Chapter 8
The Implications of Gödel's Theorem

8.1 *Pons Asinorum*

Gödel's theorem holds for any **first**-order theory that formalises the ordinary arithmetic of the natural numbers—in particular it holds of first-order Peano Arithmetic—and proves that it is **incomplete** and **undecidable**.

Gödel's theorem is one of a cluster of incompleteness theorems—Tarski's theorem, Church's theorem, Turing's theorem, the Kleene-Church theorem, Löb's theorem—which have unexpected and profound implications both for mathematics in particular and for our understanding of reason and reality in general.

Gödel's proof is difficult. It represents the *Pons Asinorum* of mathematical logic. It can be mastered only by careful study and much rumination. There are many expositions, none of them easy, few of them altogether satisfactory. Different readers approach the proof at different levels of sophistication, and with different demands for being shown the wood as a whole, or the trees in detail. In the next four sections I shall attempt to give an overview, showing the general strategy, and indicating areas of difficulty, but not properly proving that they can be overcome. Many readers may find some other exposition more helpful.[1]

[1] For a good technical exposition, see George S. Boolos and Richard C. Jef-

8.2 The Flavour of the Gödelian Argument

Gödel's theorem is a variant of the Epimenides, or Liar, Paradox. Consider the statement 'This statement is untrue': if it is true, then it is untrue; and if it is untrue, then it is true. Normally we are not greatly bothered by such a statement. We naturally ask 'What statement?', remembering that Russell was able to escape his paradox by deeming self-referential class specifications meaningless,[2] and refuse to accept that a statement has been made until the statement being referred to has itself been completed.[3] But Gödel has blocked this escape route, because, granted the resources of *Sorites* Arithmetic, he was able to devise a means whereby we can code well-formed formulae into numbers, **and** some properties of, and relations between, well-formed formulae into properties of, and relations between, numbers. In the case of the Liar Paradox it is not the 'this' but the 'true' that has to go: Tarski's theorem says that a formal logistic theory cannot contain a predicate possessing the properties of our 'true', for if it did, it would lead to a straight contradiction. Gödel's theorem, however, avoids paradox in a different way: it replaces Tarski's 'true' by the more definite and down-to-earth 'provable-in-the-system': 'provable-in-the-system' *can* be formalised, but since it is not the same as 'true', a well-formed formula can be not provable-in-the-system and still true. Indeed, if we think about it, we see that it must be. The formula cannot be provable-in-the-system, for if it were, then, being provable-in-the-system, it would be true, and being true, what it asserted would be the case, that is, that it was not provable-

frey, *Computability and Logic*, 2nd ed., Cambridge, 1980, 3rd ed., Cambridge, 1989, chs. 14 and 15. A very full exposition, with concessions to the non-numerate is given by James R. Newman and Ernest Nagel, *Gödel's Proof*, London, 1959. P.J. FitzPatrick, "To Gödel via Babel", *Mind*, 1966, pp.332-350, gives the best account of the translation problem, which I shall deal with in §8.4. A.W.Moore, *The Infinite*, London, 1990, ch.12, §2, pp.174-178, gives a fresh approach. See also J. N. Crossley and others, *What Is Mathematical Logic?*, Oxford, 1972, ch.5; Hao Wang, *Reflections on Gödel*, MIT, 1987; S. G. Shanker, *Gödel's Theorem in Focus*, London, 1988; Raymond Smullyan, *Forever Undecided: A Puzzled Guide to Gödel*, Oxford, 1988.

[2] See below, §12.2, §12.3.

[3] See Gilbert Ryle, "Heterologicality", *Analysis*, XI, 1950-51, pp.67-68.

in-the-system, which would be a contradiction. Equally, it cannot be false, for if it were false, what it asserted, namely that it was unprovable-in-the-system, would not be the case, and so if it were false, it would be provable-in-the-system, and hence true, which again would be a contradiction.

There are three crucial themes in this argument: reference, translation and diagonalization. These will be discussed in the next three sections.

8.3 Gödel Numbering

Gödel was able to block the ban on self-reference by his scheme of Gödel numbering, which enabled him to code every well-formed formula into a natural number, and thus to refer to it indirectly. Gödel's own system of Gödel numbering is easy to understand, though difficult to operate.[4] Every natural number can be factorised in terms of its prime factors. Thus $360 = 2^3 \times 3^2 \times 5^1$, which we could express succinctly as $\{3, 2, 1\}$. Gödel goes the other way. If we have a formula, and can give each type of symbol used a natural number, we can code the formula as a sequence of numbers, which in turn can be coded uniquely as the natural number that results from taking these numbers as indices of prime factors in their natural ordering. Peano Arithmetic is expressed by means of a finite number of symbols. Typically we need the brackets, the sentential connectives, the quotifiers, a stock of propositional variables, of individual variables, of predicates (including dyadic predicates for expressing relations), one particular individual—nought—and one particular function—the successor function, S. These can be arranged in an order, and can be correlated with the odd prime numbers, $3, 5, 7, 11, 13,$ *etc*.

(	)	$\rightarrow$	$\neg$	A	p	x	F	$'$	0	S
3	5	7	11	13	17	19	23	29	31	37

Any formula, whether or not it is well formed, is a string of symbols each of which can be represented by some prime number or another. Each such string can then be expressed as a single (though astronomically large) number, by forming the product of the odd prime factors in order, each being raised to the power of

[4] There are many others, which make the mathematical working easier, but the general scheme more difficult to grasp.

the number assigned to the symbol that comes in that order in the string. Thus, the formula

$$\neg\neg p \rightarrow p,$$

which has five symbols, and so could be expressed with the aid of the first five odd prime numbers, namely $3, 5, 7, 11, 13$, and would be expressed by the natural number

$$3^{11} \times 5^{11} \times 7^{17} \times 11^{7} \times 13^{17}$$

Such a number is enormous, but it is evident that every string of symbols can be thus expressed as a single number. We might reasonably hope that it would prove possible to express the property of a string of symbols being a well-formed formula as a corresponding property of the Gödel number of that string. This indeed is possible. Moreover, Gödel showed that it was possible to express the rule of substitution as a relation between the Gödel numbers of the two well-formed formulae involved, and also the rule of generalisation, and again the rule of *Modus Ponens* as a three-term relation between the Gödel numbers of the premises and the Gödel number of the conclusion. Clearly we can specify the Gödel number of each axiom of some given system.

