

Lecture XIII

Languages with a Recursively Enumerable but Nonrecursive Set of Formulae.

In dealing with formal languages, it is common to require that the set of formulae of the language be recursive. In practice, however, one hardly ever needs to use more than the fact that the set of formulae is r.e. In practice, one also hardly ever encounters languages with a recursively enumerable but nonrecursive set of formulae. However, there seems to be nothing in principle wrong with such languages, especially if one thinks, as e.g. Chomsky does, that to give a grammar for a language is to give a set of rules for generating the well-formed formulae, rather than to give a procedure for determining whether a given string of symbols is well-formed or not.

We can easily cook up a language with a non-recursive but r.e. set of formulae. For example, let S be any set which is r.e. but not recursive, and let L be the first-order language which contains no function symbols or constants and whose predicates are $\{P_1^n: n \in S\}$. L will be as required.

While this language is artificial, natural examples sometimes arise as well. In a system of Hilbert and Bernays, for example, there is, in addition to the usual logical symbols, an operator (ιy), such that $(\iota y)A(x_1, ..., x_n, y)$ denotes the unique y such that $A(x_1, ..., x_n, y)$ holds. Hilbert and Bernays thought that this really only makes sense if there is a unique y such that $A(x_1, ..., x_n, y)$ holds, so they stipulated that (ιy) could be introduced only through the rule

$$\frac{(x_1) ...(x_n)(\exists!y)A(x_1, ..., x_n, y)}{(x_1) ...(x_n)A(x_1, ..., x_n, (\iota y)A(x_1, ..., x_n, y))}$$

(where $(\exists!y)A(x1, ..., xn, y)$ means that there is a unique y such that $A(x1, ..., xn, y)$ holds, and is an abbreviation of $(\exists y)(A(x1, ..., xn, y) \wedge (z)(A(x1, ..., xn, z) \pounds z = y))$). A result of this policy is that the set of well-formed formulae of the language will in general be nonrecursive, though it will be r.e. Hilbert and Bernays were criticized on this point, though it is not clear why this is a ground for criticism.

In terms of our own formalism, we could stipulate that f_i^n be introduced when $(x_1)...(x_n)(\exists!y)A(x_1, ..., x_n, y)$ is a theorem, where i is a certain Gödel number of A, and add as a theorem $(x_1)...(x_n)A(x_1, ..., x_n, f_i^n(x_1, ..., x_n))$.

The S^m_n Theorem.

If R is a 2-place r.e. relation, then intuitively R_k should be r.e. as well; but furthermore, given k, we ought to be able to effectively find an index for R_k. This is indeed the case, and is a special case of the S^m_n theorem. More generally, let R be an m+n-place r.e. relation. In the case we have just considered, R_k is obtained from R by fixing k as a parameter; the general form of the S^m_n theorem (put informally) says that given m numbers k_1, ..., k_m, we can effectively find an index for the relation obtained from R by fixing k_1, ..., k_m as parameters. In our own formalism, the S^m_n theorem is an easy consequence of the definability in RE of substitution. We now state the S^1_1 theorem, i.e. the special case of the S^m_n theorem in which m = n = 1.

Theorem: For any 2-place r.e. relation R, there is a 1-1 recursive function ψ such that, for all k, $W_{\psi(k)} = R_k$.
Proof: Let e be an index for R. e is a Gödel number of some formula of RE $A(x_2,x_1)$ that defines R. An index of R_k, i.e. a Gödel number of a formula of RE that defines R_k, can be obtained from e via substitution. More specifically, we define the graph of ψ in RE by the formula $PS(k,y) =_{df.} (\exists p \leq y)(Num(p,k) \wedge (w<p) \sim Num(w,k) \wedge NSubst(\mathbf{0}^{(e)},y,[\mathbf{0}^{(1)},\mathbf{0}^{(2)}],p) \wedge (w<y)(\sim NSubst(\mathbf{0}^{(e)},w,[\mathbf{0}^{(1)},\mathbf{0}^{(2)}],p)))$ (the use of negation is legitimate, since the formulae it affects are equivalent to formulae of Lim). Informally, ψ assigns to k the least Gödel number of the formula obtained by substituting the least Gödel number of the numeral of k for x_2 in the formula with Gödel number e. The function thus defined is clearly 1-1, since the results of substituting different numerals for the same variable in the same formula must have different Gödel numbers.

The general form of the S^m_n theorem can be stated and proved similarly: for any m+n-place r.e. relation R, there is a 1-1 recursive function ψ such that, for all $k_1,...,k_m$, $W^{n+1}_{\psi(k1,...,km)} = R_{k1,...,km}$ (where $R_{k1,...,km}$ is $\{<y_1, ..., y_n>: R(k_1, ..., k_m, y_1, ..., y_n)\}$). As we see, the name "$S^m_n$ theorem" derives from the convention of taking m as the number of parameters and n as the number of other variables; 'S' probably stood for 'substitution' in the original conception of Kleene, to whom the theorem is due.

As a consequence of the above theorem, we have the following

Theorem: For all m and n, there is a one to one m+1-place recursive function ψ such that for all m+n-place r.e. relations R, if e is an index of R and k_1, ..., k_n are numbers, then $\psi(e, k_1, ..., k_m)$ is an index of $\{<y_1, ..., y_n>: R(k_1, ..., k_m, y_1, ..., y_n)\}$.
Proof: Apply the previous form of the S^m_n theorem to the relation W^{m+n+1}. That is, let ψ be a function such that $\psi(e, k_1, ..., k_m)$ is an index of $\{<y_1, ..., y_n>: W(e, k_1, ..., k_m, y_1, ..., y_n)\} = \{<y_1, ..., y_n>: R(k_1, ..., k_m, y_1, ..., y_n)\}$.

The second form of the S^m_n theorem can thus be seen as a special case of the first. The first form also follows directly from the second. A third form of the theorem is the standard form in most presentations of recursion theory, and the form originally proved by Kleene:

Theorem: For all m and n, there is a one to one m+1-place recursive function ψ such that if e is an index of an m+n-place partial recursive function ϕ, then $\psi(e, k_1, ..., k_m)$ is an index of the n-place function $\phi(k_1, ..., k_m, y_1, ..., y_n)$.
Proof: Apply the previous theorem to the relation $W^{m+n+2}*$, the graph of the m+n+1-place function Φ^{m+n+1}.

Given m,n, a function ψ with the property stated in the third version of the theorem (for m,n) is a function standardly called an S^m_n function.

The Uniform Effective Form of Gödel's Theorem.

We can use the S^m_n theorem to prove the uniform effective form of Gödel's theorem, i.e. that for any consistent r.e. extension Γ of Q, a sentence undecidable in Γ can be obtained (in a uniform way for all Γ) effectively from Γ itself. Specifically, given a formula A defining -K, we can find a recursive function ψ such that for all e, $\psi(e)$ is a number such that the statement that A is true of $\psi(e)$ is true but unprovable from W_e if W_e is a consistent extension of Q, and undecidable in W_e if W_e is also ω-consistent. (We say that a sentence is a theorem of W_e if it is a theorem of the set of sentences whose Gödel numbers are elements of W_e; so if W_e contains numbers other than the Gödel numbers of sentences, we ignore them.)

Recall the proof of Gödel's theorem. Let $\Gamma = W_e$ be any r.e. axiom system, and let A(x) be some Π_1 formula that defines -K. Then let $(-K)^* = \{m: \Gamma \text{ fi } A(\mathbf{0}^{(m)})\}$, the set of numbers provably in -K. Since $(-K)^*$ is r.e., for the familiar reasons, $(-K)^*$ is W_f for some f. Then the proof we are familiar with shows that $A(\mathbf{0}^{(f)})$ is true but unprovable in Γ provided that Γ is a consistent extension of Q, and undecidable if Γ is ω-consistent. Intuitively, f depends effectively on e, so f should be $\psi(e)$ for some recursive function ψ. It is the proof that this is the case that uses the S^m_n theorem.

