GÖDEL FOR CHILDREN

by Piergiorgio Odifreddi

Gödel has originally proved his result [1931] on the limitations of formal systems under special hypotheses (such as the so-called ω-consistency) and by using detailed arguments (involving the boring procedure of arithmetization, and explicit constructions of undecidable formulas). Later researches have improved on both Gödel's results and proofs, generalizing the former (starting with Rosser [1936]) and simplifying the latter (starting with Turing [1936]).

Gödel's results have fascinated logicians, philosophers and laymen, and have been popularized, with various degrees of skill, in books such as Nagel and Newman [1958], Hofstadter [1979], and Smullyan [1987]. Apparently, mathematicians have been less impressed, probably sensing these fanfares as much ado about (if not nothing, at least) little. Various technical expositions, such as Tarski, Mostowski and Robinson [1953], and Smullyan [1961], have vindicated their feelings, and showed how little is indeed needed to prove versions of Gödel's results, granted some logical background.

Usual textbooks have however continued to present these results in weak forms (holding for systems that are variously supposed to extend Peano Arithmetic, or to be axiomatizable or ω-consistent), and with complicated proofs (involving, in particular, coding devices such as Gödel's β-function, detailed arithmetizations, and explicit constructions of undecidable formulas).

Our goal is to show how, by a judicious choice of definitions and proofs, a few pages and no logical background are sufficient to obtain the strongest known version of Gödel's results (for systems suitable for arithmetic) in a fully self-contained and almost trivial way. In particular, we avoid arithmetization (unless one wants to call that an enumeration of the formulas of a system, e.g. in lexicographical order) and self-referential statements.

Our plan is the following. In Section 1 we define the class of recursive functions, by relying on Dedekind's and Peano's analysis of natural numbers.

In section 2 we strightforwardly generate the system of axioms $\mathcal{R}$ of Tarski, Mostowski and Robinson [1953], by looking for a system representing all recursive functions as defined in Section 1. Finally, in Section 3 we prove the undecidability (i.e. the nonrecursiveness) of any consistent formal system extending $\mathcal{R}$ in three words, and deduce from it the incompleteness of any axiomatizatible such system in another three words, all in accordance with the motto of *Tractatus*.

1 Recursive Functions

We start by analyzing the intuitive picture of the natural numbers, trying to characterize their structure. Something is clear: the natural numbers are all in a single discrete row, with a first but no last element. Since what matters to us is just their mutual relationship and not their ultimate individual nature, we may imagine them as obtained from a first element (the number 0), by iteration of a generation procedure (the successor operation $\mathcal{S}$). Thus the numbers are

$$0 \;\; \mathcal{S}(0) \;\; \mathcal{S}(\mathcal{S}(0)) \;\; \cdots$$

or (by using now the natural numbers metalinguistically, to indicate the number of iterations of $\mathcal{S}$)

$$0 \;\; \mathcal{S}^1(0) \;\; \mathcal{S}^2(0) \;\; \cdots$$

We simply write n for $\mathcal{S}^n(0)$.

Three axioms that we take for granted from our intuitive picture above are the following (in a first-order logic with equality):

Axioms 1.1 (Dedekind [1888])

A1 $\mathcal{S}(x) = \mathcal{S}(y) \to x = y$

A2 $0 \neq \mathcal{S}(y)$

A3 $x \neq 0 \to (\exists y)(x = \mathcal{S}(y))$.

They say that the successor induces an isomorphism between ω (the set of natural numbers) and $\omega - \{0\}$. Also, they rule out some unwanted pictures of the natural numbers, like ones with cycles, or with two infinite sequences of elements like

$$a_0 \;\; a_1 \;\; a_2 \;\; \cdots \;\; b_0 \;\; b_1 \;\; b_2 \;\; \cdots$$

Unfortunately, they leave space for structures like

$$a_0 \ a_1 \ a_2 \ \cdots \ \cdots \ b_{-2} \ b_{-1} \ b_0 \ b_1 \ b_2 \ \cdots$$

and to be able to isolate just the initial part of these structures we need to say that every element can be reached from 0 by a *finite number* of applications of $\mathcal{S}$. This seems to involve the very notion of integer that we are trying to characterize, and might seem to be circular. (It is actually impossible to do this in a first-order way).

