Gödel's incompleteness theorems

Canonical name	GodelsIncompletenessTheorems
Date of creation	2013-03-22 13:50:36
Last modified on	2013-03-22 13:50:36
Owner	mathcam (2727)
Last modified by	mathcam (2727)
Numerical id	13
Author	mathcam (2727)
Entry type	Theorem
Classification	msc 03B10
Classification	msc 03F45
Related topic	BeyondFormalism
Defines	Gödel's first incompleteness theorem
Defines	incompleteness theorem
Defines	Hilbert-Bernays derivability conditions
Defines	provability logic

Gödel's first and second incompleteness theorems are perhaps the most celebrated results in mathematical logic. The basic idea behind Gödel's proofs is that by the device of Gödel numbering, one can formulate properties of theories and sentences as arithmetical properties of the corresponding Gödel numbers, thus allowing 1st order arithmetic to speak of its own consistency, provability of some sentence and so forth.

The original result Gödel proved in his classic paper *On Formally Undecidable Propositions in Principia Mathematica and Related Systems* can be stated as

Theorem 1 *No theory* **T** *axiomatisable in the type system of PM (i.e., in Russell's theory of types) which contains Peano-arithmetic and is ω-consistent proves all true theorems of arithmetic (and no false ones).*

Stated this way, the theorem is an obvious corollary of Tarski's result on the undefinability of truth. This can be seen as follows. Consider a Gödel numbering G, which assigns to each formula ϕ its Gödel number $\ulcorner\phi\urcorner$. The set of Gödel numbers of all true sentences of arithmetic is $\{\ulcorner\phi\urcorner \mid \mathbb{N} \models \phi\}$, and by Tarski's result it isn't definable by any arithmetic formula. But assume there's a theory T an axiomatisation $\mathbf{Ax}_T$ of which is definable in arithmetic and which proves all true statements of arithmetic. But now $\exists P(P$ is a proof of x from $\mathbf{Ax}_T)$ defines the set of (Gödel numbers of) true sentences of arithmetic, which contradicts Tarski's result.

The proof given above is highly non-constructive. A much stronger version can actually be extracted from Gödel's paper, namely that

Theorem 2 *There is a primitive recursive function* **Gödel***, such that if T is a theory with a p.r. axiomatisation α, and if all primitive recursive functions are representable in T then $\mathbb{N} \models$ **Gödel**(α) but $T \nvdash$ **Gödel**(α)*

This second form of the theorem is the one usually proved, although the theorem is usually stated in a form for which the nonconstructive proof based on Tarski's result would suffice. The proof for this stronger version is based on a similar idea as Tarski's result.

Consider the formula $\exists P(P$ is a proof of x from $\alpha)$, which defines a predicate

$$\mathrm{Prov}_\alpha(x)$$

which represents provability from α. Assume we have numerated the open formulae with one variable in a sequence B_i, so that every open formula

1

occurs. Consider now the sentence $\neg\operatorname{Prov}_\alpha(B_x)$, which defines the non-provability from α predicate. Now, since $\neg\operatorname{Prov}_\alpha(B_x)$ is an open formula with one variable, it must be B_k for some k. Thus we can consider the closed sentence $B_k(k)$. This sentence is equivalent to $\neg\operatorname{Prov}_\alpha(subst(\ulcorner\neg\operatorname{Prov}_\alpha(x)\urcorner), k))$, but since $subst(\ulcorner\neg\operatorname{Prov}_\alpha(x)\urcorner), k)$ is just $B_k(k)$, it "asserts its own unprovability".

Since all the steps we took to get the undecided but true sentence $subst(\ulcorner\neg\operatorname{Prov}_\alpha(x)\urcorner), k)$ is just $B_k(k)$ were very simple mechanic manipulations of Gödel numbers guaranteed to terminate in bounded time, we have in fact produced the p.r. function **Gödel** required by the statement of the theorem.