Gödel devised one other coding facility. He needed to code not only strings of symbols but strings of strings of symbols, in order to be able to handle *sequences* of well-formed formulae, and thus to characterize *proof sequences*. We can conveniently do this by reserving the factor 2 for Gödel numbers of sequences of Gödel numbers, themselves the Gödel numbers of (invariably well-formed) formulae. Thus suppose it turned out that the Gödel number of $\neg\neg p \rightarrow p$ was 999 (of course, in fact it is much, much larger than that), and we had a proof of

$$\neg\neg\neg\neg p \rightarrow p,$$

whose own Gödel number was 1729 (again, it is much, much larger than that), the first line of which was $\neg\neg p \rightarrow p$ as an axiom, then the proof sequence would have as its Gödel number a number beginning $2 \times 3^{999} \times \ldots$ and ending, say, with 37^{1729}. With this notation, Gödel enables us to refer to sequences in general, and to express the property that a Gödel number must have if it is to be the Gödel number of a sequence that is a **proof** sequence. Granted that, it is relatively easy to define a two-term predicate of natural numbers that holds between two Gödel numbers when, but only when, the first is the Gödel number of a proof sequence whose last line is a well-formed formula whose Gödel number is the second number of the pair.

> An imaginary, and much simplified case:
>
> the Gödel number of the proof sequence is $2 \times 3^{999} \times 5^{1025} \ldots 37^{1729}$,
>
> the proof predicate would be $Pr(2 \times 3^{999} \times 5^{1025} \ldots 37^{1729}, 1729)$.

If we can carry through the coding so as to assign to every well-formed formula a Gödel number, **and** represent all the syntactical properties of well-formed formulae by means of arithmetical predicates—and it is a big IF—then we have found a way round the ban on self-reference, which was invoked to fault the simple formulations of the Liar Paradox, and can try to reformulate it in objection-proof terms.

8.4 Translation

It is easy to see that Gödel numbering enables us to **refer**, difficult to prove that it can **represent**, or **translate** the meta-logical features needed for Gödel's argument. As we saw in §8.2, Tarski's theorem shows 'true' is untranslatable into any formal system. The difference between the sentence uttered by the Liar and that constructed by Gödel is that the former is 'This statement is un**true**' whereas the latter is a coded way of saying 'This statement is un**provable**'. The former would lead to a straight contradiction IF 'true' could be expressed in terms of properties of, or relations between, Gödel numbers, and thus shows that 'true' cannot be so expressed: the latter, however, can be expressed, provided we confine 'provable' to what can be proved **within a specified formal system**, whose formation rules and rules of inference have been antecedently laid down; in that case, we avoid inconsistency by acknowledging a distinction between wide-ranging open-ended truth on the one hand and closely-defined and tightly-tied-down provability-within-a-given-formal-system on the other.

The 64,000-byte question is

Is such a translation feasible?

Gödel showed that it was. Granted the resources of a suitably formalised version of *Sorites* Arithmetic (SA), in which there was a first natural number and the successor function, and in which addition and multiplication, were defined by **recursion**.[5] The box below gives these two recursive definitions.

Recursive Definitions of Addition and Multiplication

Addition	Multiplication
$n + 0 = n$	$n \times 1 = n$
$n + Sm = S(n + m)$	$n \times Sm = (n \times m) + m$

These are **Primitive** Recursive Functions: they are functions which can be calculated by a computer. Other functions, such as exponentiation and factorial, which can be generated by certain legitimate means from these, are also Primitive Recursive functions. The Theory of Recursive Functions is complex and finicky, and of some importance for computer scientists. The over-all moral is important, but the details need not concern us here. It is helpful to compare Recursive Function Theory with an axiomatic treatment of Propositional Calculus, with the initial stock of primitive recursive functions being like the axioms, and the rules for generating new primitive recursive functions as the rules of inference. Gödel defined Primitive Recursive **predicates** and Primitive Recursive **relations** similarly. Almost all predicates and relations needed for translating important syntactical properties of Peano Arithmetic into properties of Gödel numbers are Primitive Recursive.

[5] Modern treatments use Robinson's Arithmetic Q; Q is "distinguished by the simplicity and clear mathematical content of its axioms" (G.S. Boolos and R..C. Jeffrey, *Computability and Logic*, 2nd ed., Cambridge, 1980, 3rd ed., Cambridge, 1989, p.158, quoting Tarski, Mostowski and Robinson). Its chief merit is that, although in other respects a weak language, it is tailor-made to have all recursive functions representable in it.

Clearly, if multiplication is, then being a prime number is; and if exponentiation is, then being the product of prime factors raised to various powers must be also. We have an algorithm—a mechanical rule—for telling whether a string of symbols is a well-formed formula, so it is likely that we should be able to translate it into a test for telling whether a natural number was the Gödel number of a well-formed formula. It is not obvious that the three rules of inference—the rule of substitution, the rule of generalisation and the rule of *Modus Ponens*—can be represented by Primitive Recursive Relations between the Gödel numbers of the premises and the Gödel numbers of the conclusions, but if we follow all Gödel's detailed working, we shall be convinced that they can. If that be granted, then it is obvious that Gödel numbers of the axioms can be specified individually, so that we could check whether a sequence of Gödel numbers was such that every one was either a Gödel number of one of the axioms that have been individually specified, or a Gödel number of a substitution-instance of an earlier one, or a Gödel number of a generalisation of an earlier one, or the Gödel number of a well-formed formula that follows by *Modus Ponens* from two well-formed formulae whose Gödel numbers come earlier in the sequence. Again, it is a laborious task to check in detail that this can be done, and that given some **sequence** with a Gödel number $2 \times 3^n \times 5^m \times 7^l \ldots$ the property of being a **proof** sequence is Primitive Recursive. But it can be done, and it is reasonable to take the word of competent mathematicians that it has been done. This much granted, the last steps are relatively easy. The property of being the largest prime factor of a given (even) number is clearly Primitive Recursive, as is that of being the exponent of the largest prime factor of a given (even) number. Hence the relation $Pr(x, y)$ is Primitive Recursive, where $Pr(x, y)$ is the relation that obtains between x and y, where x is the Gödel number of a proof sequence having as its last member the well-formed formula whose Gödel number is y; that is to say x is the Gödel number of a "proof" of y.