One way of restricting the possible models of axioms A1–A3 is by introducing new functions. The simplest nontrivial formalization of arithmetic that can be thus obtained is **Robinson's Arithmetic** $\mathcal{Q}$ (Robinson [1950]): it consists, in the language of first-order logic with equality, of a constant 0, functional symbols $\mathcal{S}$, $+$ and $\cdot$, and the axioms A1–A3 together with the defining axioms for $+$ and $\cdot$:

Axioms 1.2 (Grassmann [1861])

A4 $x + 0 = x$

A5 $x + \mathcal{S}(y) = \mathcal{S}(x + y)$

A6 $x \cdot 0 = 0$

A7 $x \cdot \mathcal{S}(y) = x \cdot y \, + \, x.$

More generally, we can take an operational stand, and see how we can deal with properties of natural numbers. Suppose we have a property φ of natural numbers and we wish to check that it holds for every number. From the way the numbers are generated, this follows if the property holds for 0 and it propagates through the successor operation, since every number is obtained from 0 by a finite iteration of $\mathcal{S}$. This is expressed by the **Axiom of Induction**:

Axiom 1.3 (Dedekind [1888]) *If φ is a formula with one free variable then*

A8 $\varphi(0) \wedge (\forall x)[\varphi(x) \rightarrow \varphi(\mathcal{S}(x))] \rightarrow (\forall y)\varphi(y).$

In terms of sets this means that any set containing 0 and closed under successor contains ω, or that the numerals $\mathcal{S}^n(0)$ exhaust the natural numbers.

For our purposes it is better to express this principle in the equivalent form of **Complete Induction**, which refers to the natural ordering of the natural numbers, that can be introduced for example as:

$$\begin{aligned} x \leq y &\Leftrightarrow (\exists z)(x + z = y) \\ x < y &\Leftrightarrow x \leq y \wedge x \neq y. \end{aligned}$$

We then have the following equivalent form of A8:

$$(\forall z)[(\forall x < z)\varphi(x) \rightarrow \varphi(z)] \rightarrow (\forall y)\varphi(y).$$

By writing ψ in place of $\neg\varphi$ and taking the contrapositive, Complete Induction is equivalent to the following **Least Number Principle**:

$$(\exists y)\psi(y) \rightarrow (\exists z)[\psi(z) \wedge (\forall x < z)\neg\psi(x)].$$

Its content is simply that if we know that a number with a certain property exists, then we also know that there is the least number satisfying that property. A general formulation of this principle (with parameters) in terms of functions is:

Definition 1.4 (Kleene [1936]) *A function f is defined from a relation R by* **μ-recursion**[1] *if*

1. *R is a regular predicate, i.e. $(\forall \vec{x})(\exists y)R(\vec{x}, y)$.*
2. *$f(\vec{x}) = \mu y R(\vec{x}, y)$, where $\mu y R(\vec{x}, y)$ is the least number y such that $R(\vec{x}, y)$ holds.*

Similarly, f is defined from g by μ-recursion if

1. *$(\forall \vec{x})(\exists y)(g(\vec{x}, y) = 0)$*
2. *$f(\vec{x}) = \mu y(g(\vec{x}, y) = 0)$.*

Note that the Least Number Principle can be simply written in μ-notation as

$$(\exists y)\psi(y) \rightarrow (\exists z)(z = \mu y \psi(y)).$$

We are now ready for our definition of the notion of recursiveness. The idea is simple: we take an inductive approach, by starting from the functions

[1] μ is the Greek equivalent of the first letter of 'minimum'.

corresponding to 0, $\mathcal{S}$, $+$ and $\cdot$, and by successively building up new functions using μ-recursion. We will also permit a rudimentary logical intuition to contribute to the class, both in initial functions (identities or projections, and equality) and in building rules (composition of known functions), to allow for useful manipulations. We are thus led to the following notion.