Uniform Effective Form of Gödel's Theorem: For every Π_1 formula A(x) defining -K, there is a recursive function ψ such that for all e, $A(\mathbf{0}^{(\psi(e))})$ is true but unprovable from W_e, if W_e is a consistent extension of Q, and undecidable if W_e is an ω-consistent extension of Q.
Proof: Let A(x) be a fixed Π_1 formula defining -K, and let R be the relation $\{<e, m>: A(\mathbf{0}^{(m)})$ is a theorem of $W_e\}$. If R is r.e., then by the S^1_1 theorem we can find a recursive ψ

such that $W_{\psi(e)}=R_e = \{m: R(e, m)\} = \{m: A(\mathbf{0}^{(m)})$ is a theorem of $W_e\}$ for all e. So $A(\mathbf{0}^{(\psi(e))})$ must be true but unprovable if W_e is a consistent extension of Q, and undecidable if W_e is an ω-consistent extension of Q. So we only have to prove that R is r.e., but this is clear. Let χ be a recursive function such that $\chi(m)$ is a certain Gödel number of $A(\mathbf{0}^{(m)})$. Note that $A(\mathbf{0}^{(m)})$ is a theorem of W_e iff there is a proof sequence from the sentences of W_e on which a Gödel number of $A(\mathbf{0}^{(m)})$ occurs. A proof sequence from W_e is simply a finite sequence of numbers, each of which either codes a sentence in W_e or a logical axiom, or follows from earlier terms in the sequence by a logical rule of inference. So it is clear that we can find an RE formula PS(s, e) which says that s is a proof sequence from W_e; we can then define Th(e, x) as $(\exists s)(\exists n \leq s)(PS(s, e) \wedge [n, x] \in s)$. Th(e, x) says that x is a Gödel number of a formula provable from W_e. Using the function χ above, the relation R is defined by the RE formula $Th(e, \chi(m))$.

We say that a nonrecursive r.e. set S *satisfies the uniform effective form of Gödel's theorem* just in case for some Π_1 formula A(x) defining -S, there is a recursive function ψ such that for all e, $A(\mathbf{0}^{(\psi(e))})$ is true but unprovable from W_e, if W_e is a consistent extension of Q, and undecidable if W_e is an ω-consistent extension of Q. The theorem just proved shows that the set K satisfies the uniform effective form of Gödel's theorem . However, not every nonrecursive r.e. set satisfies it. In particular, Post's simple set (defined in the exercises) does not satisfy the uniform effective form of Gödel's theorem.

The Second Incompleteness Theorem.

We shall now use the uniform effective form of Gödel's theorem to prove a version of Gödel's second incompleteness theorem, the theorem that says that a sufficiently strong r.e. axiom system cannot prove its own consistency. Our proof is based on a proof by Feferman, although it differs from that proof in an important respect. Before giving the proof, we will say a little bit about the philosophical background of Gödel's second incompleteness theorem.

In the early decades of the twentieth century, many mathematicians believed, especially because of the paradoxes, that mathematics might be in serious foundational trouble. Several leading mathematicians had then a strong interest in logic and foundations. Many of these mathematicians thought that the reason behind the trouble is that one cannot reason validly about the infinite, at least in a "natural" way, e.g., they thought that one cannot reason validly about the totality of natural numbers, as opposed to something you can reason about by reasoning about larger and larger initial segments.

Two of those leading mathematicians with strong foundational interests were Brouwer and Hilbert. Brouwer thought from the beginning that mathematics had to be radically revised, and he proposed a doctrine of what mathematical reasonings are acceptable, called

'intuitionism'. In intuitionism, infinitary constructions were not acceptable, and principles about infinite collections licensed by classical logic, like the principle that, for a given property, either all numbers have it or there is a number that is a counterexample; thus, a proof that not all numbers have a certain property does not guarantee, for the intuitionist, that there is a number without that property (this can only be shown by constructing such a number).

Some mathematicians adopted the point of view on foundations common today, i.e., the point of view that there was no problem of legitimacy with mathematics as it had been done, including set theory; in the case of the logicists, at least a certain modified logical form of set theory was legitimate. An entirely different approach to the foundational crisis was taken by Hilbert. He thought that the intuitionists were right in their worries whether mathematics as it was being done was legitimate. He further thought that the set of methods of mathematical reasoning guaranteed to be legitimate was even more restrictive than the set of methods allowed by the intuitionists. On the other hand, Hilbert did not want to change mathematics. He had the following idea. One should develop mathematics by means of formal systems, as had been done by people working in logic and foundations, and view mathematical theorems as finite strings of symbols without meaning, which could be generated in mechanical ways in the formal systems. But one should prove, by the restrictive methods allowed, that the formal systems of mathematics were consistent.

What would be the value of such a proof of consistency? Normally, the reason we don't want a formal system to be inconsistent is that not all of the theorems of an inconsistent system can be true. Since Hilbert thought that not all theorems of mathematics could be true, this was not his reason for demanding a proof of consistency. Another reason is to show that the system is not uninteresting, for an inconsistent system is uninteresting in the sense that it proves every sentence. But there were other reasons as well. We have proved for our own formalisms that if we have a Π_1 statement (x)L(x), where L(x) is a limited formula, first, we can decide, for any instance L($\mathbf{0}^{(n)}$) of L(x), whether L($\mathbf{0}^{(n)}$) is true or not. But second, and more important, that if the system is consistent, then if (x)L(x) is provable then all the instances of L(x) are true; for if some instance was false, it could be shown to be so by finite methods (limited statements, whose quantifiers involve only initial segments of the natural numbers, are the kind of statements taken to be legitimate by Hilbert), and then ~(x)L(x) would be provable, rendering the system inconsistent if (x)L(x) is provable too. In this way, a proof of consistency would provide a legitimation for theorems of the form (x)L(x).

What is known as Hilbert's Program was not merely the idea that proving consistency would be a good thing. The Program suggested by Hilbert actually included a particular and very plausible suggestion of how a proof might be attempted. At the time, it looked as if this suggestion (which we cannot explain here) really ought to work. That's why Gödel's second incompleteness theorem came as a shock, for it showed that consistency for a system could not be proved assuming that Hilbert's restricted finite methods were a subset of the methods

incorporated into the system itself. We can already see from Gödel's first incompleteness theorem that Hilbert's aim was unattainable. For if consistency was provable, then the statement that every Π_1 provable statement is true would be provable. But if this was provable, the Gödel sentence G, which is Π_1, would be such that 'G is provable $\supset$ G' would be a theorem; but G says of itself that it is not provable, so '~G $\supset$ G' would be a theorem, and so by logic G would be a theorem. And this would imply, by the first incompleteness theorem, that the system was not consistent after all.

Let us now give our proof of Gödel's second incompleteness theorem. First, let us see how to write out the first incompleteness theorem in the language of arithmetic. Pick a Π_1 formula A(x) which defines -K, and fix a recursive function ψ as in the uniform effective form of Gödel's theorem proved above. Then

For all e, if W_e is consistent and W_e extends Q, then $A(\mathbf{0}^{(\psi(e))})$ is true but unprovable,

from which it follows that

(†) For all e, if W_e is consistent and W_e extends Q, then $A(\mathbf{0}^{(\psi(e))})$ is true.

(We leave out the second part on the hypothesis of ω-consistency.) We shall write out (†) in the language of arithmetic. We have in effect already seen how to write out the statement that W_e is consistent. We have an RE formula Th(e, x) which says that x is a theorem of W_e; W_e is consistent just in case $\mathbf{0} \neq \mathbf{0}$ is not a theorem of W_e, so W_e is consistent iff e satisfies ~Th(e, $\mathbf{0}^{(n)}$), where n is a Gödel number of $\mathbf{0} \neq \mathbf{0}$; let us write Con(e) for ~Th(e, $\mathbf{0}^{(n)}$). (Alternatively, we could let Con(e) be the sentence $(\exists x)$~Th(e, x), since W_e is consistent iff at least one sentence is not provable from W_e; or we could let Con(e) be the statement that no sentence and its negation are both provable from W_e.) And we can easily write "W_e extends Q" within the system: Q has finitely many axioms A_1, ..., A_k, so let n_1, ..., n_k be their Gödel numbers; W_e extends Q just in case e satisfies Th(e, $\mathbf{0}^{(n_1)}$) $\wedge$ ... $\wedge$ Th(e, $\mathbf{0}^{(n_k)}$). Let us write "e ext. Q" for this formula. Finally, let PS(x, y) be some formula that weakly represents ψ in Q. Now consider the statement

(*) (e)(Con(e) $\wedge$ e ext. Q $\supset$ $(\exists y)$(PS(e, y) $\wedge$ A(y)))

(*) is a partial statement of the first incompleteness theorem, and therefore ought to be provable in reasonably strong systems of number theory. Now consider the theory Q+(*).