We have thus advanced the argument of the last section, and shown that Gödel numbering not only enables us to refer to well-formed formulae and sequences of well-formed formulae, but enables us also to represent as a Primitive Recursive relation the crucial syntactic relation of being a proof of a particular well-formed formula within a specified formal system.

To go to the more general provability-within-a-specified-formal-system, we need to add a quotifier, $(\mathrm{V}x)$, to obtain $(\mathrm{V}x)Pr(x,y)$; and then the predicate is no longer **Primitive** Recursive, on account of the unbounded range of $(\mathrm{V}x)$, though it retains a general sort of recursiveness, in as much as IF it does hold in a particular case, a computer, by working through each value of x in turn, would sooner or later find the value of x for which it did hold, and hence establish the fact that it did. For the sake of definiteness we work not with the wide-ranging $(\mathrm{V}x)$, but a μ-operator, the **least** x such that .... Such a predicate is called **General** Recursive; General Recursive Relations and General Recursive Functions are defined similarly. Often the 'General' is omitted, and recursive predicates/relations/functions are those that can be defined by some recursion schema, together with some innocuous operations such as substitution, and the use of a "μ-operator", which picks out the first number, if there is one, that satisfies the specified predicate, relation, or function.

We have thus vindicated the claim that thanks to Gödel numbering we can not only refer to well-formed formulae, but characterize meta-logical truths about them in arithmetical terms. Once we are assured of the adequacy of the translation—of the representability of the proof-predicate in computer-speak (that is, in some suitable formalisation of *Sorites* Arithmetic)—Gödel's argument is under way.

8.5 Diagonalization

Gödel looks like being able to refer to well-formed formulae without having actually to quote them, and to express within a formal system the meta-logical property of being unprovable-in-that-system. But it is not evident that he can actually construct a Gödelian formula G with just the tricky internal structure needed for the argument. Might he not be forever trying to tread on his own tail?

Gödel is in fact able to construct a well-formed formula with the necessary properties by means of a Diagonalization procedure that is similar in spirit to Cantor's proof that there are more real numbers than there are natural numbers:[6] the details are a mathematical *tour de force*, but not philosophically difficult. What is philosophically difficult is the integration of these manoeuvres into one coherent argument, making use of the peculiar features of G.

[6] See above, §7.2.

REFERRING	Gödel Numbering
TRANSLATION	Recursive Function Theory
SELF-REFERENCE	Diagonalization

G has to be regarded both as a formula in a calculus and as a proposition with content, making a statement about a natural number. And we also need to be able to talk about G, making precise meta-logical statements about its status—whether it is well-formed, whether it is provable within a tightly specified formal system. It is difficult to keep track of these shifts of standpoint. One useful aid is to follow a device of FitzPatrick,[7] and distinguish the different uses of symbols by using different natural languages: Latin; French; and English.

Latin is the formal system we are interested in, in this case Peano Arithmetic. (Latin is a dead language, which we study but do not normally use, though we can: a formal system may be studied as an uninterpreted calculus, though it can be interpreted, and used to express content-ful propositions.)

French is a meta-language for meta-logic. (French is a living language, which we do not just study, but *use* to make very precise statements *e.g.* about Latin syntax.)

English is for real. We make true statements in it, and can (sometimes) tell that a statement is true (*e.g.* 'Two and two makes four'). Moreover, as English linguistic chauvinists like to make out, and Frenchmen deny, English is at least as rich as French: every statement in French can be translated into English. Or, more mathematically, English must contain the whole of *Sorites* Arithmetic (SA). (Of course, it may contain more; if the system under study in Latin is Peano Arithmetic (PA), it will need to contain that; in which case, since *Sorites* Arithmetic (SA) is contained within Peano Arithmetic (PA), it will certainly be rich enough. If English is rich enough to translate Latin adequately, it is rich enough to translate French.)

Gödel was able to construct a well-formed formula, G, and considered it as a well-formed formula of a system of first-order Peano Arithmetic (PA)—a Latin sentence. We may ask, meta-logically,

[7] See P.J. FitzPatrick, "To Gödel via Babel", *Mind*, 1966, pp.332-50.

> Gödel needs to view the Gödelian sentence in three ways:
> 1. as a string of symbols—a formula—being used according to certain specified syntactic rules, but with no interpretation attached.
> 2. as making a statement in a meta-language about formulae.
> 3. as making a statement about natural numbers, like 'Five and seven make twelve' (only, much more complicated).

whether or not this string of symbols is derivable as a theorem of Peano Arithmetic. It would be like asking a French classicist whether a string of Latin words was good Latin or not. We can also consider this string of symbols as a meaningful expression, and ask whether or not it is true. It would be like asking a historian whether a statement in a letter by Cicero was true or not. We have, in effect, translated from Latin into English, and assessed the truth of the proposition expressed in the English translation. If we ask the two questions, the answers are, surprisingly, different: if we ask whether G is true, the answer turns out to be Yes; but if we pose the question in French, *Est la formule G dérivable en l'Arithmetique de Peano*? the answer turns out to be *Non.*

$$PA \nvdash G.$$

La formule G ne peut etre deduit en l'Arithmétique de Peano.

