GÖDEL’S SECOND THEOREM: ITS MEANING AND USE

by
Harvey M. Friedman
Distinguished University Professor
Mathematics, Philosophy, and Computer Science
Ohio State University

This research was partially supported by the John Templeton Foundation.

AMS Special Session
Washington, DC
March 17, 2012
I wish to thank the organizers for inviting me to give this opening lecture for the special session "The Legacy of Gödel's 2nd Incompleteness Theorem for the Foundations of Mathematics."

Gödel's Second Incompleteness Theorem (GSIT) has often been considered to be the single most striking and important result in the foundations of mathematics.

Gödel's original version is specific to an outdated formalization of mathematics based on the Russell/Whitehead Principia Mathematica, but the proof obviously works for a wide variety of formal systems. Gödel is given credit for the general form.

The usual statement of GSIT for the general public is simple enough:

PUBLIC GSIT. No sufficiently strong consistent formal system proves its own consistency.

Of course Public GSIT is not fully rigorous. Other great theorems of mathematics generally have both simple and rigorous formulations.
RIGOROUS GSIT FOR PA

Of course Public GSIT is not fully rigorous. Other great theorems of mathematics generally have both simple and rigorous formulations.

An extreme case of simplicity and rigor is

FERMAT'S LAST THEOREM. (Wiles). For integers \( n > 2 \), \( x^n + y^n = z^n \) has no solution in positive integers.

In order to quickly focus on a number of important issues, I will first discuss GSIT for Peano Arithmetic (PA).

For specificity, we will use a formalization of PA based on 0, S, <, =, and all primitive recursive function symbols.

For \( k \geq 0 \), we define the \( \Sigma_k \) formulas as usual, where bounded quantifiers are ignored.

For \( k \geq 0 \), we define \( I\Sigma_k \) as PA with induction restricted to \( \Sigma_k \) formulas.
RIGOROUS GSIT FOR PA

GSIT FOR PA. PA does not prove its own consistency (assuming PA is consistent).

GSIT for PA is of course simple, but it is not rigorous. This is because the formalization of Con(PA) is not given in the statement.

The formalization of Con(PA) is by no means a trivial matter. We are all confident that we know how to go about constructing this formalization. But note the following:

1. Generally speaking, people do not work with full formalizations of Con(PA). They use hand waving. It is hard enough to give a bug free formalization of Con(PA), let alone give bug free complete proofs of its properties. There is some research on formally verifying certain forms of GSIT. See O’Connor at http://r6.ca/Goedel/goedell1.html I believe that there are some unresolved issues here.
RIGOROUS GSIT FOR PA

1. Generally speaking, people do not work with full formalizations of Con(PA). They use hand waving. ...

2. The general view is that given a substantial amount of time, including substantial debugging, a determined logician can in fact come up with an intelligible statement and proof of GSIT for PA (and more broadly) that meets a very high standard of rigor - without resorting to proof assistants and the issues that they generate.

3. However, the general view is also that achieving 2 will not be rewarded, and so, consequently may never be fully carried out.

4. If different people actually construct Con(PA), then there are going to differ in detail.

5. As an indication of how murky this territory is of actual formalizations, roughly how big is fully formalized Con(PA)?

6. It seems clear that Con(PA) is not a beautiful mathematical object, in the normal sense.
ROLE OF HILBERT BERNAYS
DERIVABILITY CONDITIONS

The Hilbert Bernays derivability conditions are a very
good path (probably the very best path) for creating the
desired fully rigorous treatment of GSIT.

Jerosolow simplified Hilbert Bernays. Later, I made
further simplifications and streamlining, and give a
completely rigorous treatment of GSIT relative to my
simplified and streamlined version of the derivability
conditions in my paper Forty Years on his Shoulders (see
references below).

What remains is the actual construction of a “usual
Con(T)” and a fully rigorous proof that the derivability
conditions are indeed satisfied.

This should really be done, at least for the record, 80
years after the theorem was considered “proved”,
considering that it is arguably still the greatest
theorem in foundations of mathematics.

CAUTION: Under the full circumstances, any attempt to do
this that is not “perfect” is essentially worthless.
Ideal would be some imaginative theorems that say that if the formalization obeys certain transparent conditions concerning its modular construction, then it automatically satisfies the relevant derivability conditions.