Definition 1.5 (Gödel [1931], Kleene [1936a]) *The class of* **recursive functions** *is the smallest class*

1. *containing*

$$\begin{aligned} \mathcal{O}(x) &= 0 \\ \mathcal{S}(x) &= x+1 \\ \mathcal{I}_i^n(x_1,\ldots,x_n) &= x_i \qquad (1 \le i \le n) \end{aligned}$$

   together with $+$, $\cdot$, and the characteristic function of equality

$$\delta(x,y) = \begin{cases} 1 & \textit{if } x = y \\ 0 & \textit{otherwise} \end{cases}$$

2. *closed under composition, i.e. the schema that given g_1,...,g_m,h produces*

$$f(\vec{x}) = h(g_1(\vec{x}),\ldots,g_m(\vec{x}))$$

3. *closed under μ-recursion.*

This formulation of recursiveness is equivalent to the usual standard ones (see Odifreddi [1989]). Also usual are the following definitions:

- a *recursive predicate* is a predicate whose characteristic function is recursive
- a *recursively enumerable (r.e.) set* is a set which is (empty or) the image of a recursive function.

2 Representability of Functions in Formal Systems

We attack the problem in a general way, by isolating minimal conditions (which will turn out to be very weak) sufficient to represent every recursive function.

In the following, *formal system* will always mean 'system extending classical first-order logic with equality, having constants terms $\overline{n}$ (called numerals) for each n, and admitting a recursive enumeration of its formulas (with a given number of free variables)'.[2] A formal system is called:

- *axiomatizable* if the set of (numbers corresponding to) its theorems is r.e.
- *decidable* if the set of (numbers corresponding to) its theorems is recursive
- *consistent* if, for every sentence (i.e. formula without free variables) α, at most one of α and $\neg\alpha$ is provable
- *complete* if, for every sentence α, at least one of α and $\neg\alpha$ is provable.

Definition 2.1 (Tarski [1931], Gödel [1931], [1934], Tarski, Mostowski and Robinson [1953]) *Given a formal system $\mathcal{F}$ and a function f, we say that:*

1. *f is* **weakly representable** *in $\mathcal{F}$ if, for some formula φ of the language of $\mathcal{F}$,*

$$f(x_1, \ldots, x_n) = y \quad \Leftrightarrow \quad \vdash_{\mathcal{F}} \varphi(\overline{x}_1, \ldots, \overline{x}_n, \overline{y})$$

2. *f is* **representable** *in $\mathcal{F}$ if, for some formula φ,*

$$\begin{aligned} f(x_1, \ldots, x_n) = y &\quad \Rightarrow \quad \vdash_{\mathcal{F}} \quad \varphi(\overline{x}_1, \ldots, \overline{x}_n, \overline{y}) \\ f(x_1, \ldots, x_n) \neq y &\quad \Rightarrow \quad \vdash_{\mathcal{F}} \neg\varphi(\overline{x}_1, \ldots, \overline{x}_n, \overline{y}) \end{aligned}$$

3. *f is* **strongly representable** *in $\mathcal{F}$ if for some formula φ, f is representable by φ, and moreover the following uniqueness condition holds:*

$$\vdash_{\mathcal{F}} (\forall y)(\forall z)[\varphi(\overline{x}_1, \ldots, \overline{x}_n, y) \wedge \varphi(\overline{x}_1, \ldots, \overline{x}_n, z) \ \rightarrow \ y = z].$$

The relationships among the various notions are: if f is strongly representable then it is representable, and if f is representable in a consistent formal system then it is weakly representable (because if $\mathcal{F}$ is consistent and $\vdash_{\mathcal{F}} \neg\varphi$ then $\not\vdash_{\mathcal{F}} \varphi$).

[2] This last condition is trivially satisfied if the system has, for example, only a finite number of nonlogical symbols.

The two conditions

$$f(x_1,\dots,x_n)=y \Rightarrow \vdash_{\mathcal{F}} \varphi(\overline{x}_1,\dots,\overline{x}_n,\overline{y})$$
$$\vdash_{\mathcal{F}} (\forall y)(\forall z)[\varphi(\overline{x}_1,\dots,\overline{x}_n,y)\wedge\varphi(\overline{x}_1,\dots,\overline{x}_n,z) \ \rightarrow\ y=z]$$

are equivalent to the unique condition

$$\vdash_{\mathcal{F}} (\forall y)[\varphi(\overline{x}_1,\dots,\overline{x}_n,y) \ \leftrightarrow\ y=\overline{f(x_1,\dots,x_n)}].$$

Moreover, if $\mathcal{F}$ is such that

$$x\neq y \ \Rightarrow\ \vdash_{\mathcal{F}} \neg(\overline{x}=\overline{y})$$

then strong representability of f in $\mathcal{F}$ is equivalent to the unique condition

$$\vdash_{\mathcal{F}} (\forall y)[\varphi(\overline{x}_1,\dots,\overline{x}_n,y) \ \leftrightarrow\ y=\overline{f(x_1,\dots,x_n)}].$$