The reason for the answers' being different emerges if we translate again from French into English, which we can do, thanks to Gödel numbering, so that instead of being a French statement about Latin syntax, it is an English statement about a certain natural number, namely: g, the Gödel number of G, to the effect that there is no number x, such that x is even (i.e. x is the Gödel number of a **sequence**), x has the proof property (enormously complicated to spell out in simple arithmetical terms), and the index of the largest prime factor of x is g. This is a statement in English, and since it is a statement about natural numbers, and their properties, and the relations between them, expressed in *Sorites* Arithmetic (SA), it can be translated into Latin, that is to say, the formal language of Peano Arithmetic, which contains *Sorites* Arithmetic (SA). If we do this, we find that, because of the peculiar way Gödel has constructed G, the Latin translation turns out to be G itself. We can express this in English (and hence also in Latin, if we *use* it, instead of merely talking about it), using the semantic turnstile,

$$\models G$$

as a claim of ordinary **informal** arithmetic that G is true.

Gödel has constructed G to be of the form $\neg(\mathrm{V}x)(Pr(x,g))$; where g is the Gödel number of G. If we ask French questions about it, we have to say *La formule G ne peut etre deduit en l'Arithmétique de Peano*, or, in symbols, $PA \nvdash G$, but if we translate it into English, and ask if it is true, we have to say it is, $\models G$, since any other view would, as we argued earlier, lead to a contradiction. Hence, provided Peano Arithmetic is consistent, G, though true, is not provable in Peano Arithmetic. So Peano Arithmetic is **not semantically complete**.

Moreover, a further argument shows that $\neg G$ is not derivable as a theorem of Peano Arithmetic: *la formule ¬G ne peut etre deduit en l'Arithmétique de Peano,* or, in symbols, $PA \nvdash \neg G$. It follows that we could add G, not itself already a theorem, to Peano Arithmetic without inconsistency. Hence Peano Arithmetic is **not syntactically complete**; and also, since G can be neither proved nor disproved in it, Peano Arithmetic is **undecidable**.

Latin	G?
French about Latin	$PA \nvdash G$
French about Latin	*La formule G ne peut etre deduit en l'Arithmétique de Peano*
English	$\neg(\mathrm{V}x)Pr(x,g)$
Latin in use or English	$\models G$

The crucial step is the translation from French into English, that is from meta-logical propositions about well-formed formulae into arithmetical propositions about natural numbers, ascribing to the number g a complex predicate of the form $\neg(\mathrm{V}x)(Pr(x,g))$. The predicate $Pr(x,g)$ is defined, very complicatedly, in terms of *properties of natural numbers* and *relations between natural numbers*. It is a simple question of arithmetical fact whether or not a particular number has this property. And then in the particular case of whether g has the property, it turns out, thanks to the way G was constructed, that the (English) translation of (the French) $PA \nvdash G$, i.e. $\neg(\mathrm{V}x)(Pr(x,g))$, is just the proposition G in Latin, considered as a true statement of arithmetical fact. So G is True, i.e. $\models G$, iff $\neg(\mathrm{V}x)(Pr(x,g))$ iff $PA \nvdash G$. The syntactic relation $PA \nvdash G$ is viewed **meta**-logically as a true assertion about PA, which translates, via Gödel numbering, into a true assertion about a particular number, g, namely the assertion G, where G is a fully interpreted statement about ordinary natural numbers, independently of any formalisation of arithmetic. We can not only use French to talk about Latin, but translate from French into English, with all the important French terms having adequate representations in English. *La formule G ne peut etre deduit en l'Arithmétique de Peano* translates into $\neg(\mathrm{V}x)Pr(x,g)$, i.e. G is PA-unprovable, and if this were false in English, the French *La formule G n'est pas dérivable de Peano Arithmétique* would have to be false too, so that *La formule G est dérivable en l'Arithmétique de Peano* and $\vdash_{PA} G$, with the consequence that in Peano Arithmetic we could derive false propositions, and Peano Arithmetic would be unsound and inconsistent.

8.6 Conditions

Gödel's proof works only under certain conditions. It applies only to consistent systems.[8] If a system is inconsistent, then we can prove both A and $\neg A$, and from these two premises we can, granted some standard rules and theses of propositional calculus, derive any well-formed formula whatsoever, and so, in particular, the Gödelian formula G. Gödel's proof applies only to formal systems, systems, that is, which have fully spelt-out formation rules, axioms and rules of inference. They must be not only fully, but *finitely*, specified, so that after a finite number of steps it will be clear whether or not a sequence of symbols is well formed, whether or not a well-formed formula is an axiom, whether or not a well-formed formula follows from one or two others in virtue of a particular rule of inference.

[8] Originally to prove $PA \nvdash \neg G$ Gödel needed ω-consistency, but Rosser improved Gödel's proof so as to need only ordinary consistency. See J.B.Rosser, "Extensions of some theorems of Gödel and Church", *Journal of Symbolic Logic*, **1**, 1936, pp.87-91. See also further below, §8.11.

It is essential also, if a theory is to contain the means for some form of Gödel numbering, that it should include the natural numbers together with the operations of addition **and** multiplication. Peano Arithmetic and *Sorites* Arithmetic contain both because they contain the successor function, and addition and multiplication can be implicitly defined, as we have seen, by suitable axioms involving just the successor. Rather surprisingly, neither addition by itself nor multiplication by itself is enough. Presburger has shown that if we have only addition but not multiplication, the theory *is* complete and decidable. Similarly Skolem showed that if we have only multiplication but not addition, the theory is, again, complete and decidable. Even more surprisingly Tarski has shown that geometry and the elementary algebra of real numbers, *with* addition and multiplication, but *without* the general notion of natural number, is complete and decidable.[9] These results are surprising, and call for some explanation. The reason why Gödel's theorem does not apply is that in these theories we are thinking of the numbers in some other way—*e.g.* as an infinite Abelian group with addition as the composition function: they are not *counting* numbers, an endless succession of individually distinct entities, capable of being used for coding without ever being in any danger of running out. Once we have the natural numbers as counting numbers, we are subject to Gödel's theorem, because there is a potential infinity of natural numbers, and Dedekind-infinite sets can be correlated with a proper subset of themselves. But we need only a potential, not an actual, infinity of natural numbers. Gödel's theorem holds not only of Peano Arithmetic, but of the simpler *Sorites* Arithmetic. Essentially, that is, we can address Gödel's argument to a computer. We do not need to invoke the hand-waving arguments of Chapter Six to avail ourselves of the principle of recursive reasoning, in order to prove Gödel's theorem: it applies even within