Additionally, it would be valuable to prove more: if any two formalizations obeys certain transparent conditions concerning its modular construction, then they are provably equivalent in a weak system. Feferman addresses this issue (Inductively Presented Logics), but it appears to be uncomfortably complicated. See References.

Once again, the proofs need to be readable, digestible, and completely flawless.

BUT for the *statement* - not the proof - of GSIT, we of course especially recommend what is in this presentation.

SOUNDBITE: The usual formalizations of Con are sufficient to drive the Gödel Completeness Theorem, and this is, in turn, sufficient to drive GSIT.
RIGOROUS GSIT FOR PA

The usual procedure is to hand wave a construction of Con(PA) and the proof that it is not provable in PA (assuming PA is consistent). While doing this, certain features of Con(PA) are used. This procedure was followed very carefully by Hilbert and Bernays, when they placed conditions on the proof predicate.

However, we view the Hilbert/Bernays conditions as rather complicated and subtle for the statement of a fundamentally important theorem of mathematics. And it still leaves the very nasty exercise of showing that these Hilbert/Bernays derivability conditions hold of the proof predicate. (These remarks also apply to my simplification/streamlining of the derivability conditions).

We take a different approach to simplicity and rigor in GSIT. We don't get involved in proving GSIT at all. Instead, we give transparent conditions on a formalization of Con(T) that

i. hold for all usual Con(T).
ii. guarantees that the formalization is not provable in T.
RIGOROUS GSIT FOR PA

This has the effect of replacing "usual Con(T)" with a simple and rigorous notion that serves the same principal purpose.

Although all usual Con(T) will obey the conditions, not every sentence obeying the conditions will be a usual Con(T), or even provably equivalent to a usual Con(T). In fact, for some T, some sentences obeying the condition will not even imply the usual Con(T). However, any sentence obeying the conditions will be unprovable in T.

The proof that the usual Con(T) obeys the conditions will be accomplished by the Henkin proof of the Gödel completeness theorem. This argument becomes more delicate when we work with weak theories T.

In the special but commonly studied case of PA, we can use the following particularly simple condition:

A is a sentence in the language of PA such that PA is interpretable in Σ₂ + A.

Thus we are using the fundamental notion of interpretation of one theory in another, generally credited to Tarski.
RIGOROUS GSIT FOR PA

A is a sentence in the language of PA such that
PA is interpretable in $\Pi_2^1 + A$.

The use of $\Pi_2$ simplifies the argument that any usual
Con(PA) satisfies the above condition particularly
easy.

LEMMA 1. PA is interpretable in $\Pi_2$ together with any
"usual Con(PA)".

Proof: Using $\Pi_2 + \text{Con(PA)}$, we build a $\Delta_0^2$ model of PA
via a Henkin construction. Using the generous induction
in $\Pi_2$, we verify that the construction indeed produces
an interpretation of PA. The inductions in PA get
interpreted as $\Sigma_2$ inductions. QED

LEMMA 2. For all $k \geq 0$, $\Pi_{k+1}$ proves any usual Con($\Pi_k$).

Proof: $\Pi_{k+1}$ easily proves the cut elimination theorem.
Therefore, in proving Con($\Pi_k$), $\Pi_{k+1}$ need only consider
cut free proofs. $\Pi_{k+1}$ can develop a truth definition
for the relevant formulas, to which induction can be
applied. QED
**RIGOROUS GSIT FOR PA**

**LEMMA 1.** PA is interpretable in $\mathbb{I} \Sigma_2$ together with any "usual Con(PA)".

**LEMMA 2.** For all $k \geq 0$, $\mathbb{I} \Sigma_{k+1}$ proves any usual Con($\mathbb{I} \Sigma_k$).

**LEMMA 3.** Let $A$ be a sentence in the language of PA such that PA is interpretable in $\mathbb{I} \Sigma_2 + A$. Then PA does not prove $A$. (This assumes PA is consistent).

Proof: Let $A$ be as given. Assume $A$ is provable in PA.

Let $A$ be provable in $\mathbb{I} \Sigma_k$, $k \geq 2$.

