Canad. Math. Bull. 2025, pp. 1–8
http://dx.doi.org/10.4153/S000843952510091X

A classification of incompleteness statements

Henry Towsner and James Walsh

Abstract. For which choices of $X, Y, Z \in \{\Sigma^1_1, \Pi^1_1\}$ does no sufficiently strong X-sound and Y-definable extension theory prove its own Z-soundness? We give a complete answer, thereby delimiting the generalizations of Gödel's second incompleteness theorem that hold within second-order arithmetic.

1 Introduction

Gödel's second incompleteness theorem states that no sufficiently strong consistent and recursively axiomatized theory proves its own consistency. We give an equivalent restatement here:

Theorem 1.1 (Gödel) *No sufficiently strong Π^0_1-sound and Σ^0_1-definable theory proves its own Π^0_1-soundness.*

A theory is Π^0_1-sound (or, in general, Γ-sound) if all of its Π^0_1 theorems (Γ theorems) are true. This notion can be formalized in the axiom systems we consider (see Definition 2.1).

A recent result [5] lifts Gödel's theorem to the setting of second-order arithmetic, where stronger reflection principles are formalizable.

Theorem 1.2 (Walsh) *No sufficiently strong Π^1_1-sound and Σ^1_1-definable theory proves its own Π^1_1-soundness.*

Note that this latter theorem applies to all Σ^1_1-definable theories and not just to the narrower class of Σ^0_1-definable theories.

There are three classes of formulas in the statement of Theorem 1.2, leading to eight variations one could consider, including the original. In this, we consider the other seven. Table 1 records the truth-values of the statement: *No sufficiently strong X-sound and Y-definable theory proves its own Z-soundness.*

To place the ✗s on Table 1, we show how to give appropriately non-standard definitions of arbitrarily strong sound theories. Theorem 1.2 places the first ✓ on the

Received by the editors January 9, 2025; revised June 17, 2025; accepted June 17, 2025.
Published online on Cambridge Core July 24, 2025.
The first author was partially supported by the NSF Grant DMS-2054379.
AMS subject classification: **03F35, 03F40.**
Keywords: Proof theory, incompleteness, second-order arithmetic, reflection principles.

Table 1: Truth values of the statement: *No sufficiently strong X-sound and Y-definable theory proves its own Z-soundness.*

	X	Y	Z
✓	Π^1_1	Σ^1_1	Π^1_1
✗	Π^1_1	Π^1_1	Π^1_1
✗	Σ^1_1	Π^1_1	Π^1_1
✗	Π^1_1	Σ^1_1	Σ^1_1
✗	Σ^1_1	Σ^1_1	Σ^1_1
✗	Π^1_1	Π^1_1	Σ^1_1
✗	Σ^1_1	Σ^1_1	Π^1_1
✓	Σ^1_1	Π^1_1	Σ^1_1

table; for this a "sufficiently strong" theory is any extension of Σ^1_1-AC_0. For the second ✓ a "sufficiently strong" theory is any extension of ATR_0.

Both ✓s can be placed on the table via relatively simple reductions to Gödel's original second incompleteness theorem. However, in [5], it was emphasized that the first ✓ (i.e., Theorem 1.2) can be established by a self-reference-free (indeed, diagonalization-free) proof, which is desirable since applications of self-reference are a source of opacity. In particular, the first ✓ can be established by attending to the connection between Π^1_1-reflection and central concepts of ordinal analysis. To place the second ✓ on the table, we forge a connection between provable Σ^1_1-soundness and a kind of "pseduo-ordinal analysis." Whereas Π^1_1-soundness provably follows from the well-foundedness of a theory's proof-theoretic ordinal, we show that Σ^1_1-soundness provably follows from the statement that a certain canonical ill-founded linear order lacks *hyperarithmetic* descending sequences. In this way, we provide a proof with neither self-reference nor diagonalization of yet another analog of Gödel's second incompleteness theorem.

2 The proofs

2.1 Simplest cases

We begin by placing the first four ✗s on the table.

Definition 2.1 When Γ is a set of formulas, we write $\mathsf{RFN}_\Gamma(U)$ for the sentence stating the Γ-soundness of U (i.e., reflection for formulas from Γ):

$$\mathsf{RFN}_\Gamma(U) := \forall \varphi \in \Gamma\big(\mathsf{Pr}_U(\varphi) \to \mathsf{True}_\Gamma(\varphi)\big).$$

Here, True_Γ is a Γ-definable truth-predicate for Γ-formulas. For the complexity classes that we consider this truth-predicate is available already in the system ACA_0.