Proposition 2.2 *If $\mathcal{F}$ is a consistent formal system with a predicate $<$ satisfying the schemata*

1. $\neg(x<\overline{0})$
2. $x<\overline{n+1} \ \leftrightarrow\ x=\overline{0}\vee\cdots\vee x=\overline{n}$
3. $x<\overline{n} \ \vee\ x=\overline{n} \ \vee\ \overline{n}<x$

then any function representable in $\mathcal{F}$ is strongly representable in it.

Proof. Suppose ψ represents f in $\mathcal{F}$. Then the formula

$$\varphi(x_1,\dots,x_n,y)\Leftrightarrow\psi(x_1,\dots,x_n,y)\wedge(\forall z<y)\neg\psi(x_1,\dots,x_n,z)$$

strongly represents f. Indeed:

- If $f(x_1,\dots,x_n)=y$ then $f(x_1,\dots,x_n)\neq z$, for every $z<y$. By representability of f via ψ

  $$\vdash_{\mathcal{F}} \neg\psi(\overline{x}_1,\dots,\overline{x}_n,\overline{0})\wedge\cdots\wedge\neg\psi(\overline{x}_1,\dots,\overline{x}_n,\overline{y-1})\wedge\psi(\overline{x}_1,\dots,\overline{x}_n,\overline{y}).$$

  Axioms 1 and 2 (depending on whether $y=0$ or $y>0$) take care of all the $z<\overline{y}$ in the first part of the formula. Thus

  $$\vdash_{\mathcal{F}} \psi(\overline{x}_1,\dots,\overline{x}_n,\overline{y})\wedge(\forall z<\overline{y})\neg\psi(\overline{x}_1,\dots,\overline{x}_n,z)$$

  (if $y=0$ the second part is vacuously true, since there is no $z<\overline{y}$), and $\vdash_{\mathcal{F}} \varphi(\overline{x}_1,\dots,\overline{x}_n,\overline{y})$.

- If $f(x_1, \ldots, x_n) \neq y$ then, by representability, $\vdash_{\mathcal{F}} \neg\psi(\overline{x}_1, \ldots, \overline{x}_n, \overline{y})$ and so $\vdash_{\mathcal{F}} \neg\varphi(\overline{x}_1, \ldots, \overline{x}_n, \overline{y})$.

- To show the uniqueness condition we prove (see the comments after 2.1) that

$$\vdash_{\mathcal{F}} (\forall y)[\varphi(\overline{x}_1, \ldots, \overline{x}_n, y) \;\Leftrightarrow\; y = \overline{f(x_1, \ldots, x_n)}].$$

  We have $\vdash_{\mathcal{F}} \varphi(\overline{x}_1, \ldots, \overline{x}_n, \overline{f(x_1, \ldots, x_n)})$ from the first part of the proof. Suppose now $\vdash_{\mathcal{F}} \varphi(\overline{x}_1, \ldots, \overline{x}_n, y)$. By axiom 3 the only possibilities are

$$y < \overline{f(x_1, \ldots, x_n)} \;\vee\; y = \overline{f(x_1, \ldots, x_n)} \;\vee\; \overline{f(x_1, \ldots, x_n)} < y.$$

  The first one is ruled out, since from $\vdash_{\mathcal{F}} \varphi(\overline{x}_1, \ldots, \overline{x}_n, \overline{f(x_1, \ldots, x_n)})$ we have $\vdash_{\mathcal{F}} \neg\psi(\overline{x}_1, \ldots, \overline{x}_n, y)$, while from $\vdash_{\mathcal{F}} \varphi(\overline{x}_1, \ldots, \overline{x}_n, y)$ (assumed by hypothesis) we have $\vdash_{\mathcal{F}} \psi(\overline{x}_1, \ldots, \overline{x}_n, y)$, and $\mathcal{F}$ is consistent. Similarly we can rule out $\overline{f(x_1, \ldots, x_n)} < y$. Then $y = \overline{f(x_1, \ldots, x_n)}$. $\square$

The notion of representability makes sense for predicates as well:

Definition 2.3 *Given a formal system $\mathcal{F}$ and a relation R, we say that:*

