial truth definitions*. Namely, for each $n = 0, 1, 2, \dots$ we build, in a primitive recursive way, a Σ_{n+1} formula

$$Tr_n(x, y),$$

called *truth definition for* Σ_n *formulas*, which satisfies natural properties of a truth predicate.

Intuitively, when φ is a Σ_n-formula and y is a sequence encoding values of the parameters in φ then $Tr_n(\ulcorner\varphi\urcorner, y)$ defines the truth value of φ on y.

Partial truth definitions in PA

Proposition 1.

- $Tr_n(\ulcorner\varphi\urcorner, y)$ *satisfies the usual properties of truth with respect to boolean connectives, quantifiers, and rule Modus Ponens for each* $\varphi \in \Sigma_n$*, and these properties are derivable using* Σ_{n+1} *induction.*
- PA *naturally proves Tarksi's condition for any* Σ_n*-formula* φ*:*

$$Tr_n(\ulcorner\varphi\urcorner, y) \equiv \varphi(y).$$

  In particular, $\neg Tr_n(\ulcorner 0=1 \urcorner, y)$ *is naturally provable.*
- $Tr_n(\ulcorner A\urcorner, y)$ *is provable for any axiom* A *of* PA *of depth* $\leq n$*.*

Note that all the proofs in Proposition 1 are rigorous contentual arguments w/o any metamathematical assumptions about PA. The formal language of PA is used here just for bookkeeping.

A proof of Hilbert consistency for PA

Given a finite sequence S of formulas which is a legitimate PA-derivation, we first calculate n such that all formulas from S have depth $\leq n$. Then, by induction on the length of S, we check that for any formula φ in S with parameters y, the property $Tr_n(\ulcorner\varphi\urcorner, y)$ holds. This is an immediate corollary of Proposition 1, since all PA-axioms satisfy Tr_n and each rule of inference respects Tr_n. So, Tr_n serves as an invariant for all formulas from S. Since, by Proposition 1, $0=1$ does not satisfy Tr_n, $0=1$ cannot occur in S.

1. This is a rigorous mathematical proof of Hilbert consistency of PA.
2. The constructions and required properties used in this argument are formalizable in PA: partial truth definitions, compliance of truth definitions with PA-derivation rules, etc.

Comments to this proof of Hilbert consistency

Mathematically, this proof is a "deformalization" (i.e., a contentual counterpart) of the well-known formal derivation of $\mathrm{Con}(I\Sigma_n)$ in $I\Sigma_{n+1}$:

$$I\Sigma_{n+1} \vdash \mathrm{Con}(I\Sigma_n). \quad (1)$$

Note, however, that (1) alone is not sufficient for claiming a consistency proof for PA since its direct application

"consistency is provable hence consistency takes place"

requires a soundness assumption which is not appropriate since such an assumption is stronger than the desired consistency conclusion.

We have to repeat steps of (1) in a contentual reasoning.

A posteriori arithmetization: a consistency scheme

The Hilbert consistency condition

no sequence S of formulas is a derivation of a contradiction in PA

can be equivalently represented by an arithmetical **scheme** ConS(PA):

n is not a code of a proof of a contradiction in PA,

with a **numeral** parameter *n*.

Specifics of formalization for ConS(PA).

Here is a verbal description of a primitive recursive function/term $p(x)$ connecting a parameter n with the proof $p(n)$.

Given n, the Gödel number of a PA-derivation, we first calculate $r(n)$ such that all formulas from S have depth $\leq r(n)$. All quantifiers used in the description of the procedure are now bounded by $r(n)$ or other given primitive recursive functions of n.

Then, for any formula φ in S, we build a proof of $Tr_{r(n)}(\ulcorner\varphi\urcorner, y)$. Since, by Proposition 1, we have a proof of $\neg Tr_{r(n)}(0=1, y)$, we have a proof that $0=1$ is not in S.

By the description, $p(n)$ is primitive recursive and

$$\text{PA} \vdash \forall x \text{“}p(x) \textit{ is a proof that } x \textit{ does not contain } 0=1\text{”}. \qquad (2)$$

(2) does not serve as a proof of ConS(PA), but rather as a sertification that a given earlier contentual proof uses only tools formalizabe in PA.

Some morals

- The interpretation of Gödel's Second Incompleteness Theorem as yielding *the unprovability of (Hilbert) consistency of* PA *by means formalizable in* PA is a misconception which should be resisted.
- The arithmetical formula Con(PA) is not an adequate representation of Hilbert consistency of PA in the context of its provability. Consistency formulas and their relatives, such as reflection principles, are indispensible in unprovability studies. However, their impact on studies of contentual consistency proofs is limited.
- The consistency scheme ConS(PA) offers an alternative. It respects mathematical intuition, complies with Hilbert's format for consistency, and is provable by means formalizable in PA.