[9] M.Presburger, "Über die Vollständigkeit eines gewissen Systems der Arithmetik ganzer Zahlen, in welchem die Addition als einzige Operation hervortritt", *Sparawozdanie z I Kongresu matematyków krajów slowianskich, Warszawa, 1929,* Warsaw, 1930, pp.92-101, 395. T.Skolem, "Über einige Satzfunktionem in der Arithmetik", *Skrifter Utgitt av Det Norske Videnskaps-Akademi i Oslo*, I. Mat.-Naturv. Klasse, 1930, no. 7, Oslo, 1931. These and other results are conveniently summarised by Geoffrey Hunter, *Metalogic*, London, 1971, p.260.

the austere limitations of *Sorites* Arithmetic, with no commitment to any actual infinity.

Although Gödel's theorem does not hold for unnaturally weak formal systems, it does hold for all stronger first-order ones, no matter how much we strengthen them. It seems at first surprising: a stronger system, surely, should be more decidable and more complete. But, as we shall see with first- and second-order logic,[10] it does not follow, for as we strengthen a system, we increase not only its resources for answering questions, but the number and type of question that may be asked of it. Contrariwise, however, the restriction to first-order logic *is* essential. If we strengthen logic on that front, we can no longer prove Gödel's theorem. One of the consequences of Gödel's theorem is that *second*-order logic is *not* finitely axiomatizable, not the sort of logic a computer can be programmed to do.

8.7 Corollaries and Consequences

Gödel's theorem yields many further results, some of which can be argued for independently, but which are conveniently seen as corollaries and consequences of Gödel's theorem. They are summarised below.

Corollaries

1. Peano Arithmetic is not syntactically complete
2. Peano Arithmetic is not negation-complete
3. Peano Arithmetic is not decidable
4. Peano Arithmetic is not semantically complete

Further Consequences

5. Peano Arithmetic is non-monomorphic.
6. Tarski's theorem
7. Church's theorem
8. Turing's theorem
9. Gödel's second theorem

Some of these results need little discussion: others need amplification in the next two sections. Gödel's theorem has some immediate corollaries. It shows first that Peano Arithmetic (and

[10] See below, §13.5.

Sorites Arithmetic) is not **syntactically complete**. For we can add G, which is not a theorem of Peano Arithmetic, to PA, the conjunction of the axioms, without making $PA + G$ inconsistent It shows similarly that Peano Arithmetic is not **negation-complete**. For neither G nor $\neg G$ **are** theorems of Peano Arithmetic. It shows thirdly that Peano Arithmetic is not **decidable** For we cannot **tell** within Peano Arithmetic **whether or not** G or $\neg G$ are theorems. And fourthly it shows that Peano Arithmetic is not **semantically complete**: for $PA \not\vdash G$ though $\models G$.

These immediate results are really just different ways of looking at Gödel's theorem. They give rise to further, less obvious, consequences. Since neither $\vdash_{PA} G$ nor $\vdash_{PA} \neg G$ neither $PA \cup \neg G$ nor $PA \cup G$ is inconsistent, and so each has a model. Hence, as we argued in Chapter Six,[11] PA has two models, in one of which G obtains, and in the other of which $\neg G$ obtains, so that the two models are not isomorphic with each other, and Peano Arithmetic is non-monomorphic.

It should be noted, however, that this result, though it follows from Gödel's theorem, does not depend on it. We could have proved it from the Compactness theorem for first-order logic. It is only that since we were going to prove Gödel's theorem anyhow, it was more economical to use it to give us this result also.[12]

In proving Gödel's theorem, we glimpsed also a proof of Tarski's theorem, which shows that we cannot represent within Peano Arithmetic any predicate that has the characteristic properties of 'true'. For then the Liar Paradox would be unfaultable, and we should have shown how we could always derive a contradiction within Peano Arithmetic, which is to say that Peano Arithmetic is inconsistent. So, provided Peano Arithmetic is consistent, 'true' cannot be represented within it.

[11] §6.3.

[12] See above, §6.3, n.6.

8.8 Church's Theorem and Turing's Theorem

Gödel's theorem yields Church's theorem, that first-order logic is undecidable; that is, not two-way decidable in the sense of §3.4. For SA has its own Gödelian formula; suppose that first-order logic were decidable: then the well-formed formula $SA \rightarrow G$ would be decidable, and so we could tell whether or not $SA \vdash G$, that is, whether or not $\vdash_{SA} G$, contrary to Gödel's theorem.

Church's theorem is surprising. We might naturally suppose that first-order logic, being complete, was also decidable. Indeed, with the aid of Gödel numbering, we can number not only all well-formed formulae, but all sequences of well-formed formulae that constitute valid proofs in first-order logic, and work through them one by one to see if they happen to be a valid proof of any particular well-formed formula we are interested in. In this way, given sufficient time, we shall find a proof, *if any exists*. We do have **one**-way decidability. What we lack is any way of telling that a particular well-formed formula of first-order logic is **not** a theorem. We could go on grinding out proof after proof of other well-formed formulae, in each case finding that it was not the proof we wanted, but never knowing that it was 100% certain that the next proof we looked at would not be the sought-after one. We lack a definitive test for **non**-theoremhood. That is not to say that we cannot often show of a particular well-formed formula, A say, that it is not a theorem. If we have already proved $\neg A$, then, since we know that first-order logic is consistent, we can be sure that there is no proof of A. But first-order logic is not negation-complete, though it is semantically complete. Even though every well-formed formula that is true under all interpretations is a theorem, there are many well-formed formulae, such as $(\mathrm{V}x)(\mathrm{V}y)(F(x) \wedge F(y) \wedge (x \neq y))$, which are true under some, but not under all, interpretations, so that neither that well-formed formula, since it does not hold in a one-membered universe, nor its negation, which does not hold in a two-membered universe, is a theorem.