For $\Gamma \in \{\Sigma^1_1, \Pi^1_1\}$, we let $\widehat{\Gamma}$ be the dual complexity class. The following result is an immediate consequence of this definition.

Proposition 2.1 *Provably in* ACA_0*, for* $\Gamma \in \{\Sigma^1_1, \Pi^1_1\}$*, T is Γ-sound if and only if $T + \varphi$ is consistent for every true $\widehat{\Gamma}$ sentence φ.*

Theorem 2.2 *Let $\Gamma \in \{\Sigma^1_1, \Pi^1_1\}$. For any sound and arithmetically definable theory S, there is a sound and Γ-definable extension T of S such that $T \vdash \mathsf{RFN}_\Gamma(T)$.*

Proof We define $U := S + \Sigma^1_1\text{-}\mathsf{AC}_0$. Then we define

$$T(\varphi) := U(\varphi) \wedge \mathsf{RFN}_\Gamma(U).$$

That is, $\varphi \in T$ if and only if both $\varphi \in U$ and $\mathsf{RFN}_\Gamma(U)$.

Then $\Sigma^1_1\text{-}\mathsf{AC}_0 \vdash T = \varnothing \vee \big(T = U \wedge \mathsf{RFN}_\Gamma(U)\big)$. Thus, reasoning by cases, $\Sigma^1_1\text{-}\mathsf{AC}_0 \vdash \mathsf{RFN}_\Gamma(T)$. Since $T = U \supseteq \Sigma^1_1\text{-}\mathsf{AC}_0$, $T \vdash \mathsf{RFN}_\Gamma(T)$.

To see that T is Γ-definable, note that U is Γ-definable and that $\mathsf{RFN}_\Gamma(U)$ has an arithmetic antecedent and a Γ consequent.

Finally, note that T is just U, whence it is sound. ■

Remark 2.3 In the proof of Theorem 2.2, we use the Σ^1_1 choice principle only if $\Gamma = \Sigma^1_1$. Indeed, to infer that $\mathsf{RFN}_{\Sigma^1_1}(U)$ is Σ^1_1, we must pull the positively occurring existential set quantifier from $\mathsf{True}_\Gamma(\varphi)$ in front of a universal number quantifier. If $\Gamma = \Pi^1_1$, it suffices to define U as $S + \mathsf{ACA}_0$, since $\mathsf{RFN}_{\Pi^1_1}$ has a finite axiomatization in ACA_0.

2.2 Intermediate cases

We can resolve two more cases with a subtler version of the proof of Theorem 2.2. First, we recall the following useful lemma.

Lemma 2.4 *For T extending ACA_0, $\mathsf{RFN}_{\widehat{\Gamma}}(T)$ does not follow from any consistent extension of T by Γ formulas.*

Proof Suppose $T + \gamma \vdash \mathsf{RFN}_{\widehat{\Gamma}}(T)$ with $\gamma \in \Gamma$. Then $T + \gamma \vdash \mathsf{Pr}_T(\neg\gamma) \to \neg\gamma$. Hence, $T + \gamma \vdash \neg\mathsf{Pr}_T(\neg\gamma)$, i.e., $T + \gamma \vdash \mathsf{Con}(T + \gamma)$. So $T + \gamma \vdash \bot$. ■

The following theorem adds two more ✗s to our table.

Theorem 2.5 *Let $\Gamma \in \{\Sigma^1_1, \Pi^1_1\}$. For any sound and arithmetically definable theory U, there is a $\widehat{\Gamma}$-sound and $\widehat{\Gamma}$-definable extension of U that proves its own Γ-soundness.*

Proof Consider the following formulas:

$$\varphi(x) := x = \ulcorner \mathsf{RFN}_\Gamma(U) \urcorner \vee x = \ulcorner \neg\mathsf{RFN}_{\widehat{\Gamma}}(U + \mathsf{RFN}_\Gamma(U)) \urcorner$$

$$\tau(x) := U(x) \vee \Big(\mathsf{RFN}_{\widehat{\Gamma}}\big(U + \mathsf{RFN}_\Gamma(U)\big) \wedge \varphi(x)\Big).$$

Let T be the theory defined by τ.