Hilbert's consistency program

The impact of these findings to the original Hilbert's consistency program is not clear and requires additional studies. The next obvious questions in this direction are Hilbert consistency of PRA and ZF, and we can only suggest following Hilbert's advice that "one must exploit the finitary standpoint in a sharper way for the farther reaching consistency proofs."

However, to some extent, Hilbert's consistency program is already vindicated: thinking of proving consistency of a theory by means formalizable in the same theory should no longer be a taboo.

Take home foundational summary

Our starting point was the foundational problem in its entirety:

Can mathematics establish its own consistency? (3)

The prevailing wisdom so far has been "No, by Gödel's Second Incompleteness Theorem, unless mathematics is inconsistent."

We offer a new mathematically well-principled answer to (3):

Yes, for PA. *The question remains open in general.*

Constructive truth and falsity in PA

As a study of schematic reasoning in PA, we consider a theory of constructive truth/falsity. Conceptually, it reads PA-provability of a scheme $\{S(n) \mid n = 0, 1, 2, \ldots\}$ as "for each n, PA $\vdash S(n)$."

Let $t{:}Y$ be a shorthand for the standard formula $Proof(t, Y)$ stating that '*t is a proof of Y in* PA,' $\Box Y$ stand for $Provable(Y)$, i.e., $\exists x(x{:}Y)$.

Definition. An arithmetical sentence F is **constructively true** iff PA$\vdash F$. F is **constructively false** iff

$$\text{PA} \vdash \forall x \Box \neg x{:}F. \tag{4}$$

(4) is equivalent to "PA $\vdash \forall x\ v(x){:}\neg x{:}F$ for some provably total computable term $v(x)$." Indeed, assume (4). Since $u{:}F$ is decidable, given x, enumerate proofs in PA until a proof of $\neg x{:}F$ is met. By (4), such $v(x)$ is provably total. The other direction is immediate.

This notion appeared from the S4/LP formalization of BHK semantics: $\neg F$ Gödel translates to $\Box\neg\Box F$ which realizes as $v(x){:}\neg x{:}F$.

Constructive consistency

Constructive consistency of T is a formula $\mathrm{CCon}(T)$ stating that *for each number,* PA *proves that it is not a proof of a contradiction in* T:

$$\mathrm{CCon}(T) = \forall x \Box_{\mathrm{PA}} \neg x{:}_T \bot.$$

In particular, $\mathrm{CCon}(\mathrm{PA}) = \forall x \Box_{\mathrm{PA}} \neg x{:}_{\mathrm{PA}} \bot$ or, for short,

$$\mathrm{CCon}(\mathrm{PA}) = \forall x \Box \neg x{:}\bot.$$

Both $\mathrm{Con}(T)$ and $\mathrm{CCon}(T)$ are arithmetical formulas which are true iff T is consistent and in this respect they both naturally express consistency of T. However, they have different provability behavior. By G2, PA does not prove Con(PA).

The name "constructive consistency of T" is self-explanatory: it expresses the idea that consistency of each derivation x in T is confirmed constructively by a corresponding PA-proof. Besides, constructive consistency of PA is a special case of the constructive falsity condition.

The following Proposition 2 is a special instance of constructive falsity of refutable formulas. It is also an easy corollary of Feferman's general observation concerning reflection principles.

Proposition 2. PA *proves its own constructive consistency:*

$$\mathsf{PA} \vdash \mathsf{CCon}(\mathsf{PA}).$$

First, we check that $\mathsf{PA} \vdash \Box\bot \rightarrow \mathsf{CCon}(\mathsf{PA})$. Indeed, note that $\mathsf{PA} \vdash \Box\bot \rightarrow \Box\neg x{:}\bot$. By generalization,

$$\mathsf{PA} \vdash \Box\bot \rightarrow \forall x \Box\neg x{:}\bot \ .$$

Furthermore, $\mathsf{PA} \vdash \neg\Box\bot \rightarrow \mathsf{CCon}(\mathsf{PA})$. Indeed, by first-order logic, $\mathsf{PA} \vdash x{:}\bot \rightarrow \exists x(x{:}\bot)$, hence $\mathsf{PA} \vdash \neg\Box\bot \rightarrow \neg x{:}\bot$. By Σ_1-completeness of PA, $\mathsf{PA} \vdash \neg x{:}F \rightarrow \Box\neg x{:}F$, hence $\mathsf{PA} \vdash \neg\Box\bot \rightarrow \Box\neg x{:}\bot$. By generalization,

$$\mathsf{PA} \vdash \neg\Box\bot \rightarrow \forall x \Box\neg x{:}\bot \ .$$

Historically, Proposition 2 was one of the first signs that schematic reasoning in not under the G2 spell.

