Gödel's Incompleteness Theorems

by Sally Cockburn (2013)

1. Gödel Numbering

We begin with Peano's axioms for the arithmetic of the natural numbers (*i.e.* number theory):

(1) Zero is a natural number.

(2) Every natural number has an immediate successor that is also a natural number.

(3) Zero is not the immediate successor of any natural number.

(4) If two natural numbers have equal immediate successors, they are themselves equal.

(5) If a set of natural numbers A contains zero as well as the immediate successor of any natural number in A, then every natural number is in A.

Assume that S is a formal language powerful enough to represent Peano arithmetic. For the sake of simplicity, we assume that S is a somewhat richer language than ZF because it includes symbols for the unary immediate successor operation and the binary operations of addition and multiplication. The symbols of our formal system S are assigned Gödel numbers according to the following table. (Many alternate Gödel numbering systems are possible.)

symbol	Gödel number	meaning
$\neg$	1	not
$\wedge$	2	and
$\vee$	3	or
$\Longrightarrow$	4	implies
$=$	5	equals
0	6	zero
s	7	immediate successor
(	8	left parenthesis
)	9	right parenthesis
$\exists$	10	there exists
$\forall$	11	for all
$+$	12	plus
$\times$	13	times
x	17	numerical variable
x'	19	numerical variable
x''	23	numerical variable

Note that the Gödel number of each numerical variable is a prime number strictly greater than 13; we could continue this list to include countably many more numerical variables. (In what follows, for simplicity we will use y, z instead of x', x'' respectively.)

If p is a string of n symbols, with Gödel numbers $m_1, m_2, \ldots, m_n$, then the Gödel number of p is

$$GN(p) = \pi_1^{m_1} \cdot \pi_2^{m_2} \cdot \cdots \cdot \pi_n^{m_n},$$

where $\pi_1, \pi_2, \ldots, \pi_n$ are the first n primes. Note that we can assign a Gödel number to any finite string of symbols, whether or not it is a well-formed formula ("wff") in S.

Example 1. Let p be the string in the formal language S defined by

$$p := \forall y \, \exists x \, (x = sy).$$

This string is a formula with no free variables, *i.e.* a statement in S. It expresses Peano's second axiom; namely, that every number has an immediate successor. The Gödel numbers of the individual symbols are $10, 17, 9, 13, 7, 13, 4, 6, 17, 8$. Hence the Gödel number of p is

$$\begin{aligned} GN(p) &= 2^{10} \cdot 3^{17} \cdot 5^9 \cdot 7^{13} \cdot 11^7 \cdot 13^{13} \cdot 17^4 \cdot 19^6 \cdot 23^{17} \cdot 29^8 \\ &= 40949887115552894815876095999575957375662238972483734189085960792384518 \\ &\quad\, 00675312220756677800000000000. \end{aligned}$$

Exercise 1. Find the Gödel number of the string $q := \forall x\big(\neg(0 = sx)\big)$. (Note that q is Peano's third axiom.) You may leave $GN(q)$ in factored form!

Exercise 2. Express Peano's fourth axiom as a statement r in S and find $GN(r)$ (in factored form).

Exercise 3. Let $t(x)$ be the predicate, "x is an odd number". Express this predicate as a formula with one free variable, and determine its Gödel number (in factored form). (Note that 2 can be represented as $ss0$.)

A theorem in the formal system S is a statement that can be proved from the axioms using the logical rules of inference, which are displayed in Table 1. A proof of a statement p is a sequence of statements, the last of which is p. For example, we can use universal instantiation to give a proof of the statement $0 \neq 1$:

$$\begin{aligned} q &:= \forall x(\neg(0 = sx)); \\ z &:= \neg(0 = s0). \end{aligned}$$