Church's theorem forces us to distinguish one-way from two-way decidability, and the bounded and fixed-length versions of the latter, as in §3.4.

The discussion has been entirely in terms of well-formed formulae and theoremhood. It is easier to grasp the importance of what is at issue that way, but it is also, as we have now seen, easy to be misled. Mathematical logicians therefore move away from well-formed formulae to their Gödel numbers, and consider, abstractly, sets of numbers, where they call those sets of numbers that are two-way decidable "recursive", and those that are only one-way decidable "recursively enumerable". The Gödel numbers of theorems of propositional calculus are recursive; the Gödel numbers of theorems of first-order logic are recursively enumerable; and the Gödel numbers of non-theorems of first-order logic are not recursively enumerable; the Gödel numbers of theorems of Peano Arithmetic are recursively enumerable; and the Gödel numbers of non-theorems of Peano Arithmetic are not recursively enumerable.

The elucidation of decidability has been in terms of there being a *method* for deciding questions. Gödel's theorem gave rise to much discussion of what constituted a method, and the concept was elucidated in terms of a fool-proof method which a computer could be programmed to follow. Turing produced a specification of what it essentially was to be a computer, in his definition of a "Turing machine". A Turing machine can be identified with a program for an idealised computer. Notoriously, some programs do not work, but get into a loop, and just go on running without ever arriving at a result. It would be nice to have a program to check programs and eliminate all those that would not reach a conclusion. It would be nice, but we cannot have it. That is what Turing showed. Turing's theorem is an analogue of Gödel's: it states that there is no general procedure for deciding whether or not a given procedure will itself end in a definite decision, one way or the other. Note that Turing's theorem does not show that we cannot ever weed out dud programs—we can and do—but we cannot devise a *program* for weeding out them *all*. In this it closely resembles Church's theorem. Like Church's theorem too, it does not show that we cannot have a procedure for selecting all those programs which do work. We can—given enough time. Given enough time, we can run each program, and those that are such as sooner or later to arrive at a result sooner or later will. But at no time can we argue from the fact that a program has not been completed yet to the conclusion that it never will. And although we may have our suspicions that it never will, and in *some* cases may be able to spot a loop, we have no *method* which will always identify a program as non-terminating if in fact it is.

In our computer age many people find Turing's theorem more accessible than Gödel's, and prove Gödel's as a corollary of Turing's. But Gödel's raises further questions of truth and provability which are of great philosophical interest. I therefore concentrate on Gödel's theorem, and leave Turing's theorem, together with Löb's theorem, which is of considerable importance in mathematical logic,[13] and turn to the final consequence of Gödel's first theorem I shall discuss, namely Gödel's second theorem.

8.9 Gödel's Second Theorem

Gödel's second theorem highlights one crucial assumption made throughout the proof of his first theorem, that Peano Arithmetic is consistent. Granted that premise, Gödel had been able to argue in proving his first theorem not only that the Gödelian formula, G, was unprovable-in-Peano-Arithmetic, but that it was actually true. This argument was a strict deductive argument, and could be formalised in first-order logic, to yield a formal proof of G granted the consistency of Peano Arithmetic. The premise that a formal system is consistent *can* be formalised in a formal system: we cannot hope to escape *à la* Tarski by holding that consistency, like truth, cannot be expressed formally. After all, once we have expressed in a formal system what it is for a well-formed formula to be provable-in-that-system, we can say that there is no well-formed formula such that both it and its negation are provable; and indeed, it is sufficient to state simply that there is *some* well-formed formula which is not a theorem,

$$(\mathrm{V}a)(WFF(a) \wedge \neg(\mathrm{V}m)(Pr_{PA}(m,a))),$$

where a is the Gödel number of a formula A, and $WFF(a)$ expresses in arithmetical terms that A is well formed, and $Pr_{PA}(m,a)$ is the proof predicate of the formal system PA expressing the fact that m is the Gödel number of a proof in Peano Arithmetic of A. Let us abbreviate it as $Cons(PA)$. Then we can formalise the informal argument of the proof of Gödel's first theorem and say

$$\vdash_{PA} Cons(PA) \rightarrow G,$$

[13] For an account of Löb's theorem see George S. Boolos and Richard C. Jeffrey, *Computability and Logic*, 2nd ed., Cambridge, 1980, 3rd ed., Cambridge, 1989, ch. 16, esp. pp.187-188; see also ch.15 for an illuminating survey of other results, but seen as stemming from Turing's theorem rather than Gödel's.

and thus, IF we had a proof of $Cons(PA)$, that is, if

$$\vdash_{PA} Cons(PA),$$

we should, by *Modus Ponens*, have a direct proof

$$\vdash_{PA} G.$$

But this is the direct contradictory of Gödel's first theorem,

$$\nvdash_{PA} G.$$

So IF we had a proof of the consistency of Peano Arithmetic, we should end up with an inconsistency.[14] That by itself is not quite enough to prove that we cannot have a proof of the consistency of Peano Arithmetic—it could be the case that Peano Arithmetic *was* inconsistent, in which case, paradoxically there would be a proof of $Cons(PA)$, since there would be a proof of *every* well-formed formula, including $Cons(PA)$. What we have instead is the slightly weaker result that

IF Peano Arithmetic **is** consistent, then we cannot **prove that** it is, *within Peano Arithmetic*.