1. *R is* **weakly representable** *if, for some φ,*

$$R(x_1, \ldots, x_n) \;\Leftrightarrow\; \vdash_{\mathcal{F}} \varphi(\overline{x}_1, \ldots, \overline{x}_n)$$

2. *R is* **representable** *if, for some φ,*

$$\begin{aligned} R(x_1, \ldots, x_n) \;&\Leftrightarrow\; \vdash_{\mathcal{F}} \varphi(\overline{x}_1, \ldots, \overline{x}_n) \\ \neg R(x_1, \ldots, x_n) \;&\Leftrightarrow\; \vdash_{\mathcal{F}} \neg\varphi(\overline{x}_1, \ldots, \overline{x}_n). \end{aligned}$$

Note that if the characteristic function c_R of R is (weakly) represented by $\varphi(x_1, \ldots, x_n, z)$, then R is (weakly) representable by $\varphi(x_1, \ldots, x_n, \overline{1})$. Also, if $\mathcal{F}$ is such that

$$x \neq y \;\Rightarrow \vdash_{\mathcal{F}} \neg(\overline{x} = \overline{y})$$

and R is represented by $\varphi(x_1, \ldots, x_n)$, then c_R is (strongly) representable by

$$(\varphi(x_1, \ldots, x_n) \wedge z = \overline{1}\,) \;\vee\; (\neg\varphi(x_1, \ldots, x_n) \wedge z = \overline{0}\,)$$

(this follows from the comments after 2.1).[3]

Since if $f(x_1,\dots,x_n)=\mu y R(x_1,\dots,x_n,y)$ then

$$f(x_1,\dots,x_n)=y \;\Leftrightarrow\; R(x_1,\dots,x_n,y)\wedge(\forall z<y)\neg R(x_1,\dots,x_n,z),$$

the axioms of proposition 2.2 imply, by means of the same proof, that the strongly representable functions are closed under μ-operator. We turn now to the other conditions.

Proposition 2.4 *If $\mathcal{F}$ is a formal system such that*

$$x\neq y \;\Rightarrow \vdash_{\mathcal{F}} \neg(\overline{x}=\overline{y})$$

then the functions strongly representable in $\mathcal{F}$ are closed under composition.

Proof. Suppose

$$f(\vec{x})=h(g_1(\vec{x}),\dots,g_m(\vec{x}))$$

and g, h_i are strongly represented by, respectively, χ and ψ_i. Then f is strongly represented by

$$\varphi(\vec{x},y) \;\Leftrightarrow\; (\exists y_1)\dots(\exists y_m)[\psi_1(\vec{x},y_1)\wedge\cdots\wedge\psi_m(\vec{x},y_m)\wedge\chi(y_1,\dots,y_m,y)].$$

To show this we use (since we have the appropriate axioms for $\mathcal{F}$) the form of strong representability given in the comments after 2.1. Then:

- if $f(\vec{x})=y$ let $h_i(\vec{x})=y_i$ and $g(y_1,\dots,y_m)=y$, so that $\vdash_{\mathcal{F}} \psi_i(\vec{\overline{x}},\overline{y}_i)$ and $\vdash_{\mathcal{F}} \chi(\overline{y}_1,\dots,\overline{y}_m,\overline{y})$. Then $\vdash_{\mathcal{F}} \varphi(\vec{\overline{x}},\overline{y})$ and $\vdash_{\mathcal{F}} \varphi(\vec{\overline{x}},\overline{f(\vec{x})})$.
- if $\vdash_{\mathcal{F}} \varphi(\vec{\overline{x}},y)$ let $y_1,\dots,y_m$ be such that

  $$\vdash_{\mathcal{F}} \psi_1(\vec{\overline{x}},y_1)\wedge\cdots\wedge\psi_m(\vec{\overline{x}},y_m)\wedge\chi(y_1,\dots,y_m,y).$$

  By strong representability it must be $y_i=\overline{h_i(\vec{x})}$ and thus

  $$y=\overline{g(h_1(\vec{x}),\dots,h_m(\vec{x}))}=\overline{f(\vec{x})}.\quad\square$$

We are now ready to conclude our search for axioms which allow representability of every recursive function.

