Woodin's proof of the Second Incompleteness Theorem for Set Theory

Andrés E. Caicedo

November 4, 2010

As part of the University of Florida Special Year in Logic, I attended a conference at Gainesville on March 5–9, 2007, on *Singular Cardinal Combinatorics and Inner Model Theory*. Over lunch, Hugh Woodin mentioned a nice argument that quickly gives a proof of the second incompleteness theorem for set theory, and somewhat more. I present this argument here.

The proof is similar to that in Thomas Jech, *On Gödel's second incompleteness theorem*, Proceedings of the American Mathematical Society **121 (1)** (1994), 311-313. However, it is semantic in nature: Consistency is expressed in terms of the existence of models. In particular, we do not need to present a proof system to make sense of the result. Of course, thanks to the completeness theorem, if consistency is first introduced syntactically, we can still make use of the semantic approach.

Woodin's proof follows.

Argue within ZFC.

It is fairly easy to formalize first order languages in set theory so that for each formula ϕ we have a formal counterpart (a *code*) $\ulcorner\phi\urcorner$, the map $\phi \mapsto \ulcorner\phi\urcorner$ is recursive, so that the usual syntactic operations with formulas can be carried out on their formalizations, and satisfaction (for set-sized models) is defined.

The key technical tool we require is the fixed-point lemma:

Lemma 1 *For any formula $\psi(x)$ in one free variable, there is a sentence ϕ such that*

$$\phi \iff \psi(\ulcorner\phi\urcorner).$$

Proof: This is well known. Let $\tau(x)$ assert that x is the code of a formula $\mu(y)$ and that $\psi(\ulcorner\mu(\ulcorner\mu\urcorner)\urcorner)$ holds.

Let a be the code for $\tau(x)$. Then $\tau(a)$ iff $\psi(\ulcorner\tau(a)\urcorner)$, so we can take ϕ to be $\tau(a)$. □

From now on, I will abuse language and simply write ϕ for both a formula and its code.

Let $(M, \hat{\in}) \models$ ZFC. For $m, E \in M$, write

$$(m, E)^* = (\{b \in M \mid M \models b \hat{\in} m\}, \{(a, b) \in M \times M \mid M \models \text{“}a, b \hat{\in} m \,\&\, (m, E) \models aEb\text{”}\}).$$

Then $(m, E)^*$ is the actual model that m, E code within $(M, \hat{\in})$.

The following is shown by a straightforward induction on formulas:

Lemma 2 *Suppose* $(M, \hat{\in}) \models \mathsf{ZFC}$. *If* $(M, \hat{\in}) \models$ *"*$(m, E) \models \varphi$*", then* $(m, E)^* \models \varphi$.

(Of course, by considering $\neg\varphi$, it follows immediately that the converse also holds.)

We say that $P(x)$ is a *property of models of set theory* iff $P(M)$ implies $M \models \mathsf{ZFC}$.

Say that P is *hereditary* iff it is a property of models of set theory and, whenever

$$P(M) \,\&\, M \models P(N),$$

then $P(N^*)$.

Theorem 3 *Suppose that* P *is hereditary. Then, either* $P(N)$ *fails for all* N*, or else there is an* N *such that* $P(N) \,\&\, N \models \forall M \neg P(M)$.

Proof: Let $\mathrm{Th}_P = \{\varphi \mid \forall N\,(P(N) \to N \models \varphi)\}$. Using the fixed-point lemma, let ϕ be such that (ZFC proves that) $\phi \leftrightarrow (\neg\phi \in \mathrm{Th}_P)$.

Suppose $P(N) \,\&\, N \models \neg\phi$. Then $N \models \neg\phi \notin \mathrm{Th}_p$, so $N \models$ "$P(M) \,\&\, M \models \phi$" for some $M \in N$. But then $P(M^*) \,\&\, M^* \models \phi$.

We have shown that $\exists N\, P(N)$ implies $\exists N\,(P(N) \,\&\, N \models \phi)$. Fix such N, and note that $N \models \forall M\,(P(M) \to M \models \neg\phi)$.

Suppose $N \models \exists M\, P(M)$. Then $N \models \exists M\,(P(M) \,\&\, M \models \phi)$, contradiction. Hence

$$N \models \forall M \neg P(M),$$

as needed. □

Let "ZFC is consistent" be the assertion that there is a model of ZFC. The second incompleteness theorem follows at once:

Corollary 4 *Either* ZFC *is inconsistent, or else* $\exists M \models \mathsf{ZFC} +$ *"*ZFC *is inconsistent."*

Proof: Let $P(N) \equiv N \models \mathsf{ZFC}$. We claim that P is hereditary. This amounts to showing that $M \models P(N)$ implies $P(N^*)$.