Gödel's Second Incompleteness Theorem: If W_e is a consistent extension of Q+(*), then Con($\mathbf{0}^{(e)}$) is not a theorem of W_e, i.e. W_e does not prove its own consistency.
Proof: Suppose W_e extends Q* and Con($\mathbf{0}^{(e)}$) is one of its theorems. Then as (*) is a theorem of W_e, Con($\mathbf{0}^{(e)}$) $\wedge$ $\mathbf{0}^{(e)}$ ext. Q $\supset$ $(\exists y)$(PS($\mathbf{0}^{(e)}$, y) $\wedge$ A(y)) is also a theorem of

W_e; we already know that Con($\mathbf{0}^{(e)}$) is a theorem of W_e, and $\mathbf{0}^{(e)}$ ext. Q is a true sentence of RE and is therefore a theorem of Q and therefore of W_e; so $(\exists y)(PS(\mathbf{0}^{(e)}, y) \wedge A(y))$ is a theorem of W_e. Let $f = \psi(e)$. Since PS represents ψ in Q, W_e fi $PS(\mathbf{0}^{(e)}, \mathbf{0}^{(f)}) \wedge (y)(PS(\mathbf{0}^{(e)}, y) \supset y = \mathbf{0}^{(f)})$; it follows that $A(\mathbf{0}^{(f)})$ is a theorem of W_e. But we already know from the first incompleteness theorem that $A(\mathbf{0}^{(f)})$ is unprovable in W_e if W_e is a consistent extension of Q. Since W_e is an extension of Q, it follows that W_e is inconsistent.

The theorem does not show that there are no statements which might be thought of as expressing the consistency of a system which are not provable in the system, pathological statements of consistency, so to speak. To see this, let Γ be an arbitrary consistent r.e. extension of Q, let Pr'(x) be $Pr(x) \wedge x \neq \mathbf{0}^{(n)}$ (where Pr(x) is any Σ_1 formula defining the set of theorems of Γ, and n is the Gödel number of $\mathbf{0} \neq \mathbf{0}$), and let Con'_Γ be the sentence $\sim Pr'(\mathbf{0}^{(n)})$. Since Γ is consistent, $\mathbf{0} \neq \mathbf{0}$ is not a theorem of Γ, so Pr'(x) defines the set of theorems provable in Γ; if Γ is ω-consistent, then Pr'(x) weakly represents the theorems of Γ in Γ as well. So in a sense, Con'_Γ says that Γ is consistent. However, it is clear that Γ fi $\sim Pr(\mathbf{0}^{(n)})$, i.e. Γ fi Con'_Γ. Also, we know from the exercises that if we have two disjoint r.e. sets, we have weak representations of them which are provably disjoint in Q. If we take the two sets to be on the one hand the set of theorems of Γ, and on the other hand the set of sentences whose negation is a theorem of Γ, we therefore have weak representations of them which are provably disjoint in Q. We might think that the corresponding sentence expresses consistency. One of the aims of Feferman's, and of Jeroslow's, work, was to give conditions for distinguishing these pathological statements from statements for which Gödel's second incompleteness theorem goes through.

An important point about our presentation of Gödel's second incompleteness theorem, where it differs from other presentations, including Feferman's, is that in the hypothesis of the theorem we only require that a single statement (namely, the conjunction of Q and (*)) be a theorem of a system for it to fail to prove its consistency. In other presentations of the theorem, including Gödel's original presentation, the proof that a system does not prove its own consistency requires assuming that a certain sentence, different for each system, is a theorem of the system. Let G be a Gödel sentence for a system Γ which extends Q and let Con_Γ be a sentence in the language of arithmetic that says that Γ is consistent. The first incompleteness theorem states that if Γ is consistent, then G is true but unprovable, so in particular, if Γ is consistent, then G is true. So if Γ is a powerful enough system to prove the first incompleteness theorem, then Γ fi $Con_\Gamma \supset G$. If Γ fi Con_Γ, then Γ fi G; since G is true but unprovable from Γ, it follows that Con_Γ is not a theorem of Γ. This is how the second incompleteness theorem was originally proved, as a corollary of the first incompleteness theorem. Thus, the unprovability of consistency for different Γ's under this presentation is proved under the hypothesis that different sentences are provable in these different Γ's — if Γ and Δ are different systems, then to conclude that neither Γ nor Δ prove

their consistency one must assume that Γ fi $Con_\Gamma \supset G$, and that Δ fi $Con_\Delta \supset D$ (where D is a Gödel sentence for Δ).

On our approach, taking Con_Γ to be $Con(\mathbf{0}^{(e)})$, where e is an index for Γ, we give a single sentence (*) such that any consistent r.e. Γ which extends Q + (*) fails to prove Con_Γ. Without any job of formalization at all, it is shown that any extension of Q+(*) satisfies the second incompleteness theorem. And a system that does not contain Q+(*) is not sufficient for elementary number theory, since it should be clear that the methods used in class can be regarded as methods of elementary number theory.

This much we can say without any formalization at all. And we can presume that some systems are strong enough to contain elementary number theory, and therefore to prove Q+(*). So we know enough at this point to state the main philosophical moral of the second incompleteness theorem -a system in standard formalization strong enough to contain elementary number theory cannot prove its own consistency. Strictly speaking we have stated this only for formalisms whose language is the first-order language of arithmetic, but the technique is easily extended to first-order systems in standard formalization with a richer vocabulary. Some ideas as to how to consider such systems will become clear when we discuss the Tarski-Mostowski-Robinson theorem in a later lecture.

If one wishes to consider a specific system, such as the system we have called 'PA', we can say in advance that it satisfies the conditional statement that if it contains elementary number theory, it cannot prove its own consistency in the sense of $Con(\mathbf{0}^{(e)})$ above. However, we have a task of formalization if we wish to show that the system contains elementary number theory or at any rate Q+(*). Here is one of the misleading features of the name 'Peano arithmetic' that has been used for this system: it gives the impression that by definition the system contains elementary number theory, when in fact it requires a detailed formalization to show that this is so. If, for example, the properties of exponentiation or factorial could not be developed in it, it would not contain elementary number theory after all. We have seen the basic idea of how to do this, but the formalization here is not trivial. Thus it does require a considerable task of formalization to show that (*) can be proved in PA, and hence that the appropriate statement $Con(\mathbf{0}^{(e)})$ is not provable in PA. But it requires no formalization at all to claim that any system in standard formalization containing elementary number theory fails to prove its own consistency.

Lecture XIV

The Self-Reference Lemma.

When Gödel proved the incompleteness theorem, he used the fact that there is a sentence G with Gödel number n which is provably equivalent to the sentence ~Pr($\mathbf{0}^{(n)}$) saying that the formula with Gödel number n is not a theorem. Thus in a sense G says of itself that it is unprovable. We have already pointed out that it is difficult to even remember how G is constructed, and that Gödel's theorem is more naturally motivated by considering the properties of the sentence ~Prov($\mathbf{0}^{(n)}$,$\mathbf{0}^{(n)}$), where n is the Gödel number of ~Prov(x,x). In this sense, Gödel's use of the fact about "self-reference", had the negative effect of making his proof appear somewhat mysterious. On the other hand, it had the positive effect of calling attention to the fact that the argument for the existence of G does not depend in any way on the choice of the predicate ~Pr(x), and establishes a more general claim (which, although not stated by Gödel, can be reasonably attributed to him), usually referred to as 'the self-reference lemma'.

Self-Reference Lemma. Let A(x) be any formula in one free variable in the language of arithmetic (or RE). Then there is a sentence G of the language of arithmetic (of RE) such that G ≡ A($\mathbf{0}^{(n)}$) is a theorem of Q, where n is a Gödel number of G.

(In the case of RE, this could be made precise in two ways: either showing that the translation of G ≡ A($\mathbf{0}^{(n)}$) into the narrow language of arithmetic is provable in Q or showing that the appropriate sentence in the broad language of arithmetic is provable in the appropriate formalization of Q.)

Intuitively, G says of itself that it has the property A(x). To prove a version of the first incompleteness theorem using the lemma, let Γ be any consistent r.e. extension of Q, and let Pr(x) be a formula that defines the set of theorems of Γ in RE. Use the self-reference lemma to obtain a sentence G such that G ≡ ~Pr($\mathbf{0}^{(n)}$) is a theorem of Q and hence of Γ, where n is a Gödel number of G. If G is a theorem of Γ, then Pr($\mathbf{0}^{(n)}$) is a true sentence of RE, and hence is provable in Q and therefore in Γ; since Γ fi G ≡ ~Pr($\mathbf{0}^{(n)}$), Γ fi ~G, so ~G is also a theorem of Γ and Γ is inconsistent. Since we are assuming that Γ is consistent, G is not a theorem of Γ. However, since G says of itself that it is not a theorem of Γ, G is true; or more formally, ~Pr($\mathbf{0}^{(n)}$) is true since G is not a theorem of Γ, G ≡ ~Pr($\mathbf{0}^{(n)}$) is a theorem of Q and is therefore true, so G is true. So G is true but unprovable. The proof of the self-reference lemma reveals that G is a Π_1 sentence; from this it follows that if Γ is ω-consistent, ~G is not provable either.