CCon(PA) vs. Con(PA)

By G2, $\forall x \neg x{:}\bot$ is not internally provable. So, there is no p such that

$$\mathrm{PA} \vdash p{:}\forall x \neg x{:}\bot.$$

Constructive consistency offers a more flexible approach: it allows the aforementioned certification p to depend on x, $p = p(x)$ and we can ask whether

$$\mathrm{PA} \vdash \forall x \;\, p(x){:}\neg x{:}\bot.$$

In a general form this is a question of whether

$$\mathrm{PA} \vdash \forall x \exists y (y{:}\neg x{:}\bot),$$

i.e.

$$\mathrm{PA} \vdash \mathrm{CCon(PA)}$$

which was answered affirmatively in Proposition 2.

Provability of CCon(PA) is not an answer

However, the argument

$$\text{PA } \textit{is consistent because } \text{PA} \vdash \text{CCon(PA)}$$

is circular since it relies on soundness of PA.

Here we face the **deformalization problem**: given that a statement is formally provable in a theory T, produce a rigorous mathematical proof of this statement. This does not necessarily work, e.g., when T is inconsistent, or T is not sound, like $T = \text{PA} + \neg\text{Con(PA)}$, etc.

A general deformalization can work for sound T's, but the assumption of soundness is stronger than the assumption of consistency.

Deformalization also can work on a case-by-case basis: given a specific derivation d in T, repeat its steps countentually and check whether the corresponding assumptions are acceptable.

Normal forms of constructive falsity

Theorem [Normal Form Theorem] F *is constructively false iff*

$$\mathsf{PA} \vdash \mathrm{Con}(\mathsf{PA}) \rightarrow \neg\Box F.$$

Equivalently F is constructive false iff $\mathsf{PA} \vdash \Box F \rightarrow \Box\bot$.

Adequacy Theorem

Adequacy Theorem.

1. PA $\vdash F$ *yields "F is constructively true";*
2. PA $\vdash \neg F$ *yields "F is constructively false";*
3. *"constructively true" and "constructively false" are mutually exclusive;*
4. *"constructively true/false" do not coincide with "provable/refutable";*
5. *"constructively true" and "constructively false" are monotone in the Lindenbaum algebra of* PA*: if* PA $\vdash F \rightarrow G$*, then*
   - *"F is constructively true" yields "G is constructively true,"*
   - *"G is constructively false" yields "F is constructively false."*

Inconsistency is not constructively false.

Theorem 1.

1. $\mathrm{Con(PA)} = \neg\Box\bot$ *is true and constructively false.*
2. $\neg\mathrm{Con(PA)} = \Box\bot$ *is false, but not constructively false.*

Proof. 1. Con(PA) is true in the standard model since PA is sound, hence consistent. Furthermore, since, by the formalized Löb's Theorem,

$$\mathrm{PA} \vdash \Box\neg\Box\bot \to \Box\bot,$$

2. Immediate from Normal Form Theorem, since $\mathrm{PA} \nvdash \Box\Box\bot \to \Box\bot$: otherwise, by Löb's Theorem, $\mathrm{PA} \vdash \Box\bot$ which is not the case.

Rosser sentences

By Rosser's Theorem, there is a sentence R, for which independence in PA follows from simple consistency of PA: if PA is consistent, then nether R nor its negation $\neg R$ is provable.

Theorem 2. *Rosser sentences R and $\neg R$ are both constructively false.*

Proof. The proof of Rosser's Theorem is syntactic and can be formalized in PA:

$$\mathrm{PA} \vdash \neg\Box\bot \rightarrow (\neg\Box R \ \wedge \neg\Box\neg R).$$

By Normal Form Theorem, both R and $\neg R$ are constructively false.

Theorem 3. *There is a true independent in* PA *sentence which is not constructively false.*

Proof. Using the fixed-point lemma, find a sentence L such that

$$\mathrm{PA} \vdash L \leftrightarrow \text{“}L \textit{ is constructively false.”}$$

Formally,

$$\mathrm{PA} \vdash L \leftrightarrow (\Box L \to \Box \bot). \tag{5}$$

If $\mathrm{PA} \vdash L$, then $\mathrm{PA} \vdash \Box L$ and, by (5), $\mathrm{PA} \vdash \Box \bot$ which is not the case. If $\mathrm{PA} \vdash \neg L$, then, by Adequacy Theorem item 2, L is constructively false, hence, $\mathrm{PA} \vdash \Box L \to \Box \bot$. By the fixed point (5), $\mathrm{PA} \vdash L$ - a contradiction in PA. So, L is independent and not constructively false.

Note that L is classically true: otherwise $\Box L$ is false and $\Box L \to \Box \bot$ is vacuously true. By the fixed point (5), L ought to be true as well.

Summary table of classical and constructive truth/falsity

Intersection of classes	**Example**
True and constructively true	$0 = 0$
True and constructively false	$\text{Con}(\text{PA}), R$
True and neither	Constructive Liar L
False and constructively true	$\emptyset$
False and constructively false	$0 = 1$, $\neg R$
False and neither	$\neg\text{Con}(\text{PA})$

Further details can be found in
Artemov, S., 2019. The Provability of Consistency. arXiv:1902.07404.