The first version of the proof can be used to show that also many non-axiomatisable theories are incomplete. For example, consider $PA +$ all true Π_1 sentences. Since Π_1 truth is definable at Π_2-level, this theory is definable in arithmetic by a formula α. However, it's not complete, since otherwise $\exists p(p$ is a proof of x from $\alpha)$ would be the set of true sentences of arithmetic. This can be extended to show that no arithmetically definable theory with sufficient expressive power is complete.

The second version of Gödel's first incompleteness theorem suggests a natural way to extend theories to stronger theories which are exactly as sound as the original theories. This sort of process has been studied by Turing, Feferman, Fenstad and others under the names of ordinal logics and transfinite recursive progressions of arithmetical theories.

Gödel's second incompleteness theorem concerns what a theory can prove about its own provability predicate, in particular whether it can prove that no contradiction is provable. The answer under very general settings is that a theory can't prove that it is consistent, without actually being inconsistent.

The second incompleteness theorem is best presented by means of a provability logic. Consider an arithmetic theory T which is p.r. axiomatised by α. We extend the language this theory is expressed in with a new sentence forming operator $\square$, so that any sentence in parentheses prefixed by $\square$ is a sentence. Thus for example, $\square(0 = 1)$ is a formula. Intuitively, we want $\square(\phi)$ to express the provability of ϕ from α. Thus the semantics of our new language is exactly the same as that of the original language, with the additional rule that $\square(\phi)$ is true if and only if $\alpha \vdash \phi$. There is a slight difficulty here; ϕ might itself contain boxed expressions, and we haven't yet provided any semantics for these. The answer is simple, whenever a boxed expression $\square(\psi)$ occurs within the scope of another box, we replace it with the arithmetical statement $\operatorname{Prov}_\alpha(\psi)$. Thus for example the truth of $\square(\square(0 = 1))$ is equiva-

2

lent to $\alpha \vdash \mathrm{Prov}_\alpha(\ulcorner 0 = 1 \urcorner)$. Assuming that α is strong enough to prove all true instances of $\mathrm{Prov}_\alpha(\ulcorner \phi \urcorner)$ we can in fact interpret the whole of the new boxed language by the translation. This is what we shall do, so formally $\alpha \vdash \phi$ (where ϕ might contain boxed sentences) is taken to mean $\alpha \vdash \phi*$ where $\phi*$ is obtained by replacing the boxed expressions with arithmetical formulae as above.

There are a number of restrictions we must impose on α (and thus on $\Box$, the meaning of which is determined by α). These are known as Hilbert-Bernays derivability conditions and they are as follows

- if $\alpha \vdash \phi$ then $\alpha \vdash \Box(\phi)$

- $\alpha \vdash \Box(\phi) \to \Box\Box(\phi)$

- $\alpha \vdash \Box(\phi \to \psi) \to (\Box\phi \to \Box\psi)$

A statement $Cons$ asserts the consistency of α if its equivalent to $\neg\Box(0 = 1)$. Gödel's first incompleteness theorem shows that there is a sentence $B_k(k)$ for which the following is true $\neg\Box(0 = 1) \to \Diamond(B_k(k)) \wedge \Diamond(\neg(B_k(k))$, where $\Diamond$ is the dual of $\Box$, i.e. $\Diamond(\phi) \leftrightarrow \neg\Box(\neg\phi)$. A careful analysis reveals that this is *provable* in any α which satisfied the derivability conditions, i.e. $\alpha \vdash \neg\Box(0 = 1) \to \Diamond(B_k(k)) \wedge \Diamond(\neg(B_k(k)))$. Assume now that α can prove $\neg\Box(0 = 1)$, i.e. that α can prove its own consistency. Then α can prove $\Diamond(B_k(k)) \wedge \Diamond(\neg(B_k(k)))$. But this means that α can prove $B_k(k)$! Thus α is inconsistent.