Notice that we often state the Gödel theorems saying that the sentence obtained is one which is true but unprovable. If the self-reference lemma is stated for the language of

arithmetic, we know that the predicate Tr(x) saying that x is the Gödel number of a true sentence cannot be defined in arithmetic itself. We know also that the opposite situation holds for the language RE. Either way, we have the following corollary which, like the lemma itself, holds for both the language of arithmetic and the language RE:

Corollary: Let A(x) be any formula in one free variable in the language of arithmetic (or RE). Then there is a sentence G of the language of arithmetic (or of RE) with Gödel number n such that $G \equiv A(\mathbf{0}^{(n)})$ and $A(\mathbf{0}^{(n)}) \equiv Tr(\mathbf{0}^{(n)})$ are both true.

There are numerous ways of proving the self-reference lemma. Given our Gödel numbering, G cannot actually be the sentence $A(\mathbf{0}^{(n)})$, since the Gödel number of $A(\mathbf{0}^{(n)})$ must be larger than n. However, it is possible to devise a different Gödel numbering such that for every formula A(x), there is a number n such that $A(\mathbf{0}^{(n)})$ gets Gödel number n. (This method of proving the self-reference lemma was discovered independently by Raymond Smullyan and the author.) If we add extra constants to our language, then we can prove a version of the self-reference lemma for the expanded language. Specifically, let L* be the language obtained from the language of arithmetic by adding the constants a_2, a_3, ... (a_1 is already in L). Interpret the new constants as follows: if n is a Gödel number of a formula $A(x_1)$, then interpret a_{n+1} as the least Gödel number of $A(a_{n+1})$. Then the sentence $A(a_{n+1})$ says of itself that it is A. Note that if m_n is the Gödel number of $A(a_{n+1})$, the sentence $a_{n+1} = \mathbf{0}^{(m_n)}$ is true under this interpretation. If we let Q* be the axiom system obtained from Q by adding as axioms all sentences of the form $a_{n+1} = \mathbf{0}^{(m_n)}$, then Q* fi $A(a_{n+1}) \equiv A(\mathbf{0}^{(m_n)})$ for all n, so we can let G be the sentence $A(a_{n+1})$. So if we chose to work in the language L* rather than L, we could get the self-reference lemma very quickly; moreover, L* does not really have greater expressive power than L, since L* simply assigns new names to some things that already have names in L. Using this version of the self-reference lemma it is also possible to prove Gödel's incompleteness theorem, as we have seen in an exercise.

The proof of the self-reference lemma essentially due to Gödel employs the usual Gödel numbering and constructs the sentence G in a more complicated way. Let A(x) be given. Let ϕ be a recursive function such that if y is the Gödel number of a formula $C(x_1)$, then $\phi(n, y)$ is the Gödel number of $C(\mathbf{0}^{(n)})$. Let B(x, y, z) represent ϕ in Q, and let A'(x, y) be the formula $(\exists z)(B(x, y, z) \wedge A(z))$. If y is the Gödel number of a formula $C(x_1)$, then A'(n, y) holds iff the Gödel number of $C(\mathbf{0}^{(n)})$ satisfies A(x). (We can read A'(x, y) as "y is A of x"; for example, if A(x) is "x is provable", then A'(x, y) is "y is provable of x".) Let m be the Gödel number of $A'(x_1, x_1)$, and let G be the sentence $A'(\mathbf{0}^{(m)}, \mathbf{0}^{(m)})$. ($A(x_1, x_1)$ says that x_1 is A of itself, and G says that "is A of itself" is A of itself.) We shall show that Q fi $G \equiv A(\mathbf{0}^{(n)})$, where n is the Gödel number of G. Note that G is really $(\exists z)(B(\mathbf{0}^{(m)}, \mathbf{0}^{(m)}, z) \wedge A(z))$, where B represents ϕ in Q. Note also that $\phi(m, m)$ is the Gödel number of G itself, since m is the Gödel number of $A'(x_1, x_1)$ and G is $A'(\mathbf{0}^{(m)}, \mathbf{0}^{(m)})$; so Q fi $B(\mathbf{0}^{(m)}, \mathbf{0}^{(m)},$

$\mathbf{0}^{(n)}) \wedge (y)(B(\mathbf{0}^{(m)}, \mathbf{0}^{(m)}, y) \supset y = \mathbf{0}^{(n)})$. So Q fi $A(\mathbf{0}^{(n)}) \supset (\exists z)(B(\mathbf{0}^{(m)}, \mathbf{0}^{(m)}, z) \wedge A(z))$, i.e. Q fi $A(\mathbf{0}^{(n)}) \supset G$; and Q fi $G \supset B(\mathbf{0}^{(m)}, \mathbf{0}^{(m)}, \mathbf{0}^{(n)}) \wedge A(\mathbf{0}^{(n)})$, so Q fi $G \supset A(\mathbf{0}^{(n)})$. Therefore, Q fi $G \equiv A(\mathbf{0}^{(n)})$.

The proof of the self-reference lemma that will be the preferred one in our treatment is perhaps the standard one nowadays, and uses some of the recursion theory that we have already developed. It is as follows. Let the formula A be given. Let $PH(x_1, x_2, y)$ be a formula that functionally represents Φ in Q (recall that Φ is a function that enumerates the unary partial recursive functions). Let ψ be a recursive function such that $\psi(m)$ is a certain Gödel number of $(\exists y)(PH(\mathbf{0}^{(m)}, \mathbf{0}^{(m)}, y) \wedge A(y))$. That there is one such recursive function is clear by the familiar reasons. In fact, we may naturally let ψ be just an S_n^m function for the given formula $(\exists y)(PH(x_2, x_2, y) \wedge A(y))$ (which we may take to have number e). Let f be an index of ψ. Let G be the sentence $(\exists y)(PH(\mathbf{0}^{(f)}, \mathbf{0}^{(f)}, y) \wedge A(y))$. $\Phi(f, f) = \psi(f)$ = a Gödel number of $(\exists y)(PH(\mathbf{0}^{(f)}, \mathbf{0}^{(f)}, y) \wedge A(y))$ = a Gödel number of G. Letting $n = \psi(f)$, Q fi $G \supset A(\mathbf{0}^{(n)})$ (since Q fi $G \supset PH(\mathbf{0}^{(f)}, \mathbf{0}^{(f)}, \mathbf{0}^{(n)}) \wedge A(\mathbf{0}^{(n)})$, as PH functionally represents Φ in Q), and Q fi $A(\mathbf{0}^{(n)}) \supset G$ (since Q fi $PH(\mathbf{0}^{(f)}, \mathbf{0}^{(f)}, \mathbf{0}^{(n)})$). Thus, Q fi $G \equiv A(\mathbf{0}^{(n)})$.

Through a similar proof we can obtain an effective version of the self-reference lemma:

Self-Reference Lemma. Effective Version: There is a recursive function ϕ such that for all formulae A(x) of the language of arithmetic (RE) in one free variable, if m is a Gödel number of A(x), then $\phi(m)$ is a Gödel number of a sentence G_m of the language of arithmetic (RE) such that Q fi $G \equiv A(\mathbf{0}^{(\phi(m))})$.

Proof: Let $PH(x_1, x_2, x_3, y)$ be a formula that functionally represents Φ^3 in Q (recall that Φ^3 is a function that enumerates the 2-place partial recursive functions). Let ψ be a 2-place recursive function such that if p is a Gödel number of a formula A(y), $\psi(q,p)$ is a certain Gödel number of $(\exists y)(PH(\mathbf{0}^{(q)}, \mathbf{0}^{(q)}, \mathbf{0}^{(p)}, y) \wedge A(y))$. This may be taken again to be an S_n^m function. Let f be an index of ψ, and let $\phi(p)=\psi(f,p)$. Then $\phi(p)$ will be a code of the sentence $G_p = (\exists y)(PH(\mathbf{0}^{(f)}, \mathbf{0}^{(f)}, \mathbf{0}^{(p)}, y) \wedge A(y))$, if p is a Gödel number of A(y). So if p is a Gödel number of A(y), $\Phi(f, f, p) = \psi(f,p) = \phi(p)$ = a Gödel number of $(\exists y)(PH(\mathbf{0}^{(f)}, \mathbf{0}^{(f)}, \mathbf{0}^{(p)}, y) \wedge A(y))$ = a Gödel number of G_p. Letting $r = \phi(p)$, Q fi $G_p \supset A(\mathbf{0}^{(r)})$ (since Q fi $G_p \supset PH(\mathbf{0}^{(f)}, \mathbf{0}^{(f)}, \mathbf{0}^{(p)}, \mathbf{0}^{(r)}) \wedge A(\mathbf{0}^{(r)})$, as PH functionally represents Φ^3 in Q), and Q fi $A(\mathbf{0}^{(r)}) \supset G_p$ (since Q fi $PH(\mathbf{0}^{(f)}, \mathbf{0}^{(f)}, \mathbf{0}^{(p)}, \mathbf{0}^{(r)})$).