Claim *T is $\widehat{\Gamma}$-definable via τ.*

By inspection.

Claim *T is $\widehat{\Gamma}$-sound.*

Since U is sound, $U + \mathsf{RFN}_\Gamma(U)$ is sound, so $\mathsf{RFN}_{\widehat{\Gamma}}(U + \mathsf{RFN}_\Gamma(U))$ holds, and therefore externally, we see that T is the theory:

$$U + \mathsf{RFN}_\Gamma(U) + \neg\mathsf{RFN}_{\widehat{\Gamma}}(U + \mathsf{RFN}_\Gamma(U)).$$

In particular, T has the form $U' + \neg\mathsf{RFN}_{\widehat{\Gamma}}(U')$ where U' is sound. Suppose that $U' + \neg\mathsf{RFN}_{\widehat{\Gamma}}(U') \vdash \sigma$ where σ is false $\widehat{\Gamma}$. Then $U' + \neg\sigma \vdash \mathsf{RFN}_{\widehat{\Gamma}}(U')$. So $\mathsf{RFN}_{\widehat{\Gamma}}(U')$ follows from a consistent extension of U' by Γ formulas, contradicting Lemma 2.4.

Claim $T \vdash \mathsf{RFN}_\Gamma(\tau)$.

From our external characterization of T, we see that

$$T \vdash \neg\mathsf{RFN}_{\widehat{\Gamma}}(U + \mathsf{RFN}_\Gamma(U)).$$

Hence, T proves that τ defines the theory U. Again, appealing to our external characterization of T, $T \vdash \mathsf{RFN}_\Gamma(U)$. Thus, $T \vdash \mathsf{RFN}_\Gamma(\tau)$. ∎

2.3 Limitations

The presentation τ of theory T defined in Theorem 2.5 is clearly somewhat pathological, in part because T cannot discern the identity of τ. Before continuing to the final case, we want to illustrate that such pathologies are inevitable. We use a proof technique suggested at the end of [5].

Proposition 2.6 *Let T be a Γ-definable extension of Σ^1_2-AC_0 that proves Theorems 1.2 and 2.8. Suppose that there is a Γ presentation τ of T such that T proves $\mathsf{RFN}_{\widehat{\Gamma}}(\tau)$. Then both of the following hold:*

(1) *There is a theorem A of T such that $T \vdash \neg\tau(A)$.*
(2) *There is a Γ presentation $\tau^\star$ of T such that T proves $\neg\mathsf{RFN}_{\widehat{\Gamma}}(\tau^\star)$.*

Proof Suppose that each of the following holds:

(1) T is definable by a Γ formula τ;
(2) T extends Σ^1_2-AC_0;
(3) T proves Theorems 1.2 and 2.8;
(4) T proves the $\widehat{\Gamma}$-soundness of τ.

Let σ be a sentence axiomatizing Σ^1_2-AC_0. We have assumed $T \vdash \sigma$. We also have that $T \vdash \mathsf{RFN}_{\widehat{\Gamma}}(\tau)$. Let $A_1, \dots, A_n$ be the axioms of T that are used in the T-proof of $\sigma \wedge \mathsf{RFN}_{\widehat{\Gamma}}(\tau)$. Thus,

$$\vdash (A_1 \wedge \cdots \wedge A_n) \to \big(\sigma \wedge \mathsf{RFN}_{\widehat{\Gamma}}(\tau)\big).$$

Claim $T \vdash \tau(A_1 \wedge \cdots \wedge A_n) \to \neg\mathsf{RFN}_{\widehat{\Gamma}}(\tau)$.

Reason in T. Suppose $\tau(A_1 \wedge \cdots \wedge A_n)$. Then τ extends Σ^1_2-AC_0 and τ proves $\mathsf{RFN}_{\widehat{\Gamma}}(\tau)$. Since τ is a Γ formula, Theorem 1.2 (if $\Gamma = \Sigma^1_1$) or Theorem 2.8 (if $\Gamma = \Pi^1_1$) entails that τ is not $\widehat{\Gamma}$-sound.