That is Gödel's second theorem.

It is important to note the italicised condition. Only formal proofs within Peano Arithmetic (and hence within any weaker system) are excluded. We can in fact prove that Peano Arithmetic is consistent by transfinite induction: this was done by Gentzen.[15] But the principle of transfinite induction is not one that is available in Peano Arithmetic itself. We may also be able to argue for the consistency of Peano Arithmetic informally, and to this, again, there is no objection. All that is excluded by Gödel's second theorem is there being a formal proof within Peano Arithmetic itself. That is, indeed, a remarkable result, but it does not mean that we must utterly despair of the consistency of arithmetic, or take up an attitude of resolute agnosticism to the question.

[14] In which case, paradoxically, we could prove every well-formed formula to be a theorem, including that asserting that PA was consistent. In §12.10 many axioms of Zermelo-Fraenkel set theory are reported to be proved consistent provided ZF is itself, but ZF can be proved consistent only if it is not!

[15] G.Gentzen, "Die Wiederspruchfreiheit der reinen Zahlentheorie", *Mathematische Annalen*, **112**, 1936, pp.493-565; tr. in M.E.Szabo, ed., *The Collected Papers of Gerhard Gentzen*, London and Amsterdam, 1969.

8.10 Mechanism

Gödel's and Turing's theorems establish certain limitations on formal systems and computer programs. Gödel's theorem goes further, and shows that although the Gödelian formula, G, was unprovable-in-the-system, it was none the less true. That seems to suggest that the mental powers of an ordinary mathematician go beyond the formal inferences of a formal system, and cannot be simulated by a computer, no matter how sophisticated its programming. Alternatively, we might argue directly from Tarski's theorem, that since truth cannot be adequately expressed in a formal system, and since we evidently possess the concept of truth, we cannot be just the embodiment of some formal system. The mechanist, who wishes to deny this, would be reduced to denying that we really have a concept of truth at all. Some are prepared to maintain just this, but then have difficulty in claiming that mechanism is true.

These arguments have been developed in detail, and have been accepted as cogent by some, and vehemently rejected by others.[16] Much of the controversy turns on what exactly is meant by a machine, and by mechanism. Often also the critics have misunderstood the structure of the argument. Just as Gödel does not produce an absolutely undecidable well-formed formula, but only, after having had a formal system specified, works out a well-formed formula which is unprovable-in-that-system, so the philosopher arguing against mechanism, does not produce something that no machine can be programmed to do, but only offers a scheme of disproof, whereby any claim to represent a mind by a particular machine can be shown to fail for that particular machine.

The issues raised by this controversy are of considerable philosophical interest generally—it makes quite a difference to determinist and materialist views of the mind, if no mechanist account of the mind can be adequate: but that discussion belongs to meta-

[16] The original article was "Minds, Machine and Gödel", *Philosophy*, **36**, 1961, pp.112-127; and the argument was set forth more fully in J.R.Lucas, *The Freedom of the Will*, Oxford, 1970. The most noteworthy recent supporters of the argument are Dale Jacquette, "Metamathematical Criteria for Minds and Machines", *Erkenntnis*, **27**, 1987, pp.1-16; and Roger Penrose, *The Emperor's New Mind*, Oxford, 1989, and *Shadows of the Mind*, Oxford, 1994.

physics or the philosophy of mind, and here we need to confine ourselves to questions bearing on the philosophy of mathematics. These are first and foremost the nature of proof, and, stemming from that, the nature of mathematical knowledge.[17]

8.11 Gödel's Theorem and Provability

Gödel's theorem shows that for any reasonable system, L, some well-formed formulae are true which are not provable-in-L. Not only are they true, but we can see that they are true. Nor is this seeing some *recherché* sense-experience: rather, it is the result of careful argument. Gödel's theorem is proved. An intelligent man can follow his proof, and be rationally convinced by it that the Gödelian formula, G, is true. In some sense of 'prove' G is proved to be true. Gödel's theorem thus suggests that mathematical truth outruns provability-in-any-formal-system, and that beyond provability-in-any-particular-formal-system there are further possibilities of proof.

In Chapter 6 we were led to extend *Sorites* Arithmetic by the rule of recursive reasoning, but the implication of Gödel's theorem is more profound. Gödel originally needed ω-consistency to prove $PA \nvdash \neg G$ on the strength of being able, for any m, to prove Gm. But we need to walk warily: if we were to allow ourselves to prove-in-the-system $\vdash (\mathrm{A}n)Gn$, we should land ourselves in inconsistency. We have to distinguish the formal $\vdash$ from the informal 'therefore' (which we symbolize as $p \;\;\Vdash q$ in §13.3). This is puzzling. "Why", the reader may ask, "was it all right in Chapter 6 to allow a general licence to infer from for EACH $m, \vdash_L Fm$, to $\vdash_L (\mathrm{A}n)Gn$, but not here to infer from for EACH $m, \vdash_L Gm$, to $\vdash_L (\mathrm{A}n)Gn$?" The answer is that the method of proof is different in the two cases. The premises in the former case, $\vdash_L Fm$, were proved for each successive m, by recursion, in a fairly standard

[17] Since the views expressed in this section are highly controversial, it is only fair to refer the reader to criticisms of the Gödelian Argument, which are listed by Chalmers at
http://www.artsci.wustl.edu/~philos/papers/chalmers.biblio.4.html
In particular there are many in *Journal of Behavioral and Brain Sciences*, **13:4**, 1990, and **16:3**, 1993. Judson C. Webb, *Mechanism, Mentalism and Metamathematics; An Essay on Finitism*, Dordrecht, 1980, is a sustained attack on the arguments given here.

way, like, though not *exactly* like, the proofs of $\vdash_L Ft$ for particular instances of t. But in the case of the Gödel formulae, although for each m, $\vdash_L \neg Pr(m, g)$, the proof is not the same in each case: it is not a straightforward standard generalisation, nor a recursive proof, but a global *Reductio ad Absurdum.* Working out the proof predicate will be different in each case, but we can be sure that in each case it will turn out not to be a proof of G.