The proofs of the self-reference lemma do not depend on the fact that A has only one free variable. Noting this allows us to state a more general version of the self-reference lemma in which G is allowed to have free variables.

Self-Reference Lemma with Free Variables: Let $A(x, y_1, ..., y_m)$ be a formula of the language of arithmetic (or RE) with all free variables shown; then there is a formula $G(y_1, ..., y_m)$ of the language of arithmetic (or of RE) such that Q fi $(y_1)...(y_m)(G(y_1, ..., y_m) \equiv$

$A(\mathbf{0}^{(n)}, y_1, ..., y_m))$, where n is a Gödel number of G.

The version of the self-reference lemma in which G does not have free variables is simply the special case of this lemma in which n = 0. Naturally, there is an effective version of the self-reference lemma with free variables.

A corollary of the self-reference lemma with free variables is the following:

Corollary: Let A(x, y) be a formula of the language of arithmetic (or RE) with all free variables shown; then there is a formula G(y) of the language of arithmetic (or of RE) with Gödel number n such that $(y)(G(y) \equiv A(\mathbf{0}^{(n)},y))$ and $(y)(A(\mathbf{0}^{(n)},y) \equiv Sat(\mathbf{0}^{(n)},y))$ are both true.

(In the case of RE, Sat(x,y) is W(x,y). In the case of the language of arithmetic, Sat(x,y), which we use to mean that y satisfies the formula of the language of arithmetic with Gödel number x, is not itself a formula of the language.)

The self-reference lemma with free variables might be given the name "self-reference lemma with parameters", but this name is more appropriate for the following variant of the lemma.

Self-Reference Lemma With Parameters: For any formula A(x), there is a recursive function ψ and a formula PS(x, y) that represents ψ in Q, such that for all m, $\psi(m)$ is the Gödel number of the formula $(\exists z)(PS(\mathbf{0}^{(m)}, z) \wedge A(z))$, and furthermore this formula is provably equivalent in Q to $A(\mathbf{0}^{(\psi(m))})$.

Proof: Let χ be a recursive function such that if m is the Gödel number of a formula $B(x_1, x_2)$, then $\chi(m, n, p)$ is the Gödel number of the formula $B(\mathbf{0}^{(n)}, \mathbf{0}^{(p)})$. Let CH(x, y, z, w) be a formula that represents χ in Q. Let n be the Gödel number of the formula $(\exists x_3)(CH(x_1, x_1, x_2, x_3) \wedge A(x_3))$, and let PS(x, y) be the formula $CH(\mathbf{0}^{(n)}, \mathbf{0}^{(n)}, x, y)$. PS represents the function $\psi(x) = \chi(n, n, x)$; to prove the theorem, we only have to show that $\psi(m)$ is the Gödel number of the formula $(\exists z)(PS(\mathbf{0}^{(m)}, z) \wedge A(z))$, for any m. Since n is the Gödel number of $(\exists x_3)(CH(x_1, x_1, x_2, x_3) \wedge A(x_3))$, it follows that $\psi(m) = \chi(n, n, m)$ = the Gödel number of $(\exists x_3)(CH(\mathbf{0}^{(n)}, \mathbf{0}^{(n)}, \mathbf{0}^{(m)}, x_3) \wedge A(x_3))$, which is the formula $(\exists x_3)(PS(\mathbf{0}^{(m)}, x_3) \wedge A(x_3))$.

Now, notice that $(\exists z)(PS(\mathbf{0}^{(m)}, z) \wedge A(z))$ is provably equivalent to $A(\mathbf{0}^{(\psi(m))})$. Thus, writing G(x) for $(\exists z)(PS(x, z) \wedge A(z))$, we have

$$Q \text{ fi } G(\mathbf{0}^{(m)}) \equiv A(\mathbf{0}^{(\psi(m))})$$

for all m, where $\psi(m)$ is a Gödel number of $G(\mathbf{0}^{(m)})$.

An alternative proof of the self-reference lemma with parameters consists in noting that we

may take G(x) to be the formula $(\exists y)(PH(\mathbf{0}^{(f)}, \mathbf{0}^{(f)}, x, y) \wedge A(y))$, as in the proof of the effective form of the self-reference lemma, and ψ to be the function ϕ of that same proof.

The Recursion Theorem

Kleene seemed to use the term 'the recursion theorem' as an ambiguous term for two theorems that he proved. Later, the two theorems came to be called Kleene's first and second recursion theorems. Generally speaking, the second recursion theorem is the more powerful of the two. Nowadays, it is usually called '*the* recursion theorem'. We discuss this theorem here (the first recursion theorem will come later). In terms of our formalism, it is simply the self-reference lemma for the language RE with formulae of two variables. Recall that for any 2-place relation R and number e, R_e is the set $\{n: R(e, n)\}$.

Recursion Theorem: For any 2-place r.e. relation R, there is an e such that $W_e = R_e$.

Before proving the recursion theorem, it is worth noting that the result is somewhat surprising. Any r.e. relation R can be thought of as enumerating a subclass of the r.e. sets (namely, the class $\{R_e: e \in \mathbf{N}\}$). We may thus call such a relation a *subnumeration* or the r.e. sets. The recursion theorem says that every subnumeration coincides with W at some point. Offhand, we might have thought that we could obtain a subnumeration which did not coincide with W at any point at all; R might be some scrambling of W, for example. The recursion theorem shows that this is not so.

Note that, since W_e is the set of numbers satisfying the RE formula with Gödel number e, the second recursion theorem says that for any r.e. relation R there is an RE formula A(x) with Gödel number e such that for all n, n satisfies A just in case R(e,n). Since R is itself defined by some RE formula B, this is just to say that for any RE formula B(y, x) of two free variables, there is an RE formula A(x) of one free variable such that for all n, A(x) is true of n iff $B(\mathbf{0}^{(e)}, x)$ is true of n, and so, that $(x)(A(x) \equiv B(\mathbf{0}^{(e)},x))$ is true, where e is the Gödel number of A(x). That is, the recursion theorem is really the self-reference lemma with free variables for RE in the case of one free variable. We can thus prove the recursion theorem by imitating the proof of the self-reference lemma, by considering an S_n^m function for the RE formula $(\exists z)(PH(x_2, x_2, z) \wedge B(z, y))$. This was also the inspiration for Kleene's original proof of the recursion theorem, although he was not working with RE, but with a different formalism. We shall give a proof which, although based essentially on the same underlying facts, is shorter and more common in textbooks.

Proof of the Recursion Theorem: Let R be any 2-place r.e. relation. Consider the relation $S(x, y) = R(\Phi(x, x), y)$. S is an r.e. relation, so apply the S_1^1 theorem to obtain a recursive function ψ such that for all m, $W_{\psi(m)} = S_m = R_{\Phi(m, m)}$. Since ψ is recursive, it has

an index f. Let $e = \psi(f)$; $W_e = W_{\psi(f)} = S_f = R_{\Phi(f, f)} = R_e$ (since $\Phi(f, f) = \psi(f) = e$).

This proof is breathtakingly short. It only uses the fact that W is an enumeration for which the statement of the S_n^m theorem holds.

In the same way that there is an effective version of the self-reference lemma with free variables, there is an effective form of the recursion theorem that is easy to state and prove: there is a recursive function ϕ such that for any 2-place r.e. relation R with index e, $W_{\phi(e)}=R_{\phi(e)}$.Of course, the effective version, like the noneffective, can be proved for all appropriate formalisms, and not just for RE.

The recursion theorem can be generalized to n+1-place r.e. relations. If R is an n+1-place relation, then let R_e be the relation $\{<x_1, ..., x_n>: R(e, x_1, ..., x_n)\}$; the general form of the recursion theorem states that for every n+1-place r.e. relation R, there is an e such that $W_e^{n+1} = R_e$.

Besides being surprising, the recursion theorem has curious consequences. Let R(x, y) be the relation W(x+1, y). Then $W_e = R_e = W_{e+1}$ for some e; so W enumerates the r.e. sets in such a way that at least one such set is listed two consecutive times. More generally, we see that for all n there is an e such that $W_e = W_{e+n}$; so W has many repetitions. (It is natural to ask whether a repetition-free enumeration of the r.e. sets exists; it turns out that such enumerations do exist, but are hard to construct.) Also, we can find a number e such that $W_e = \{e\}$; just let R(e, x) be the identity relation. Since this relation is certainly r.e., we can use the recursion theorem to find an e such that W(e, x) iff x = e, i.e. we can find a formula A(x) which is satisfied only by its own Gödel number.