Since $T \vdash \mathsf{RFN}_{\widehat{\Gamma}}(\tau)$, the claim implies that $T \vdash \neg\tau(A_1 \wedge \cdots \wedge A_n)$.

On the other hand, consider $\tau^\star(x) := \tau(x) \vee x = \ulcorner A_1 \wedge \cdots \wedge A_n \urcorner$. Note that $\tau^\star$ is a Γ definition of T. Yet, we have just shown that $T \vdash \neg\mathsf{RFN}_{\widehat{\Gamma}}(\tau^\star)$. ■

Remark 2.7 Note that in the proof, we need only assume that T extends Σ^1_2-AC_0 if $\Gamma = \Pi^1_1$. If $\Gamma = \Sigma^1_1$, it suffices to assume that T extends Σ^1_1-AC_0 since Theorem 1.2 applies to extensions of Σ^1_1-AC_0. Likewise, we need not assume that T proves *both* Theorems 1.2 and 2.8. It suffices to assume that T proves Theorem 1.2 (if $\Gamma = \Sigma^1_1$) or that T proves Theorem 2.8 (if $\Gamma = \Pi^1_1$).

2.4 Hardest case

The only remaining case is the dual form of Theorem 1.2.

Theorem 2.8 *No Σ^1_1-sound and Π^1_1-definable extension of* ATR_0 *proves its own Σ^1_1-soundness.*

First, we give a short proof that was discovered by an anonymous referee.

Proof Let T be a Σ^1_1-sound and Π^1_1-definable extension of ATR_0 that proves its own Σ^1_1-soundness. Let Φ be the (conjunction of) the finitely many statements used in the proof (assume that a single sentence axiomatizing ATR_0 is among them). The sentence $\Phi \in T$ is true Π^1_1. Hence, $\Phi + \Phi \in T$ is consistent and $\Phi + \Phi \in T \vdash \mathsf{RFN}_{\Sigma^1_1}(T)$. By running this same argument inside $\Phi + \Phi \in T$, we conclude that $\Phi + \Phi \in T \vdash \mathsf{Con}(\Phi + \Phi \in T)$. Yet $\Phi + \Phi \in T$ is a consistent and finitely axiomatized extension of ATR_0, which contradicts Gödel's second incompleteness theorem. ■

Note that a dual version of this proof also establishes Theorem 1.2.

For the rest of this section, we will give an alternate proof. In [5], Theorem 1.2 was proved using concepts from ordinal analysis. In short, a connection is forged between Π^1_1-soundness and well-foundedness of proof-theoretic ordinals. Since we are now interested in Σ^1_1-soundness, we forge an analogous connection between Σ^1_1-soundness and *pseudo-well-foundedness*, where an order is pseudo-well-founded if it lacks hyperarithmetic descending sequences.

For the rest of this section, assume that T is a Σ^1_1-sound and Π^1_1-definable extension of ATR_0. In what follows, $\mathsf{PWF}(x)$ is a predicate stating that x encodes a recursive pseudo-well-founded order (that is, a linear order with no hyperarithmetic decreasing sequence). A universal quantifier over Hyp can be transformed into an existential set quantifier in the theory ATR_0 [4, Theorem VIII.3.20]. It follows that the statement $\mathsf{PWF}(x)$ is T-provably equivalent to a Σ^1_1 formula.

We will define $\prec_T$ to hold on pairs (e, α) where $e \in \textsc{Rec}$ and $\alpha \in dom(\prec_e)$. We define $(e, \alpha) \prec_T (e', \beta)$ to hold if

there is some $f \in \textsc{Hyp}$ so that $\mathsf{Emb}(f, \prec_e \restriction \alpha + 1, \prec_{e'} \restriction \beta)$ and $T \vdash \mathsf{PWF}(\prec_{e'})$.

Here, we write $\prec_e \restriction \alpha + 1$ for the restriction of the relation $\prec_e$ to $\{\gamma \in dom(\prec_e) \mid \gamma \preceq_e \alpha\}$.

To prove that $T \nvdash \mathsf{RFN}_{\Sigma^1_1}(T)$, it suffices to check that $T \vdash \mathsf{RFN}_{\Sigma^1_1}(T) \to \mathsf{PWF}(\prec_T)$ and that $T \nvdash \mathsf{PWF}(\prec_T)$. Let's take these one at a time.