We can extend Gödel numbering to sequences of strings. If $p_1, p_2, \ldots p_n$ are strings in S with Gödel numbers $GN(p_1), GN(p_2), \ldots, GN(p_n)$, then the Gödel number of the sequence $\sigma = (p_1, p_2, \ldots p_n)$ is

$$GN(\sigma) = \pi_1^{GN(p_1)} \cdot \pi_2^{GN(p_2)} \cdot \cdots \cdot \pi_n^{GN(p_n)},$$

where $\pi_1, \pi_2, \ldots, \pi_n$ are the first n primes. For example, the Gödel number of the proof that $0 \neq 1$ is

$$GN(p, z) = 2^{GN(q)} \cdot 3^{GN(z)}.$$

Table 1. Rules of Inference

Name	Rule of Inference
modus ponens	$p \implies q$ $\underline{p \qquad\quad}$ $\therefore q$
modus tollens	$p \implies q$ $\underline{\neg q \qquad\quad}$ $\therefore \neg p$
addition	$\underline{p \qquad\quad}$ $\therefore p \vee q$
simplification	$\underline{p \wedge q \quad}$ $\therefore p$
conjunction	p $\underline{q \qquad}$ $\therefore p \wedge q$
hypothetical syllogism	$p \implies q$ $\underline{q \implies r}$ $\therefore p \implies r$
disjunctive syllogism	$p \vee q$ $\underline{\neg p \qquad\quad}$ $\therefore q$
universal instantiation	$\underline{\forall x \left(\phi(x)\right) \qquad\qquad\qquad}$ $\therefore \phi(n)$ for any $n \in \mathbb{N}$
existential generalization	$\underline{\phi(n) \text{ for some } n \in \mathbb{N}}$ $\therefore \exists x \left(\phi(x)\right)$

We can now assign a Gödel number to any logical symbol, any numerical variable, any string or any sequence of strings in the formal system S. Conversely, given a number, can we determine if it is the Gödel number of some symbol or string or sequence of strings?

Exercise 4. Show that $100 = 2^2 \cdot 5^2$ cannot be the Gödel number of a symbol, a variable, a string or a sequence of strings.

Exercise 5. The number $8,100,000$ is a Gödel number; of what?

2. Translating Meta-Mathematics into Arithmetic

The point of Gödel numbering is that by representing symbols, strings and sequences of strings in S with numbers, we can translate meta-mathematical sentences into purely arithmetical ones.

Exercise 6. Show that a string p begins with "$\neg$" if and only if $GN(p)$ is divisible by 2, but not by 4.

This exercise illustrates how a meta-mathematical sentence - a sentence about the formal system S - translates into an arithmetical sentence via Gödel numbering. In this case, the meta-mathematical sentence is a predicate, namely,

$$\phi(p) := p \text{ begins with "}\neg\text{"}.$$

The corresponding arithmetical sentence is the predicate

$$T[x] := x \text{ is divisible by 2, but not by 4.}$$

Exercise 7. Express $T[x]$ in the formal language S.

The biconditional statement in Exercise 6 could be expressed as:

$$\phi(p) \text{ if and only if } T[GN(p)].$$

Gödel showed that virtually all meta-mathematical sentences about the formal system can be translated into arithmetical sentences via Gödel numbering. For example, the meta-mathematical predicate "p is a well-formed formula" can be translated into a purely arithmetical property of the Gödel number $GN(p)$, that is

$$p \text{ is a well-formed formula if and only if } Wff[GN(p)].$$

Similarly, recall that the proof of a statement p in S in a sequence of statements, σ, the last of which is p. We can formulate this as a meta-mathematical binary relation:

$$\psi(p, \sigma) := \text{ statement } p \text{ is proved by sequence } \sigma.$$

Both the statement and the proof have Gödel numbers, $GN(p)$ and $GN(\sigma)$ respectively. Gödel proved that the meta-mathematical relation ψ can be represented by a purely arithmetical relation Prf between the Gödel numbers $GN(p)$ and $GN(\sigma)$:

$$\psi(p, \sigma) \text{ if and only if } Prf\big(GN(p), GN(\sigma)\big). \tag{1}$$

The arithmetical relation Prf is complicated and a powerful computer would be required to verify it for a given pair of Gödel numbers, but the important point is that theoretically, at least, it *can* be done. Since it is a purely arithmetical relation between natural numbers, it can be expressed in the formal language S, as a well-formed formula with two free (numerical) variables. (This is where we require the assumption that S is powerful enough to represent arithmetic). From this arithmetical relation, we can define the arithmetical predicate (formula with one free variable)

$$Pr[x] := \exists y \Big(Prf(x, y)\Big).$$

In other words, for a given statement p in S,

$$p \text{ is provable in } S \text{ if and only if } Pr[GN(p)]. \tag{2}$$

The meta-mathematical property that a given statement is provable in S (*i.e.* is a theorem in S) can be translated into a purely arithmetical property of its Gödel number! In fact, we get a stronger, stranger result.