More generally still, let ψ be any recursive function; by letting $R(x, y) = W(\psi(x), y)$, we see that $W_e = W_{\psi(e)}$ for some e. So we have the following

Theorem: For every recursive function ψ, there is an e such that $W_e = W_{\psi(e)}$.

This theorem looks superficially like a fixed-point theorem, and we will sometimes refer to it as 'the fixed-point version of the recursion theorem'. Notice, however, that it is not quite a fixed point theorem. A fixed point theorem states that a function F has a fixed-point, i.e. there is an a such that F(a) = a. On the one hand, the theorem does not show that ψ itself has a fixed point, since we can have $\psi(e) \neq e$ and $W_e = W_{\psi(e)}$. On the other hand, the "function" $F(W_e) = W_{\psi(e)}$ is not really a function at all, since its value depends not only on its argument, the set W_e, but also on the index e (we can have $W_e = W_f$ and $W_{\psi(e)} \neq W_{\psi(f)}$). By contrast, Kleene's first recursion theorem, which we shall eventually prove, really is a fixed-point theorem.

There is also a version of the recursion theorem for Φ. In fact, there are two versions, corresponding to the first version and to the fixed-point version.

Recursion Theorem for Partial Recursive Functions: (a) For all partial recursive ψ

there is an e such that $\Phi(e, x) = \psi(e, x)$, all x; and (b) for all total recursive ψ there is an e such that $\Phi(e, x) = \Phi(\psi(e), x)$, all x.

Proof: For (a), recall that Φ is really a uniformization of the relation W^3. Let PS(x, y, z) be the graph of ψ. Find an e such that $W_e = PS_e$, i.e. for all y and z, W(e, y, z) iff PS(e, y, z) iff $\psi(e, y) = z$; then W_e is single-valued, so W(e, y, z) iff $\Phi(e, y) = z$; so $\Phi(e, y) = \psi(e, y)$.

(b) is immediate from (a): let $\chi(x, y) = \Phi(\psi(x), y)$, and let e be an index of χ; then $\Phi(e, x) = \chi(e, x) = \Phi(\psi(e), x)$.

Form (b) is the form that usually is referred to as 'the recursion theorem' in the literature.

The recursion theorem is interesting mainly because the relation R can itself involve W, as we saw in the case $R(x, y) = W(\psi(x), y)$. To illustrate why this is useful, we shall give a proof, using the recursion theorem, that the factorial function is recursive. (This illustrates, by the way, why the theorem is called 'the recursion theorem'.) To show this, it suffices to show that the graph of the factorial function is recursive. If R is a relation such that

$$(*) \qquad R(x, y) \equiv (x = 0 \wedge y = 1) \vee (\exists n)(\exists z)(x = n+1 \wedge y = (n+1)\cdot z \wedge R(n, z)),$$

then R is the graph of the factorial function. (This can be seen by showing, by induction on x, that there is exactly one y such that R(x, y), and y = x!.) So we only have to find an r.e. relation R that satisfies (*). If R is r.e., then $R = W^3_e$ for some e, so an appropriate R exists just in case

$$W(e, x, y) \equiv (x = 0 \wedge y = 1) \vee (\exists n)(\exists z)(x = n+1 \wedge y = (n+1)\cdot z \wedge W(e, n, z))$$

holds for some e. Setting $S(e, x, y) \equiv (x = 0 \wedge y = 1) \vee (\exists n)(\exists z)(x = n+1 \wedge y = (n+1)\cdot z \wedge W(e, n, z))$, we see that S is r.e. and that y = x! is recursive if

$$W(e, x, y) \equiv S(e, x, y)$$

for some e; but by the recursion theorem, such an e exists. We can similarly show that the Ackermann function is recursive. More generally, we can use the recursion theorem to find partial recursive functions that satisfy arbitrary systems of equations. For example, consider the system consisting of the two equations

$$\psi(0) = 1$$
$$\psi(n+1) = \psi(n)\cdot(n+1)$$

We can use an argument similar to the one given above to show that there is a partial recursive function satisfying these equations. In this case, we see that the function in question is total. In general, however, we cannot guarantee this. For example, let our

system of equations consist of just the equation $\psi(x) = \psi(x)+1$. This does indeed have a solution, namely the function which is undefined everywhere.

So far, we have not used the recursion theorem to prove anything that could not be proved already using the generated sets theorem. However, there are some important applications of the recursion theorem that go beyond the generated sets theorem. Unfortunately, these applications are not as easy to state as the ones just given, and presuppose some knowledge of transfinite ordinals. Just as we can define functions on the natural numbers by ordinary induction, we can define functions on the ordinals by transfinite induction; and if α is a limit ordinal, $f(\alpha)$ will in general depend on the infinitely many values $f(\beta)$ for $\beta < \alpha$. Thus, we cannot use the generated sets theorem to show that such a function is recursive, since we cannot generate $f(\alpha)$ until we have generated $f(\beta)$ for all $\beta < \alpha$, and at no stage have we actually generated infinitely values. Nonetheless, we can intuitively define f by a system of equations. For example, we might define ordinal exponentiation by

$$\alpha^0 = 1$$
$$\alpha^{\beta+1} = \alpha^\beta \cdot \alpha$$
$$\alpha^\beta = \sup\{\alpha^\gamma: \ \gamma < \beta\} \text{ when } \beta \text{ is a limit.}$$

In fact, we can use the recursion theorem to show that this system of equations defines a recursive function on the recursive ordinals (i.e. those ordinals which are order types of recursive well-orderings), in essentially the way we showed that the factorial function is recursive. (However, for this to make sense we need a way of coding up the recursive ordinals as natural numbers.) Thus, we can use the recursion theorem to get around the problem that the value of α^β depends on that of infinitely α^γ's for $\gamma < \beta$ when β is a limit. (Since what we are really defining is an *index* e of the ordinal exponentiation function, the set $\{\alpha^\gamma: \ \gamma < \beta\}$ is coded up in a finite way in terms of α, β and e; in effect, this is what allows us to talk about infinitely many values of the function at once.)

Exercises

1. (a) Let S be an r.e. set. Prove that there is a 1-1 recursive function χ such that for all m, $W_{\chi(m)} = \mathbf{N}$ if $m \in S$ and $W_{\chi(m)} = \emptyset$ if $m \notin S$.

(b) Show that K is 1-1 complete. (This is a result that has been long awaited.)

2. (a) Show that an r.e. set S is nonrecursive iff there is a total function ψ such that for all x, $\psi(x) \in S$ iff $\psi(x) \in W_x$. S is called *completely creative* if ψ is recursive, and *1-1 completely creative* if ψ is also 1-1. Observe that K is 1-1 completely creative, where ψ is the identity function.

(b) Prove that every completely creative set is many-one complete, and that every 1-1 completely creative set is 1-1 complete.

(c) Prove that if $S_1 \leq_m S_2$ and S_1 is completely creative, then so is S_2. Also show that if $S_1 \leq_1 S_2$ and S_1 is 1-1 completely creative, then so is S_2.

(d) Show that every many-one complete set is completely creative, and that every 1-1 complete set is 1-1 completely creative.

3. (a) Recall the set S of exercise 6 in Lecture XII. There is a formula $(x_2)L(x_1, x_2)$ with L in Lim that defines the complement of S. Why? Prove that if Γ is a consistent r.e. extension of Q, then only finitely many sentences of the form $(x_2)L(\mathbf{0}^{(m)}, x_2)$ are provable in Γ even though infinitely many such sentences are true. Hence conclude that if Γ is ω-consistent, all but a finite number of true sentences of the form $(x_2)L(\mathbf{0}^{(m)}, x_2)$ are undecidable.

(b) Prove that if the effective form of the Gödel theorem holds for an r.e. set T which is not recursive (in the sense in which it holds for K), then there is an infinite r.e. set that is disjoint from T. Conclude that though the noneffective form of the Gödel theorem holds for the set S of exercise 6 of the midterm assignment, S does not satisfy the effective form. (An r.e. set T with properties (b) and (c) of exercise 6 is called 'simple'. Observe that exercise 6 shows that every simple set is neither recursive nor 1-1 complete. Property (a) of the present exercise also follows from the fact that the set is simple.)

(c) Also show that if T is a nonrecursive r.e. set and T is completely creative, then T satisfies the effective form of Gödel's theorem. (It follows that K satisfies the effective form of Gödel's theorem, which we have already seen to be the case.)