Interpretations map inconsistencies to inconsistencies. This tells us that Con($\mathbb{I} \Sigma_2 + A$) $\Rightarrow$ Con($\mathbb{I} \Sigma_{k+1}$) is provable well within $\mathbb{I} \Sigma_2$. Hence $\mathbb{I} \Sigma_2 + \text{Con}(\mathbb{I} \Sigma_k)$ proves Con($\mathbb{I} \Sigma_{k+1}$).

By the Lemma, $\mathbb{I} \Sigma_{k+1}$ proves Con($\mathbb{I} \Sigma_k$). Hence $\mathbb{I} \Sigma_{k+1}$ proves Con($\mathbb{I} \Sigma_{k+1}$). By ordinary GSIT, $\mathbb{I} \Sigma_{k+1}$ is inconsistent. This is a contradiction. QED

**RIGOROUS GSIT FOR PA.** Consider the condition "A is a sentence in the language of PA such that PA is interpretable in $\mathbb{I} \Sigma_2 + A". Any sentence obeying this condition is unprovable in PA (assuming PA is consistent). The "usual Con(PA)" obey this condition.
We can replace $I\Sigma_2$ by any $I\Sigma_k$, $k \geq 0$. However, if we do this, the first claim in Rigorous GSIT for PA has to be argued much more carefully, and will be dependent on a sharper idea of "usual Con(PA)".

We now give a very general form of GSIT. The system $I\Sigma_0(exp)$, which I like to call EFA (exponential function arithmetic), has the language $0,S,+,\cdot,exp,\lt,=,$ and consists of axioms of successor, defining axioms, induction for bounded formulas, and the axioms for logic. EFA is finitely axiomatizable.

EFA is arguably the weakest theory where we can be fully confident that our idea of “usual Con(T)” can be appropriately supported.

We consider consistent recursively axiomatizable theories T in many sorted logic. We use the condition

\[ A \text{ is a } \Pi^0_1 \text{ sentence in the language of EFA such that } T \text{ is interpretable in } EFA + A. \]
RIGOROUS GSIT

A is a \( \Pi^0_1 \) sentence in the language of EFA such that T is interpretable in EFA + A.

We let SEFA = superexponential function arithmetic, based on the language 0,S,+,\cdot,\exp,\text{superexp},<,=, and consists of axioms of successor, defining axioms, induction for bounded formulas, and the axioms for logic. SEFA is finitely axiomatizable.

LEMMA 4. Let T be a consistent recursively axiomatizable system in many sorted logic. T is interpretable in EFA together with any "usual Con(T)".

Proof: We make the usual Henkin construction of a model of T, using any recursive enumeration of the axioms of T. However, we have to be more careful in that we only have EFA. In particular, we don't know that the Henkin construction can be carried out indefinitely. However, it can be carried out as much as it can, which may along a cut. EFA is enough to support the argument. QED
LEMMA 4. Let $T$ be a consistent recursively axiomatizable extension of EFA in many sorted logic. $T$ is interpretable in $I\Sigma_0$ together with any "usual $\text{Con}(T)$".

$L-Con(T)$ asserts that if $A$ is a $\Sigma^0_1$ sentence and $T$ proves $A$, then $A$ is true.

LEMMA 5. SEFA proves $L-Con(EFA)$.

Proof: SEFA proves the cut elimination theorem for predicate calculus - and is in fact equivalent to it over EFA, or even over $I\Sigma_0$. So we can assume in SEFA that we have a cut free proof in EFA of the $\Sigma_1$ sentence $A$. SEFA can provide valuation of terms in $0,S,+,\cdot$, and satisfaction for the relevant subformulas in a cut free proof. Induction is then applied to get the truth of $A$. QED
LEMMA 4. Let T be a consistent recursively axiomatizable extension of EFA in many sorted logic. T is interpretable in $\Sigma_0$ together with any "usual Con(T)".

LEMMA 5. SEFA proves 1-Con(EFA).

LEMMA 6. Let T be a consistent recursively axiomatizable extension of SEFA in many sorted logic. Suppose A is a $\Pi^0_1$ sentence in the language of EFA such that T is interpretable in EFA + A. Then A is not provable in T.

Proof: Let T,A be as given. Suppose A is provable in T. Let A be provable in the finite fragment K of T. We can assume that K extends SEFA.