Lemma 1. For any statement p in S, p is provable in S if and only if $Pr[GN(p)]$ is provable in S.

Proof. If p is provable in S, then there is a sequence σ of statements in S, the last of which is p, *i.e.* $\psi(p, \sigma)$. By (1), $Prf\big(GN(p), GN(\sigma)\big)$. By existential generalization, then, $\exists y \Big(Prf(GN(p), y)\Big)$, and this by definition is $Pr[GN(p)]$. The sequence of statements,

$$\Big(Prf\big(GN(p), GN(\sigma)\big),\ Pr[GN(p)]\Big)$$

constitutes a proof of $Pr[GN(p)]$ in S.

Now assume $Pr[GN(p)]$ is provable in S. This means that we can prove in S that the number $GN(p)$ has the arithmetical property Pr. By (2), this is true only if p is provable in S. □

Given an arbitrary arithmetical property $F[x]$, we might ask what it 'means' for the Gödel number $GN(p)$ of a string p to have that property; that is, what does $F[GN(p)]$ 'say' about p? For example, what does it 'mean' about p if the Gödel number of p is a perfect square? Does it have to 'mean' anything? The following theorem, proved by Gödel, provides a curious partial answer.

Theorem 1 (Diagonalization Lemma)**.** *Let $F[x]$ be an arithmetical predicate in S, that is, a predicate about a numerical variable x. Then there exists a statement q in S such that the biconditional " $q \iff F[GN(q)]$" is provable in S.*

A traditionalist might interpret this as saying, "for any arithmetical property F, there is a statement q that says that its Gödel number has property F". A formalist, however, would steer clear of such an interpretation; "$x \iff F[GN(x)]$" is merely a well-formed formula with one free variable. The Diagonalization Lemma simply says that there is some statement q that, when substituted in for the free variable, creates a biconditional statement that can be derived from the axioms of the system using the logical rules of inference.

We now apply the Diagonalization Lemma, not to the predicate $Pr[x]$, but to its negation $\neg Pr[x]$ (which is clearly also a predicate in S). The statement whose existence is asserted by the Diagonalization Lemma is commonly denoted G. Thus we have:

Corollary 2. *There exists a statement G in S such that " $G \iff \neg Pr[GN(G)]$" is provable in S.*

Again, it is tempting to attach a self-reflexive, meta-mathematical meaning to this biconditional statement, such as "G says of itself that it is not provable in S". However, from a formalist perspective, G is just a well-formed formula in S with no free variables; all G 'says' is some (probably hideously complicated) statement in arithmetic.

3. Consistency and Completeness

Definition. Let S be a formal system. Then S is

(1) *consistent* if and only if there is no statement p in S such that both p and $\neg p$ are provable in S;

(2) *complete* if and only if for every statement p in S, either p or $\neg p$ is provable in S.

We can express the definition of consistency in the formal language S with the statement

$$\neg \exists p \Big(Pr[GN(p)] \wedge Pr[GN(\neg p)] \Big),$$

or with the logically equivalent statement

$$\forall p \Big(\neg Pr[GN(p)] \vee \neg Pr[GN(\neg p)] \Big).$$

(Note here that we are quantifying over well-formed formulas with no free variables.)

Exercise 8. Express the definition of completeness in the formal language S.

Note that if S is both consistent and complete, then for every statement p in S, exactly one of p or $\neg p$ is provable in S. This is obviously the gold standard for formal systems! Gödel showed that no formal system S that is powerful enough to represent Peano arithmetic can meet this gold standard.

GÖDEL'S FIRST INCOMPLETENESS THEOREM

If S is consistent, then S is incomplete.

Proof. We will show that if S is consistent, then neither G nor $\neg G$ is provable in S.

Claim 1. If S is consistent, then G is not provable in S.

Proof of Claim 1. We use proof by contrapositive (that is, *modus tollens*). Assume G is provable in S. Since the biconditional $G \iff \neg Pr[GN(G)]$ is provable in S, by the inference rule of simplification, the implication $G \implies \neg Pr[GN(G)]$ is provable in S. By *modus ponens*, $\neg Pr[GN(G)]$ is provable in S. However, we assumed G is provable in S and so by Lemma 1, $Pr[GN(G)]$ is provable in S. Since $\neg Pr[GN(G)]$ and $Pr[GN(G)]$ are both provable in S, by definition S is inconsistent.