Lecture XV

The Recursion Theorem with Parameters.

Let R be a 3-place r.e. relation, or in other words, a subnumeration of the 2-place r.e. relations. For any given m, let R^m be the relation $\{<e, x>: R(e, x, m)\}$; R^m is a subnumeration of the r.e. sets. It follows from the recursion theorem that for any given m, there is an e such that $W_e = R^m_e$. But more is true: we can find e effectively from m.

Recursion Theorem with One Parameter: For any 3-place r.e. relation R, there is a recursive function ψ such that for all m, $W_{\psi(m)} = R^m_{\psi(m)}$ i.e. for all x and m, $W(\psi(m), x)$ iff $R(\psi(m), x, m)$.

Proof: First, let χ be a recursive function such that $W_{\chi(m, a)} = R^m_{\Phi(a,a)}$ for all m, a. Since $R^y_{\Phi(x,x)}$ is an r.e. relation, such a χ exists by the S^m_n theorem (taking m and a as the parameters). Next, let ϕ be a recursive function such that $\Phi(\phi(m), a) = \chi(m, a)$ for all m, a; ϕ is easily obtainable from a two place function α guaranteed by the S^m_n theorem for partial recursive functions, by taking an index of χ as fixed as the first argument of α. Finally, let $\psi(m) = \Phi(\phi(m), \phi(m)) = \chi(m, \phi(m))$. Then $W_{\psi(m)} = W_{\chi(m, \phi(m))} = R^m_{\Phi(\phi(m),\phi(m))} = R^m_{\psi(m)}$, all m.

This proof should be compared to the proof of the parameter-free recursion theorem; all we have done is to make the number f of that proof depend effectively on m. The theorem can be generalized to more than one parameter via the usual methods, i.e. either by imitation of the proof for one parameter, or via the pairing function.

The more usual statement of the theorem is this: for all 2-place recursive χ there is a 1-place recursive function ψ such that for all m, $W_{\psi(m)} = W_{\chi(\psi(m), m)}$. This follows from the version we have just proved: simply let $R(y, x, m)$ iff $W(\chi(y, m), x)$, and find a ψ such that $W_{\psi(m)} = R^m_{\psi(m)}) = W_{\chi(\psi(m), m)}$.

The recursion theorem with parameters has even spookier applications than the parameter-free version.

Arbitrary Enumerations.

We shall now take a different approach to the S^m_n and recursion theorems, by considering arbitrary enumerations of the r.e. sets rather than simply the specific relation W. This approach has the virtue of making the recursion theorem appear less mysterious than the usual presentation.

For most applications of either the recursion theorem or the S^m_n theorem, we don't need

any specific properties of the relation W except that it is an enumeration. For most applications of the S_n^m and recursion theorems, it suffices to have available the fact that there is some enumeration of the r.e. sets with the properties stated in the S_n^m and recursion theorems for W. Eventually, the approach that we will develop establishes that W has these properties, but it first "cooks up" enumerations with those properties. One can find in the literature the awareness that it is possible to cook up enumerations with the S_n^m property; however, the rest of the theory does not appear in the literature and is due to the author, who developed it without knowing that it had been developed for the S_n^m case.

Let W' be an enumeration of the r.e. sets. For each k, we can easily obtain an enumeration W'^k of the k-place relations from W' via the pairing function. The *diagonal enumeration* of an enumeration of the two-place r.e. relations $W'^2(x,z,y)$, $Diag(W'^2)$, is the relation $W'^2(x, x, y)$. We say that W' is a *recursion enumeration* (or that it has the *recursion property*) if for all r.e. two-place relations R there is an e such that $W'_e = R_e$. We also say that a subnumeration S is a *recursion subnumeration* if for all r.e. two-place relations R there is an e such that $S_e = R_e$; every recursion subnumeration is an enumeration: let A be an r.e. set and let R be the r.e. relation such that R(e,x) iff x is in A; then $A=R_e$ for every e; since S is a recursion subnumeration, there is an e such that $S_e = R_e=A$.

Theorem: For any enumeration in two variables $W'^2(x,z,y)$, its diagonal enumeration $Diag(W'^2)$ is a recursion enumeration.
Proof: That $W'^2(x,z,y)$ is an enumeration means that for every r.e. two-place relation R there is an e such that for all z,y, R(z,y) iff $W'^2(e,z,y)$. In particular, for every R there is an e such that for every y, R(e,y) iff $W'^2(e,e,y)$, i.e. for every R, $R_e=Diag(W'^2)_e$. This proves that $Diag(W'^2)$ is a recursion subnumeration of the r.e. sets, and hence, by our previous result, that it is a recursion enumeration.

This proof of the existence of a recursion enumeration of the r.e. sets from the existence of an enumeration of the two-place r.e. relations is as breathtakingly short as the standard proof that W has the recursion property, if not more so. However, it is much more natural and less mysterious than the latter. Suppose you had an enumeration of the 2-place r.e. relations, and you wanted to construct an enumeration of the r.e. sets with the recursion property. Each 2-place r.e. relation can be thought of as a list of r.e. sets, and the given enumeration of the r.e. relations can be thought of as a list of all these lists; in constructing an enumeration with the recursion property, what you really want to do is to construct a list $W^\dagger$ of r.e. sets which coincides with each of the other r.e. lists at some point. If R is the eth such list, what could be more natural than having $W^\dagger$ coincide with R at the eth place? This is just what we have done in defining $Diag(W'^2)$ above.

We say that W' is a *fixed-point enumeration* (or that it has the *fixed-point property*) if for all total recursive functions ψ there is an e such that $W'_e = W'_{\psi(e)}$. In calling these 'fixed

point enumerations' we are referring to the fact that the fixed-point version of the recursion theorem resembles a fixed point theorem (as we have pointed out, however, it is not really a fixed point theorem). By a proof similar to the proof of the fixed point theorem from the recursion theorem, we can prove the following

Theorem: Every recursion enumeration is a fixed-point enumeration.

The converse fails; that is, there are fixed-point enumerations which are not recursion enumerations.

Let us now define the notion of an enumeration that satisfies the S_n^m theorem. We can say that W' is a *substitution enumeration* (or that it has the *substitution property*) if for any 2-place r.e. relation R there is a 1-1 recursive function ψ such that $W'_{\psi(e)} = R_e$ for all e. Another way of stating the definition of a substitution enumeration is as follows. If R and S are subnumerations of the r.e. sets (i.e. 2-place r.e. relations), and ψ is a recursive function, let us say that ψ is a *translation* of R into S (in symbols, $\psi: R \rightarrow S$) if for all e, $R_e = S_{\psi(e)}$. Let us say that a subnumeration S is *maximal* if for every r.e. R, there is a recursive ψ such that $\psi: R \rightarrow S$; if we can require ψ to be 1-1, then we say that S is *1-1 maximal*. Translation is analogous to reducibility, and maximality (1-1 maximality) is analogous to m-completeness (1-completeness). Clearly, an enumeration is a substitution enumeration just in case it is 1-1 maximal. (A 1-1 maximal enumeration can also be called an *effective* enumeration) Assuming that an enumeration W' exists, it will follow that every maximal subnumeration S is an enumeration, because there will be a recursive function ψ such that for all e, $W_e = S_{\psi(e)}$, and so S enumerates the r.e. sets.

We shall now show that given any enumeration, we can find a 1-1 maximal enumeration.

Theorem: If W' is an enumeration of the r.e. sets, the relation W"([e, n], x) which holds iff $W'^2(e, n, x)$ is a 1-1 maximal enumeration.
Proof: Let W' be an arbitrary enumeration. Let W" be the enumeration such that W"([e, n], x) = $W'^2(e, n, x)$; W" is called the *pairing contraction* of W'^2. (Formally, W" is the r.e. relation defined by $(\exists e)(\exists n)(z = [e, n] \wedge W'^2(e, n, x))$. Note that $W''_z = \emptyset$ when z is not of the form [e, n].) To see that W" is 1-1 maximal, let R be any r.e. relation, and let $R = W'^2_{e_0}$. Let $\psi(n) = [e_0, n]$. $W''(\psi(n), x)$ iff $W'^2(e_0, n, x)$ iff R(n, x), so $W''_{\psi(n)} = R_n$. Since ψ is 1-1, ψ is a 1-1 translation of R into W".