Three stages of generalisation:

1. Standard generalisation;
   from $\vdash_L Ft$ infer $\vdash_L (\forall x)Fx$
   from $\vdash_L$ for ANY $\vdash_L$ ALL
2. Proof depends on case, but in a standard way;
   from for EVERY $m, \vdash_L Gm$ infer $\vdash_L (\forall n)Gn$;
   from $\vdash_L$ for EVERY $\vdash_L$ ALL
3. Proof depends on case, but in a non-standard way;
   from for EVERY $m, \vdash_L Gm$ infer that $\vdash Gm$ holds generally,
   but do not infer $\vdash_L (\forall n)Gn$;
   from $\vdash_L$ for EACH do not infer as a theorem of L, $\vdash_L$ ALL although the informal inference from EACH to ALL *is* valid.

Of course we could formalise the proof of Gödel's theorem, but only in *another* formal system, which would in turn have proofs which were not formalisable in it. However far we go in formalising Gödelian proofs, and adding them to formal systems, the new system will still be subject to the Gödelian style of argument, which will yield yet further well-formed formulae that cannot be derived in the new system, but can never the less be seen to be true.

The "incompletability of mathematics", as it may be called, is well expressed by Myhill.

> Gödel's argument establishes that there exist, for any correct formal system containing the arithmetic of natural numbers, correct inferences which cannot be carried out in that system.[18]

[18] J.Myhill, "Some Remarks on the Notion of Proof", *Journal of Philosophy*, LVII, 1960, p. 462.

It is of great significance. In the first place it tells against Formalism. We cannot give a satisfactory account of provability or mathematical truth in terms of the syntactic properties of any formal system adequate for arithmetic. Even *Sorites* Arithmetic is incomplete, and contains well-formed formulae that are unprovable-in-SA, but can be proved and seen to be true from outside *Sorites* Arithmetic. The whole programme of accounting for mathematical inference in formal, syntactic terms alone, loses credibility: and we feel impelled to look elsewhere for the inwardness of mathematical endeavour. Post argues,

> For the entire development should lead away from the purely formal as the ideal of a mathematical science, with a consequent return to postulates that are to be self-evident properties of the new meaningful mathematical science under consideration.[19]

Gödel's theorem has been taken to imply that we simply have an innate ability to recognise a valid proof when we see one, much as the Intuitionists claim. But what Gödel's theorem shows is only that the concept of proof cannot be completely formalised, not that it cannot be formalised at all. In fact, Gödel's theorem argues against Intuitionism. Intuitionists, as we saw,[20] lay great emphasis on mathematicians' activity in actually proving theorems, and construe truth in terms of provability, from which it would follow that the Principle of Bivalence—that every proposition is either true, or else false—is questionable. But once we recognise that truth outruns provability, the validity of the Principle of Bivalence can no longer be impugned on those grounds, and the intuitionist critique of classical logic loses its edge.

More generally, the fact that mathematical truth outruns provability within a formal system argues for the creativity of mathematical inference, and perhaps its objectivity, but poses problems about the logic of argument. Kant argued that moral actions and moral judgements must be universalisable, and it has seemed to many modern philosophers that this is a requirement of rationality

[19] E.L. Post, "Recursively Enumerable Sets of Positive Integers and their Decision Problem", *Bulletin of American Mathematical Society*, **50**, 1944, p.416, fn 100; quoted by Robin Gandy "The Confluence of Ideas in 1936", *The Universal Turing Machine*, Oxford, 1988, p.96.

[20] §7.4.

generally. But if every inference is valid only if it is in accordance with some maxim, it would seem that reason must be entirely rule-governed. The resolution of the difficulty lies in distinguishing the order of the quotifiers. The version of universalisability espoused by Kant and his modern successors, requires that there be some rule which every action (in the case of Kant) or every inference (in modern accounts) accords with:

there is a *Rule* such that for every *Inference*

(*Inference* accords with *Rule*),

(V *Rule*)(A *Inference*)(*Inference* accords with *Rule*),

and this requirement is one that Gödel's theorem shows to be not always satisfiable. All that rationality really requires is the weaker condition:

for every *Inference* there is a *Rule* such that

(*Inference* accords with *Rule*),

(A *Inference*)(V *Rule*)(*Inference* accords with *Rule*).

We cannot formulate in advance rules of inference that will cover every valid inference: but given an inference, we can detect the hitherto unformulated principle it exemplifies. Inferences are necessarily not essentially one-off, in the way that some choices are, and the particular configuration of individual objects in space. It is not that some particular Gödelian formula just happens to be true but unprovable: the argument that establishes its truth can be used to establish the truth of other well-formed formulae, and if some putative Gödelian formula does not have its truth established, we are entitled to ask "Why?", and have some difference pointed out to us, to distinguish its case from the ones where the argument does apply. Universalisability is a real requirement of rationality, but less rigid than commonly supposed, allowing us to take a more generous view of the nature of mathematical truth.[21]

[21] See, more fully, J.R.Lucas, "The Lesbian Rule", *Philosophy*, 1955, XXX, pp.195-213.

Gödel's Theorems show:

1. First-order Peano Arithmetic is not decidable
2. First-order Peano Arithmetic is not complete
3. First-order Logic is not decidable
4. First-order Peano Arithmetic has non-standard models (see §6.3)
5. Second-order Peano Arithmetic is not recursively axiomatizable
6. Second-order Logic is not recursively axiomatizable
7. First-order Peano Arithmetic cannot be proved consistent by the methods of First-order Peano Arithmetic alone
8. Truth outruns provability
9. Verificationist arguments against the Principle of Bivalence are invalid
10. However much proofs are formalised, there are further proofs not fully formalised, but evidently cogent
11. Hilbert's Programme cannot be carried through to a successful conclusion
12. Reason is creative
13. Synthetic *a priori* truths are possible