Claim 2. If S is consistent, then $\neg G$ is not provable in S.

Proof of Claim 2. Assume that $\neg G$ is provable in S. Since the biconditional $G \iff \neg Pr[GN(G)]$ is provable in S, by simplification again, the implication $\neg Pr[GN(G)] \implies G$ is provable in S. By *modus tollens*, $Pr[GN(G)]$ is provable in S. By Lemma 1, G is provable in S. Since both $\neg G$ and G are provable in S, by definition S is inconsistent. □

Gödel's Second Incompleteness Theorem states that a consistent formal system powerful enough to express Peano arithmetic cannot prove its own consistency. This does *not* mean that such a system is inconsistent, or that its consistency can never be proved; it only means that to prove the consistency of such a formal system, we need another (stronger) formal system.

To talk about a formal system S proving its own consistency, we must first formulate a statement in S that expresses the fact that S is consistent. For this, we need to take a detour.

In classical logic, a contradiction is defined to be a compound statement (*i.e.* a statement made up of other constituent statements, combined using logical connectives) such that the final column in its truth table consists entirely of F's. In other words, the truth value of a contradiction is always 'false', no matter what the truth values of the constituent statements are. Clearly, for any statement p, the compound statement $p \wedge \neg p$ fits this definition of a contradiction. Now, if statement r is a contradiction, then for any statement q, the truth table for implication shows that $r \implies q$ is always true. Hence, if we can prove r, then we can prove any q.

The formalists, however, wanted to replace notions of truth and falsity with provability and non-provability. To replace the classical notion of contradiction, we introduce into any formal system a new symbol, which we will denote by X, which has the status of a formula with no free variables (*i.e.* a statement). We also introduce two new rules of inference associated with X:

$$\frac{p \wedge \neg p}{\therefore X} \qquad\qquad \frac{X}{\therefore \forall q(q)}$$

Exercise 9. Using the rules of inference for X and in Table 1, prove that if a formal system S is inconsistent, then every statement q in S is provable in S. (This is sometimes called the *Principle of Explosion.*)

The Principle of Explosion implies that if S is inconsistent, then for any statement p, *both* p and $\neg p$ are provable in S. Conversely, if *every* statement p in S has the property that both p and $\neg p$ are provable in S, then in particular *there exists* a statement p in S such that both p and $\neg p$ are provable in S (because universal quantification is stronger than existential quantification). Hence we have proved both directions of the biconditional:

$$S \text{ is inconsistent} \iff \forall p \Big(Pr[GN(p)] \wedge Pr[GN(\neg p)] \Big)$$

Taking the contrapositive yields:

$$S \text{ is consistent} \iff \exists p \Big(\neg Pr[GN(p)] \vee \neg Pr[GN(\neg p)] \Big)$$

In the first section, we showed that $z := \neg(0 = s0)$ (which is the expression in S of the arithmetical statement $0 \neq 1$) is provable in S. If S is consistent, then by definition $\neg z := 0 = s0$ is not provable in S. Conversely, if we can demonstrate that $\neg z$ is not provable in S, then by the above, S is consistent. We let

$$Con := \neg Pr[GN(0 = s0)].$$

Then Con is a statement in S that expresses the meta-mathematical statement that S is consistent.

GÖDEL'S SECOND INCOMPLETENESS THEOREM.

If S is consistent, then Con is not provable in S.

Proof. Again, we use proof by contrapositive. Assume Con is provable in S. The proof of Claim 1 in the proof of Gödel's First Incompleteness Theorem can be formalized in S; that is, the statement $Con \Longrightarrow \neg Pr[GN(G)]$ is provable in S. Hence, by our assumption and *modus ponens*, $\neg Pr[GN(G)]$ is provable in S. Since the biconditional $G \Longleftrightarrow \neg Pr[GN(G)]$ is provable in S, the implication $\neg Pr[GN(G)] \Longrightarrow G$ is provable in S. By *modus ponens* again, G is provable in S. By the contrapositive of Claim 1, S is inconsistent. □

4. References

(1) Torkel Franzén, *Gödel's Theorem: An Incomplete Guide to its Use and Abuse*, A. K. Peters Ltd, 2005.

(2) Robert S. Wolf, *A Tour through Mathematical Logic*, Mathematical Association of America, 2005.