This is because for each true axiom ϕ of ZFC, $M \models$ "$N \models \phi$" implies $N^* \models \phi$, and $M \models$ "$P(N) \to N \models \phi$." □

Corollary 5 *Either there are no* ω*-models of* ZFC *or else there is an* ω*-model of* ZFC *without* ω*-models of* ZFC.

Proof: Let $P(N) \equiv N$ is an ω-model of ZFC. Suppose that $P(M) \,\&\, M \models P(N)$. Then $N^* \models \mathsf{ZFC}$ and $M \models \omega^N \cong \omega$ so $\omega^{N^*} \cong \omega^M \cong \omega$, and $P(N^*)$ follows. □

Corollary 6 *Either there are no transitive models of* ZFC *or else there is a transitive model without transitive models.*

PROOF: Let $P(N) \equiv N$ is a transitive model of ZFC. Let $M \models P(N)$, M transitive. Then $N \subseteq M$, so N is really transitive. $\square$

Remark 7 *If M is an ω-model of* ZFC*, then*

$$M \models \text{“}\exists N \models \mathsf{ZFC}\text{”}.$$

This is because ZFC *proves the completeness theorem, and therefore “*ZFC *is consistent” is (equivalent to) the arithmetic statement “There is no proof from* ZFC *of $0 = 1$”. But the existence of M implies that this statement is true. Now note that, since M is an ω-model, it is correct for arithmetic statements.*

Remark 8 *Similarly, if M is a transitive model of* ZFC*, then*

$$M \models \text{“There is an } \omega\text{-model of } \mathsf{ZFC}\text{”}.$$

This is because of Mostowski's absoluteness theorem: Any transitive model of set theory is correct about Σ^1_1 statements. See, for example, section 13 of Akihiro Kanamori, **The higher infinite: Large cardinals in set theory from their beginnings***, Springer, second edition (2008).*

Note that the statement “There is an ω-model of ZFC*” is Σ^1_1, as it can be expressed by saying that there is a real x coding a model of* ZFC*, and there is a real y coding an order isomorphism of ω onto the natural numbers of the model coded by x.*

Since any transitive model is an ω-model, the existence of M implies that the Σ^1_1 statement that there is an ω-model is true. But then it holds in M.

The following is a version for ω-models of ZFC of a result of Steel on ω-models of second order arithmetic, see John Steel, *Descending Sequences of Degrees*, The Journal of Symbolic Logic, **40 (1)**, (Mar., 1975), 59–61.

If M is a model of set theory, and $M \models$ “N is a model of T” for some theory T, we can think of N as a code for N^*. We use “code” in a slightly more general fashion in the statement below:

Corollary 9 *There is no sequence $(M_n \mid n < \omega)$ of ω-models of* ZFC *such that for all n, there is a code for $(M_m \mid m > n)$ in M_n.*

PROOF: Let $P(N) \equiv$ “N is an ω-model of ZFC and there is a sequence $(M_n \mid n \in \omega)$ of ω-models of ZFC such that $N = M_0$ and, for all n, there is a code for $(M_m \mid m > n)$ in M_n”.

Obviously, P is hereditary, since being an ω-model is. It follows that either the corollary holds, or else there is an N such that $P(N)$ but $N \models \forall M\, \neg P(M)$.

However, the second possibility is impossible, since if $(M_n \mid n \in \omega)$ is a sequence witnessing $P(N)$, then $(M_{n+1} \mid n \in \omega)$ is a sequence witnessing $P(M_1)$. But the code $\vec{M}^*$ for this sequence is in $N = M_0$ and so, in N, P holds of the code for M_1, as witnessed by $\vec{M}^*$. Contradiction. □

Finally, I sketch how to recover the full version of the second incompleteness theorem from Woodin's proof.

Theorem 10 *If T is a consistent recursively enumerable theory that interprets* PA*, then T cannot prove its own consistency.*

PROOF: There are a few additional wrinkles, since we only assume that the theory under consideration interprets PA:

1. It is not clear how to even state the completeness theorem within PA. However, completeness is provable within the system ACA_0 of second order arithmetic (this is shown in Stephen Simpson, **Subsystems of second order arithmetic**, Cambridge University Press, second edition (2010), see for example Theorem IV.3.3). Moreover, for arithmetic statements, ACA_0 is conservative over PA.

2. Although it is fairly straightforward that ACA_0 is conservative over PA for first order statements, we need that this is provable within PA. This can be done in a few ways. See, for example, Joseph Shoenfield, **Mathematical Logic**, A K Peters (2001).

From the above, it follows that arguing within T, we can implement Woodin's proof. □

Typeset using LaTeX2WP.