[3]The axioms are needed even for simple representability of c_R, because when $c_R(x_1,\dots,x_n)\neq z$ and $z\neq 0,1$ we need to know $\overline{z}\neq\overline{0},\overline{1}$ to be able to infer that the formula intended to represent c_R is not provable.

Theorem 2.5 (Gödel [1936], Mostowski [1947], Tarski, Mostowski and Robinson [1953]) *In any formal system $\mathcal{F}$ with a predicate $<$ and functions $+$ and $\cdot$ satisfying the following schemata, any recursive function is (strongly) representable (and thus, if the system is consistent, also weakly representable):*

B1 $\neg(\overline{x} = \overline{y})$, *for* $x \neq y$

B2 $x < \overline{n} \ \vee \ x = \overline{n} \ \vee \ \overline{n} < x$

B3 $\neg(x < \overline{0})$

B4 $x < \overline{n+1} \ \leftrightarrow \ \ x = \overline{0} \vee \cdots \vee x = \overline{n}$

B5 $\overline{x} \ + \ \overline{y} \ = \ \overline{x+y}$

B6 $\overline{x} \ \cdot \ \overline{y} \ = \ \overline{x \cdot y}$

Proof. We proceed by induction on Definition 1.5. We have just proved that closure under composition is implied by B1, and closure under μ-recursion follows from B2–B4, as in 2.2. It then suffices to note that:

- Equality is representable because if $x = y$ then obviously $\vdash_{\mathcal{F}} \overline{x} = \overline{y}$, and if $x \neq y$ then $\vdash_{\mathcal{F}} \neg(\overline{x} = \overline{y})$ by B1. Then its characteristic function is representable too.

- $\mathcal{O}$ is representable by
$$\varphi(x, z) \ \Leftrightarrow \ z = \overline{0}.$$

- Sum is representable by
$$\varphi(x, y, z) \ \Leftrightarrow \ x + y = z.$$
Indeed, if $x + y = z$ then $\vdash_{\mathcal{F}} \overline{x+y} = \overline{z}$ and (by B5) $\vdash_{\mathcal{F}} \overline{x} + \overline{y} = \overline{z}$, i.e. $\vdash_{\mathcal{F}} \varphi(\overline{x}, \overline{y}, \overline{z})$. And if $x + y \neq z$ then $\vdash_{\mathcal{F}} \neg(\overline{x+y} = \overline{z})$ by B1 and $\vdash_{\mathcal{F}} \neg(\overline{x} + \overline{y} = \overline{z})$ by B5, i.e. $\vdash_{\mathcal{F}} \neg\varphi(\overline{x}, \overline{y}, \overline{z})$.

- Product is similarly represented by
$$\varphi(x, y, z) \ \Leftrightarrow \ x \cdot y = z.$$

- Successor is representable by
$$\varphi(x, z) \ \Leftrightarrow \ x + \overline{1} = z.$$

- Identities $\mathcal{I}_i^n$ are obviously represented by

$$\varphi(x_1, \ldots, x_n, z) \;\Leftrightarrow\; z = x_i. \quad \square$$

The axioms B1–B6 define a theory $\mathcal{R}$ with infinitely many axioms (which, by Tarski, Mostowski and Robinson [1953], is not finitely axiomatizable). Since $\mathcal{R}$ is our reference point for the undecidability and incompleteness results, it pays to stop for a moment to consider examples of systems extending it.

Robinson's Arithmetic $\mathcal{Q}$ (see above) is an extension of $\mathcal{R}$, and thus provides an example of a finitely axiomatized theory in which every recursive function is representable. Robinson has proved that if any one of the axioms of $\mathcal{Q}$ is removed, then some recursive function is not strongly representable. Thus $\mathcal{Q}$ *is a minimal finitely axiomatizable theory in which every recursive function is strongly representable.* See Tarski, Mostowski and Robinson [1953] for details.

Another interesting example of (definitional) extension of $\mathcal{R}$ is the system of Set Theory with the following three axioms:

S1 $(\forall x)(\forall y)[(\forall z)(z \in x \leftrightarrow z \in y) \;\rightarrow\; x = y]$

S2 $(\exists x)(\forall y)(y \notin x)$

S3 $(\forall x)(\forall y)(\exists z)(\forall v)(v \in z \;\leftrightarrow\; v \in x \vee v = y)$.