By the interpretation, Con(EFA + A) $\Rightarrow$ Con(K) is provable in EFA. By Lemma 5, SEFA + A proves Con(EFA + A). Hence SEFA + A proves Con(K). Therefore K proves Con(K). By ordinary GSIT, K is inconsistent. This violates the consistency of T. QED
RIGOROUS GSIT

LEMMA 4. Let $T$ be a consistent recursively axiomatizable extension of EFA in many sorted logic. $T$ is interpretable in $I\Sigma_0$ together with any "usual $\text{Con}(T)$".

LEMMA 5. EFA proves $1^-\text{WCon}(I\Sigma_0)$.

LEMMA 6. Let $T$ be a consistent recursively axiomatizable extension of EFA in many sorted logic. Suppose $A$ is a $\Pi^0_1$ sentence in the language of EFA such that $T$ is interpretable in $I\Sigma_0 + A$. Then $A$ is not provable in $T$.

RIGOROUS GSIT FOR EXTENSIONS OF SEFA. Let $T$ be a consistent recursively axiomatizable extension of SEFA in many sorted logic. Consider the condition "$A$ is a $\Pi^0_1$ sentence in the language of EFA such that $T$ is interpretable in EFA + $A$". Any sentence obeying this condition is unprovable in $T$. The "usual $\text{Con}(T)$" obey this condition.
RIGOROUS GSIT

Experts sometimes construct a “usual Con(T)” over the weak system $I\Sigma_0$ (or even less). We give a corresponding form of Rigorous GSIT.

LEMMA 7. Let $T$ be a consistent recursively axiomatizable system in many sorted logic. $T$ is interpretable in $I\Sigma_0$ together with any "usual expert Con(T)".

Proof: We make the usual Henkin construction of a model of $T$, using any recursive enumeration of the axioms of $T$. However, we have to be yet more careful in that we only have $I\Sigma_0$. This is why they are called experts. QED

We will need the following two technical notions:
WCon($T$) = weak consistency of $T$, asserts that there is no cut free proof of $K \Rightarrow 1 = 0$, where $K$ is a finite sequence of axioms from $T$.

1-WCon($T$) asserts that if $A$ is a $\Sigma^0_1$ sentence and $T$ proves $K \Rightarrow A$, where $K$ is a finite sequence of axioms from $T$, then $A$ is true.
LEMMA 7. Let T be a consistent recursively axiomatizable system in many sorted logic. T is interpretable in $I\Sigma_0$ together with any "usual expert Con(T)".

LEMMA 8. EFA proves $1-WCon(I\Sigma_0)$.

Proof: We can assume in EFA that we have a cut free proof in $I\Sigma_0$ of the $\Sigma_1$ sentence A. EFA can provide valuation of terms in $0,S,+,$, and satisfaction for the relevant subformulas in a cut free proof. Induction in EFA is then applied to get the truth of A. QED

LEMMA 9. Let T be a consistent recursively axiomatizable extension of EFA in many sorted logic. Suppose A is a $\Pi^0_1$ sentence in the language of $I\Sigma_0$ such that T is interpretable in $I\Sigma_0 + A$. Then A is not provable in T.
**RIGOROUS GSIT**

**Lemma 7.** Let $T$ be a consistent recursively axiomatizable system in many sorted logic. $T$ is interpretable in $I\Sigma_0$ together with any "usual expert $\text{Con}(T)$".

**Lemma 8.** EFA proves $1\text{-WCon}(I\Sigma_0)$.

**Lemma 9.** Let $T$ be a consistent recursively axiomatizable extension of EFA in many sorted logic. Suppose $A$ is a $\Pi^0_1$ sentence in the language of $I\Sigma_0$ such that $T$ is interpretable in $I\Sigma_0 + A$. Then $A$ is not provable in $T$.

**Proof:** Let $T,A$ be as given. Suppose $A$ is provable in $T$. Let $A$ be provable in the finite fragment $K$ of $T$. We can assume that $K$ extends EFA.