Claim $T \vdash \mathsf{RFN}_{\Sigma^1_1}(T) \to \mathsf{PWF}(\prec_T)$.

Proof Reason in T. Suppose $\neg\mathsf{PWF}(\prec_T)$. That is, there is a hyp descending sequence f in $\prec_T$. Let $f(n) = (e_n, \beta_n)$. Thus, we have

$$\forall n\ (e_{n+1}, \beta_{n+1}) \prec_T (e_n, \beta_n).$$

By the definition of $\prec_T$, this is just to say:

$$\forall n\ \exists g \in \textsc{Hyp}\ \mathsf{Emb}(g, f(n+1), f(n)),$$

where we abuse notation to write $\mathsf{Emb}(g, f(n+1), f(n))$ for $\mathsf{Emb}(g, \prec_{e_{n+1}} \upharpoonright \beta_{n+1} + 1, \prec_{e_n} \upharpoonright \beta_n)$ to emphasize the role of f in the statement.

The formula $\mathsf{Emb}(g, f(n+1), f(n))$ is Σ^1_1 in the parameter f; this is an application of Σ^1_1-AC_0, which is a consequence of ATR_0 [4, Theorem V.8.3].

ATR_0 proves that $\textsc{Hyp}$ satisfies Σ^1_1 choice, and therefore proves

$$\exists g \in \textsc{Hyp}\ \forall n\ \mathsf{Emb}(g_n, f(n+1), f(n)).$$

Note that g is technically a set encoding the graphs of the countably many functions g_n in the usual way.

Using arithmetic comprehension, we form the composition $g_\star$ of the functions encoded in g—$g_\star(0) = g_0(\beta_1)$, $g_\star(1) = g_0(g_1(\beta_2))$ and so on. The function $g_\star$ is a hyp descending sequence in $\prec_{e_0}$, so $\prec_{e_0}$ is not pseudo-well-founded. Since $f(1) \prec_T f(0)$, we also have $T \vdash \mathsf{PWF}(\prec_{e_0})$. Recall that $\mathsf{PWF}(\prec_{e_0})$ is a Σ^1_1 claim. Hence, $\neg\mathsf{RFN}_{\Sigma^1_1}(T)$. ■

Before addressing the second claim, let's record a dual form of Rathjen's formalized version of Σ^1_1 bounding [3, Lemma 1.1].

Lemma 2.9 *Suppose $H(x)$ is a Π^1_1 formula, such that*

$$\mathsf{ATR}_0 \vdash \forall x\big(H(x) \to \mathsf{PWF}(x)\big).$$

Then for some $e \in \textsc{Rec}$, $\mathsf{ATR}_0 \vdash \mathsf{PWF}(e) \wedge \neg H(e)$.

Remark 2.10 Note that the dual form of Lemma 2.9 has a diagonalization-free proof (with ACA_0 in place of ATR_0) [5, Lemma 4.22]. Kreisel noted (as discussed by Harrison [2, pp. 527–529]) that when a proof can be formalized in Σ^1_1-AC_0, then the proof of the dual result (where all quantifiers are restricted to Hyp) is also valid. This is a proof in ATR_0 since ATR_0 proves that Hyp satisfies Σ^1_1-AC_0. Since the proof of [5, Lemma 4.22] is somewhat involved, we produce here an alternate proof of Lemma 2.9 that incorporates some diagonalization, though we emphasize that diagonalization is not strictly necessary.

Proof [2, Theorem 1.3] implies that PWF (the set of pseduo-well-founded recursive linear orders) is Σ^1_1-complete; note that Harrison does not use self-reference or any other form of diagonalization in his proof, which is the mere application of Kreisel's aforementioned trick (Remark 2.10) to the proof that well-foundedness is

Π^1_1-complete for recursive linear orders. Hence, there is a total recursive function $\{k\}$ such that:

$$\neg H(n) \iff \mathsf{PWF}\big(\{k\}(n)\big).$$

Since the reduction of Π^1_1 predicates to $\mathcal{O}$ can be carried out in ACA_0, *a fortiori* it can be carried out in Σ^1_1-AC_0. When we restrict all quantifiers to Hyp, we thereby get a proof of the dual result for $\mathcal{O}^\star$, which is the set of notations for recursive linear orderings with no hyperarithmetic descending sequences introduced in [1]. Hence,

$$\mathsf{ATR}_0 \vdash \neg H(x) \leftrightarrow \mathsf{PWF}\big(\{k\}(x)\big).$$