They express, respectively, extensionality, existence of the empty set $\emptyset$, and existence of the generalized successor $x \cup \{y\}$, and again provide a finitely axiomatized theory in which every recursive function is representable. See Smielew and Tarski [1950] for details.

In particular, the results below will apply to all usual systems for arithmetic (like Peano's) and set theory (like Zermelo-Fraenkel's).

3 The Undecidability and Incompleteness Results

The following is a general formulation of the so-called **Gödel's First Theorem**.

Theorem 3.1 (Post [1922], Gödel [1931], [1934], Rosser [1936], Church [1936], Tarski, Mostowski and Robinson [1953])

1. *Every consistent formal system $\mathcal{F}$ extending $\mathcal{R}$ is undecidable.*

2. If, moreover, $\mathcal{F}$ is axiomatizable, then $\mathcal{F}$ is incomplete.

Proof. Let $\{\psi_n\}_{n\in\omega}$ be a recursive enumeration of the formulas in the language of $\mathcal{F}$, with one free variable. If $\mathcal{F}$ is decidable, the diagonal set

$$n \in F \;\Leftrightarrow \vdash_{\mathcal{F}} \psi_n(\overline{n})$$

is recursive, and then so is $\overline{F}$.[4] Every recursive set is representable in $\mathcal{F}$ by 2.5 and hence, by consistency of $\mathcal{F}$, weakly representable. Then there is an a such that

$$n \in \overline{F} \;\Leftrightarrow \vdash_{\mathcal{F}} \psi_a(\overline{n}).$$

For $n = a$ we get a contradiction.

Let $\{\varphi_n\}_{n\in\omega}$ be a recursive enumeration of the formulas in the language of $\mathcal{F}$ without free variables. If $\mathcal{F}$ is axiomatizable formal system then, by definition, the set of its theorems is an r.e. set. If $\mathcal{F}$ were complete then we would know that either a sentence is a theorem, or its negation is. But then $\mathcal{F}$ would be decidable: to know whether a sentence φ is a theorem, generate the theorems until either the sentence or its negation appear.[5] $\square$

The proof of undecidability uses a formula ψ_a that is represents $\overline{F}$: thus $\psi_a(\overline{n})$ says that $n \in \overline{F}$, i.e. that $\psi_n(\overline{n})$ is not provable in $\mathcal{F}$. For $n = a$, $\psi_a(\overline{a})$ thus says that $\psi_a(\overline{a})$ (i.e. itself) is not provable in $\mathcal{F}$. In other words, the meaning of $\psi_a(\overline{a})$ is 'I am not provable in $\mathcal{F}$'.

As far as incompleteness is concerned, the proof given in 3.1 is indirect, and does not explicitly exhibit undecidable sentences, which are neither provable nor disprovable. On the other hand, even the sentences obtained by other proofs (using self-referential diagonalization) are regarded as somewhat artificial, from a mathematical point of view.[6]

[4] The recursive predicates are obviously closed under negation: if c_R is the characteristic function of R, then $\delta(c_R(x), 0)$ is the characteristic function of $\neg R$.

[5] This is a recursive operation, since it can be written as

$$\mu x.\,(\varphi_{f(x)} = \varphi \;\vee\; \varphi_{f(x)} = \neg\varphi).$$

And the recursive predicates are closed under disjunction: the function

$$d(x,y) = \delta(\delta(x+y,0),0)$$

is 0 exactly when the two arguments x and y are 0, i.e. when $x+y=0$. Thus, if c_R and c_S are the characteristic functions of R and S, then $d(c_R(x), c_S(x))$ is the characteristic function of $R \vee S$.

[6] Great efforts have been made to obtain less artificial undecidable sentences, for various systems in common use: two extreme and classical examples are the **Continuum Hy-**

The fatal consequences for formalism embodied in Theorem 3.1 can be expressed as follows: *any formal system is inadequate, being either inconsistent, or undecidable (and hence, if formal, also incomplete), or not sufficiently strong* (to prove at least the elementary arithmetical facts expressed by the axioms of $\mathcal{R}$).

While the formulation of 3.1 allows the treatment of mathematical theories (for example, as noted at the end of Section 2, of usual arithmetical or set-theoretical formal systems), it does not directly cover purely logical systems like the predicate calculus. But at this stage a simple observation will do the job.