By the interpretation, $\text{Con}(I\Sigma_0 + A) \Rightarrow \text{WCon}(K)$ is provable in EFA, with a specific bound on the complexity of the formulas for the left side. Hence EFA proves $\text{WCon}(I\Sigma_0 + A) \Rightarrow \text{WCon}(K)$. By Lemma 8, EFA $+ A$ proves $\text{WCon}(I\Sigma_0 + A)$. Hence EFA $+ A$ proves $\text{WCon}(K)$. Therefore $K$ proves $\text{WCon}(K)$. By ordinary GSIT, which works for $\text{WCon}$, we have that $K$ is inconsistent. This violates the consistency of $T$. QED
RIGOROUS GSIT FOR EXTENSIONS OF EFA. Let T be a consistent recursively axiomatizable extension of EFA in many sorted logic. Consider the condition "A is a $\Pi^0_1$ sentence in the language of $I\Sigma_0$ such that T is interpretable in $I\Sigma_0 + A$". Any sentence obeying this condition is unprovable in T. The "usual expert Con(T)" obey this condition.
We mention two additional settings for GSIT. One is theories of finite sequences of nonnegative integers. This supports formalizations where the coding is eliminated or greatly curtailed. A variant of EFA is used in this context. $\Sigma_0^0$ is replaced by a basic theory of finite sequences, which, like $\Sigma_0^0$, is not sufficient to develop exponentiation.

Another important setting is set theory. Here one can talk about structures and relativizations directly, and so important features arise.

Let $T$ be a finite set of sentences in $\in,=\). By the Set Theoretic Satisfiability of $T$, we mean the following sentence in set theory $(\in,=)$, given by relativization:

there exists $D,R$, where $R$ is a set of ordered pairs from $D$, such that $(D,R)$ satisfies each element of $T$. 
ALTERNATIVE RIGOROUS GSIT

Let RST (rudimentary set theory) be the following convenient set theory in $\in, =$.

a. Extensionality.
b. Pairing.
c. Union.
d. Cartesian product.
e. Separation for bounded formulas.

It can be shown that RST is finitely axiomatizable.

RIGOROUS GSIT FOR SET THEORETIC SATISFIABILITY. Let $T$ be a consistent finite set of sentences in $\in, =$ which implies RST. $T$ does not prove the Set Theoretic Satisfiability of $T$.

COROLLARY 7. Let $T$ be a consistent set of sentences in $\in, =$, which implies RST. Let $A$ be a sentence in $\in, =$ such that $T + A$ proves the set theoretic satisfiability of each finite subset of $T$. Then $T$ does not prove $A$.

It does not appear that we can obtain any reasonable form of GSIT for PA readily from Rigorous GSIT for Set Theoretic Satisfiability.
WEAK GSIT

There is a weak form of GSIT which has a particularly transparent proof. We will not discuss a Rigorous Form of this weak form of GSIT.

Let T be a consistent recursively axiomatizable extension of $\Sigma_0$. A provably recursive function of T is a recursive function $f:N \to N$ such that the following holds. There is a $\Sigma^0_1$ formula $A(n,m)$, where

i. T proves $(\forall n)(\exists ! m)(A(n,m))$.
ii. $f(n) = m \iff (n,m)$ is true.

GSIT FOR 1-CON. Let T be a 1-consistent recursively axiomatizable extension of $\Sigma_0$. Then T does not prove any usual 1-Con(T). Furthermore, there is a provably recursive function of $T + 1\text{-Con}(T)$ that is not a provably recursive function of T.

Proof: We define $h:N \to N$. Suppose $n$ codes a proof in $T$ of $(\forall n)(\exists ! m)(A(n,m))$. Then $(\exists ! m)(A(n,m))$ is true, since T is 1-consistent. Set $h(n) = m+1$. Otherwise, set $h(n) = 0$. Then $h$ is a provably recursive function of $T + 1\text{-Con}(T)$, and not a provably recursive function of T. QED
APPLICATIONS OF GSIT

Is there a proof of GSIT, say for PA, close to this simple? Is there a fundamental feature of PA that is changed by adding Con(PA)?

I now mention two applications of GSIT.

THEOREM 8. (Feferman). Let A be a consistent sentence. There is a consistent sentence B such that A is interpretable in B and B is not interpretable in A.