By the recursion theorem and the S-m-n theorem, there is an integer e so that ATR_0 proves that $\forall i[\{e\}(i) \simeq \{\{k\}(e)\}(i)]$ (where $\simeq$ means that if either side converges then both sides converge and are equal). Working in ATR_0, $\neg\mathsf{PWF}(e)$ implies $\neg\mathsf{PWF}(\{k\}(e))$, which implies $H(e)$, which implies $\mathsf{PWF}(e)$, which is a contradiction. So $\mathsf{ATR}_0 \vdash \mathsf{PWF}(e)$. (Not that this implies $e \in \textsc{Rec}$ by the definition of $\mathsf{PWF}(e)$.)

Similarly, $H(e)$ implies $\neg\mathsf{PWF}(\{k\}(e))$, which is equivalent to $\neg\mathsf{PWF}(e)$, which we have already ruled out. So $\mathsf{ATR}_0 \vdash \neg H(e)$. ∎

Claim $T \nvdash \mathsf{PWF}(\prec_T)$.

Proof Suppose that T proves $\mathsf{PWF}(\prec_T)$. From the definition of $\prec_T$, it follows that:

$$T \vdash \big(\exists f \in \textsc{Hyp}\ \mathsf{Emb}(f, \prec_x, \prec_T)\big) \to \mathsf{PWF}(\prec_x).$$

The formula $\exists f \in \textsc{Hyp}\ \mathsf{Emb}(f, \prec_x, \prec_T)$ consists of an existential hyp quantifier before a Π^1_1 matrix (the matrix is Π^1_1 since $\prec_T$ refers to provability in T and T is Π^1_1-definable). Hence, there exists a Π^1_1 formula $\pi(x)$ such that:

$$\mathsf{ATR}_0 \vdash \pi(x) \leftrightarrow \exists f \in \textsc{Hyp}\ \mathsf{Emb}(f, \prec_x, \prec_T).$$

By Lemma 2.9, there is some e so that

$$\mathsf{ATR}_0 \vdash \mathsf{PWF}(\prec_e) \wedge \neg\pi(e).$$

Hence, $\mathsf{ATR}_0 \vdash \neg\exists f \in \textsc{Hyp}\ \mathsf{Emb}(f, \prec_e, \prec_T)$. Moreover, since ATR_0 is sound, we infer that $\neg\exists f \in \textsc{Hyp}\ \mathsf{Emb}(f, \prec_e, \prec_T)$ is true.

On the other hand, since T extends ATR_0, we infer that $T \vdash \mathsf{PWF}(\prec_e)$. Hence, the map $\alpha \mapsto (e, \alpha)$ is a canonical hyp embedding of $\prec_e$ into $\prec_T$. So $\neg\exists f \in \textsc{Hyp}\ \mathsf{Emb}(f, \prec_e, \prec_T)$ is false after all. Contradiction. ∎

It follows from the claims that $T \nvdash \mathsf{RFN}_{\Sigma^1_1}(T)$, which completes the proof of Theorem 2.8.

Acknowledgements Thanks to Hanul Jeon for discussion. Thanks to an anonymous referee for many helpful suggestions. Thanks, in particular, for discovering an error in an earlier draft and for suggesting the simple proof of Theorem 2.8, which we have included.

References

[1] S. Feferman and C. Spector, *Incompleteness along paths in progressions of theories*. J. Symb. Logic **27**(1962), 383–390.
[2] J. Harrison, *Recursive pseudo-well-orderings*. Trans. Amer. Math. Soc. **131**(1968), 526–543.
[3] M. Rathjen, *The role of parameters in bar rule and bar induction*. J. Symb. Logic **56**(1991), 715–730.
[4] S. G. Simpson, *Subsystems of second order arithmetic*. Second Edition, Cambridge University Press, New York, 2009.
[5] J. Walsh, *An incompleteness theorem via ordinal analysis*. J. Symb. Logic **18**(2024), 80–96.

Department of Mathematics, University of Pennsylvania, Philadelphia, PA 19104, United States
e-mail: htowsner@math.upenn.edu

Department of Philosophy, New York University, New York, NY 10012, United States
e-mail: jmw534@nyu.edu