Theorem 3.2 (Church [1936a], Turing [1936]) *The Predicate Calculus is undecidable.*

Proof. Let $\mathcal{Q}$ be Robinson's Arithmetic: since $\mathcal{Q}$ is a consistent extension of $\mathcal{R}$, it is undecidable. Let ψ be the conjunction of its axioms (recall, and this is the crucial point, that $\mathcal{Q}$ is finitely axiomatized). By the Deduction Theorem of Predicate Calculus, $\vdash_{\mathcal{R}} \varphi$ holds if and only if $\psi \to \varphi$ is provable in Predicate Calculus, and thus any decision procedure for this would give one for $\mathcal{Q}$, contradicting 3.1. $\square$

Bibliography

Church, A.

[1936] An unsolvable problem of elementary number theory, *Am. J. Math.* 58 (1936) 345–363, also in Davis [1965], pp. 89–107.

[1936a] A note on the Entscheidungsproblem, *J. Symb. Log.* 1 (1936) 40–41, also in Davis [1965], pp. 110–115.

Cohen, P.J.

[1963] The independence of the continuum hypothesis, *Proc. Nat. Acad. Sci.* 50 (1963) 1143–1148.

Davis, M.

[1965] *The undecidable*, Raven Press, 1965.

Gödel, K.

[1931] Über formal unentscheidbare Sätze der Principia mathematica und verwandter Systeme I, *Monash. Math. Phys.* 38 (1931) 173–198, transl.

pothesis for Set Theory (Gödel [1938], Cohen [1963]), and a finite version of **Ramsey's Theorem** for Peano Arithmetic (Paris and Harrington [1977]).

in Davis [1965], pp. 5–38.

[1934] On undecidable propositions of formal mathematical systems, *mimeographed notes*, 1934, in Davis [1965], pp. 41–81.

[1936] Über die Lange der Beweise, *Ergeb. Math. Koll.* 7 (1936) 23–24, transl. in Davis [1965], pp. 82–83.

[1938] The consistency of the axiom of choice and of the generalized continuum hypothesis, *Proc. Nat. Acad. Sci.* 24 (1938) 556–557.

Hofstadter, D.R.

[1979] *Gödel, Escher, Bach: an Eternal Golden Braid*, Basic Books, 1979.

Kleene, S.K.

[1936] General recursive functions of natural numbers, *Math. Ann.* 112 (1936) 727–742, also in Davis [1965], pp. 237–252.

[1936a] A note on recursive functions, *Bull. Am. Math. Soc.* 42 (1936) 544–546.

Mostowski, A.

[1947] On definable sets of positive integers, *Fund. Math.* 34 (1947) 81–112.

Nagel, E., and Newman, J.R.

[1958] *Gödel's proof*, New York University Press, 1958.

Odifreddi, P.G.

[1989] *Classical recursion theory*, North Holland, 1989.

Paris, J.B., and Harrington, L.

[1977] A mathematical incompleteness in $\mathcal{PA}$, in *Handbook of Mathematical Logic*, Barwise ed., North Holland, 1977, pp. 1133–1142.

Post, E.L.

[1922] Absolutely unsolvable problems and relatively undecidable propositions. Account of an anticipation, in Davis [1965], pp. 340–433.

Robinson, R.M.

[1950] An essentially undecidable axiom system, *Proc. Int. Congr. Math.* (1950) 729–730.

Rosser, B.J.

[1936] Extensions of some theorems of Gödel and Church, *J. Symb. Log.* 1 (1936) 87–91, also in Davis [1965], pp. 230–235.

Smielew, W., and Tarski, A.

[1950] Mutual interpretability of some essentially undecidable theories, *Proc. Int. Congr. Math.* (1950) 734.

Smullyan, R.M.

[1961] *Theory of formal systems*, Princeton University Press, 1961.

[1987] *Forever undecided: a puzzle guide to Gödel*, Knopf, 1987.

Tarski, A.

[1931] Sur les ensembles définissables de nombres réels, *Fund. Math.* 17 (1931) 210–239.

Tarski, A., Mostowski, A., and Robinson, R.M.

[1953] *Undecidable theories*, North Holland, 1953.

Turing, A.M.

[1936] On computable numbers with an application to the Entscheidungsproblem, *Proc. London Math. Soc.* 42 (1936) 230–265, corrections ibid. 43 (1937) 544–546, also in Davis [1965], pp. 116–154.