Proof: Let A be consistent. Let B = EFA ∧ Con(A). A is interpretable in B. Since B is true, B is consistent. Suppose B is interpretable in A. Then EFA |- Con(A) ⇒ Con(B). Hence B |- Con(B). Therefore B is inconsistent, by GSIT. Contradiction. QED
APPLICATIONS OF GSIT

GSIT is used in an essential way all through Concrete Mathematical Incompleteness. The latest is Invariant Maximality:

EVERY ORDER INVARIANT SUBSET OF $\mathbb{Q}[0,16]^{32}$ HAS A $\mathbb{Z}^+\uparrow$ INVARIANT MAXIMAL SQUARE.

It is an exercise that this is provably equivalent to the satisfiability of a sentence in predicate calculus, over RCA$^0$. So it is $\Pi^0_1$ in light of its logical form.

The statement follows from small large cardinals, and so is believed to be consistent with ZFC. But why is it unprovable in ZFC?

Within ACA', I use this statement to construct a model of ZFC, thereby proving Con(ZFC). By GSIT, the statement cannot be proved in ZFC.
FINITE GSIT

For any reasonable system $T$, and positive integer $n$, the finite consistency statement $\text{Con}_n(T)$ expresses that “every inconsistency in $T$ uses at least $n$ symbols”. I gave a lower bound of $n^{1/4}$ on the number of symbols required to prove in $\text{Con}_n(T)$ in $T$, provided $n$ is sufficiently large. A more careful version of the argument gives the lower bound of $n^{1/2}$ for sufficiently large $n$. We called this Finite GSIT.

Pudlak gave a much more careful analysis of Finite GSIT, establishing an $(n(\log(n))^{-1/2})$ lower bound and an $O(n)$ upper bound, for systems $T$ satisfying certain reasonable conditions.

It would be very interesting to extend Finite GSIT in several directions. One direction is to give a treatment of a good lower bound for a proof of $\text{Con}_n(T)$ in $T$, which is along the lines of the Hilbert Bernays derivability conditions, adapted carefully for Finite GSIT.

Another direction to take Finite GSIT is to give some versions which are not asymptotic. I.e., they involve specific numbers of symbols that are argued to be related to actual mathematical practice.
FINITE GSIT

Although the very good upper bound of $O(n)$ is given in Pudlak for a proof of $\text{Con}_n(T)$ in $T$, at least for some reasonable systems $T$, the situation seems quite different if we are talking about proofs in $S$ of $\text{Con}_n(T)$, where $S$ is significantly weaker than $T$. For specificity, consider how many symbols it takes to prove $\text{Con}_n(\text{ZF})$ in $\text{PA}$, where $n$ is large. We believe that there is no subexponential upper bound here. There are connections with $P = ? \ NP$. 
THE CONSISTENCY OF PA

We now come to the proposition that PA is consistent.

We claim that PA is interpretable in a finite set of basic mathematical assertions that are quoted and used all the time in mathematics. Thus Con(PA) follows immediately from the consistency of these finite set of basic mathematical assertions.

We already know how to do this in Reverse Mathematics over the base theory RCA\(_0\). Firstly, PA is interpretable in ACA\(_0\). So PA is interpretable in RCA\(_0\) + "every infinite sequence from \([0,1]\) has a convergent subsequence".

But what about in Strict Reverse Mathematics?
A promising approach is to use integers, finite sequences of integers, sets of finite sequences of integers, and polynomial functions from the $\mathbb{Z}^n$ into the $\mathbb{Z}^m$. Have axioms for discrete order ring, sup norm, minimal norm achieved in a nonempty set, length, i-th term, closure under union and complement, appropriate axioms for polynomial functions, and the set of all values of any polynomial function on a set forms a set.

This will yield a system corresponding to ACA$_0$, and so PA is interpretable in this system.

This SRM result could be improved on in many ways.

An inconsistency in PA is far less likely than spontaneous disintegration of the sun, annihilation of human life by black holes, gamma ray bursts landing on Earth, comets landing on Earth, practical finite P = NP, perpetual motion machines, practical time travel, practical cold fusion, realization of Jurassic Park, and million year human life spans.

I believe that futures markets would confirm my belief.
S. Feferman, Arithmetization of metamathematics in a general setting, Fundamenta Mathematicae 49 (1960), 35–92.