Once we know that W" is a substitution enumeration, it follows that it is a recursion enumeration (and therefore a fixed-point enumeration). In fact, the standard proof of the recursion theorem using the S_n^m theorem establishes that every substitution enumeration is a recursion enumeration, since it doesn't appeal to any properties of W besides its being an enumeration. Actually, the following is also true:

Theorem: If W'_1 and W'_2 are enumerations such that for some recursive ψ, ψ: $W'_1 \rightarrow W'_2$, then, if W'_1 has the recursion property, W'_2 has also the recursion property.
Proof: That W'_1 has the recursion property means that for all r.e. two-place relations R there is an e such that $W'_{1e} = R_e$; that ψ: $W'_1 \rightarrow W'_2$ means that for all e, $W'_{1e} = W'_{2\psi(e)}$. We want to prove that for all r.e. two-place relations R there is an e such that $W'_{2e} = R_e$. Let R be an r.e. two-place relation. There is an e such that for all x, $W'_1(e,x)$ iff R(e,x). Consider the relation R'(x,y) which holds iff $R(\psi(x),y)$. This is an r.e. relation and so there is an e such that for all y, $W'_1(e,y)$ iff R'(e,y) iff $R(\psi(e),y)$ iff $W'_2(\psi(e),y)$. So $\psi(e)$ is such that $W'_{2\psi(e)} = R_{\psi(e)}$, and W'_2 has the recursion property.

The theorem has as an immediate corollary that a maximal enumeration must have the recursion property, since any recursion enumeration gets translated into it.

We mentioned that not every fixed-point enumeration is a recursion enumeration. A fixed-point enumeration which is maximal is also a recursion enumeration.

As we said, most of the results in recursion theory that use W really only depend on the fact that there is an enumeration with certain properties (specifically, the substitution property, the recursion property, and the recursion property with parameters); as far as recursion theory is concerned, little is gained by showing that the particular enumeration W has these properties, since a cooked up enumeration with those properties will in general do the job as well.

Lecture XVI

The Tarski-Mostowski-Robinson Theorem

Recall from lecture X that if Γ is a set of true sentences in the language of arithmetic, then every r.e. set is weakly representable in Γ. Specifically, if A(x) is a formula of RE that defines a set S, then Q ⊃ A(x) weakly represents S in Γ: if n ∈ S, then Q fi A($\mathbf{0}^{(n)}$), so *a fortiori* Γ, Q fi A($\mathbf{0}^{(n)}$), and so by the deduction theorem, Γ fi Q ⊃ A($\mathbf{0}^{(n)}$); if, on the other hand, Γ fi Q ⊃ A($\mathbf{0}^{(n)}$), then Q ⊃ A($\mathbf{0}^{(n)}$) is true (since it follows from a set of true sentences), and Q is true, so A($\mathbf{0}^{(n)}$) is true and therefore n ∈ S. It follows that every r.e. set is 1-1-reducible to the set of theorems of Γ; if Γ is r.e., then the set of theorems of Γ is 1-complete. But whether or not Γ is r.e., Γ is undecidable.

Alfred Tarski, Andrzej Mostowski, and Raphael Robinson generalized this result, developing a technique for showing that various theories are undecidable. The theorem summing up this technique that we will state here, which says that certain theories are 1-1 complete, can be reasonably attributed to Bernays. We will call our basic result the 'Tarski-Mostowski-Robinson theorem', since it is essentially due to them, although Myhill and Bernays deserve credit for stating it in this form.

The basic idea behind the proof of the Tarski-Mostowski-Robinson theorem is to weaken the hypothesis that Γ be true (in the standard model of the language of arithmetic) in such a way that the argument of the last paragraph still goes through. We shall prove the theorem in stages, finding successively weaker hypotheses.

First, note that we can find a slight weakening of the hypothesis already. We already know that if Γ is a true theory in a language with two three-place predicates A and M for addition and multiplication (or, from an exercise, even with a single three-place predicate for exponentiation) then Γ is 1-complete. Weakening the hypothesis still further: suppose Γ is a theory in some language L' which contains the language L of arithmetic (or simply the language {A, M}) but contains extra vocabulary. Then the reasoning still goes through, as long as Γ has a model whose restriction to L is the standard model of L (or isomorphic to it). To see this, we need only verify that if Γ fi Q ⊃ A($\mathbf{0}^{(n)}$) then n ∈ S (where A(x) defines S in RE and 'Q' is some appropriate formulation of Q if the language considered is {A,M}). So suppose Γ fi Q ⊃ A($\mathbf{0}^{(n)}$) and I is a model of Γ whose restriction to L is the standard model of L. Then Q ⊃ A($\mathbf{0}^{(n)}$) is true in I and therefore in the standard model, since Q ⊃ A($\mathbf{0}^{(n)}$) is a sentence of L. So we have the result that if Γ is a theory in some language L' which contains the language L of arithmetic (or simply which contains {A,M}) and Γ has a model whose restriction to L is the standard model of L (or isomorphic to it) then Γ is 1-complete.

Even in this form, the result is difficult to apply in practice, since, first, some theories we

might want to apply it to are formulated in languages which do not contain the language of arithmetic; and second, few if any interesting theories whose languages extend the language of arithmetic have models whose restriction to this language is isomorphic to the structure of the natural numbers. The full Tarski-Mostowski-Robinson theorem will show that the theories of various sorts of algebraic structures (e.g. groups, rings, etc.) are undecidable; to use the form of the theorem just mentioned to show that the theory of some class C of structures is undecidable, the structure <**N**, 0, ', +, ·> must be a member of C, and few if any such classes that have actually been studied include this structure. For example, we cannot yet show that the theory of rings is undecidable, since the natural numbers under addition and multiplication do not form a ring, as they are not closed under additive inverse.

However, the integers do form a ring, and moreover they include the natural numbers as a part. This suggests another weakening of the hypothesis that Γ is true: roughly, we shall show that as long as Γ has a model I such that the natural numbers under addition and multiplication are a submodel of I and they can be "picked out" using the language of Γ, then Γ is 1-complete. Actually, we shall prove a result that turns out to be equally powerful: we shall show that if Γ is a theory in some first-order language L, and L' is a language obtained from L by adding finitely many constants, and Γ has a model I in the language L' such that the natural numbers under addition and multiplication (or a structure isomorphic to this) are definable as a submodel of I, then the set of theorems of Γ in L is a set to which all r.e. sets are 1-1 reducible.

Tarski-Mostowski-Robinson Theorem: Let Γ be a theory in some first-order language L, and let L' be obtained from L by adding finitely many constants (possibly 0). Suppose Γ has a model I in the language L' such that the natural numbers are definable as a submodel of I. Then the set of theorems of Γ in L is a set to which all r.e. sets are 1-1 reducible.

The proof of the theorem will occupy us for the most part of the rest of this lecture.

As a first step to spelling the content of the theorem out, let Γ be a theory in some first-order language L, and let L' be obtained from L by adding finitely many constants. What does it mean to say that Γ has a model I in L' such that the natural numbers under addition and multiplication (or a structure isomorphic to this) are definable as a submodel of I? It means that there is a model I in L' of Γ and there are formulae N'(x), A'(x, y, z), and M'(x, y, z) of L' such that the structure $<I_{N'},I_{A'},I_{M'}>$ is the structure of the natural numbers under addition and multiplication (or a structure isomorphic to it), where $I_{N'}$={a: a satisfies N'(x) in I}, $I_{A'}$={<a,b,c> ∈ $I_{N'}^3$: <a,b,c> satisfies A'(x, y, z) in I}and $I_{M'}$={<a,b,c> ∈ $I_{N'}^3$: <a,b,c> satisfies M'(x, y, z) in I}.

If N', A' and M' are not already primitive predicate letters in L, we add corresponding predicates N, A, M to L and sentences (x)(N(x) ≡ N'(x)), (x)(y)(z)(A(x,y,z) ≡ A'(x,y,z)), (x)(y)(z)(M(x,y,z) ≡ M'(x,y,z)) as "definitional" axioms to Γ. We also add symbols for zero and successor and definitional axioms for them, as follows: (x)(N(x) ⊃ (x=**0** ≡ A(x,x,x)))

for zero, $(y)(x)(N(x) \wedge N(y) \supset (x'=y \equiv (\exists w)(N(w) \wedge (z)\sim A(z,z,z) \wedge M(w,w,w) \wedge A(x,w,y))))$ for successor. The resulting theory is the set of consequences in $L' \cup \{N,\mathbf{0},',A,M\}$ of Γ plus the finite set D of definitional axioms.

Now, if B is a sentence, let B_N be the result of restricting all of B's quantifiers to N; that is, B_N comes from B by replacing $(\exists x)...$ by $(\exists x)(N(x) \wedge ...)$ throughout and $(x)...$ by $(x)(N(x) \supset ...)$ throughout. B_N is called the *relativization* of B to N. It is simple enough to show that B_N holds in I iff B holds in the submodel of I defined by N. Call Q_N the theory whose theorems are the consequences of a conjunction of the relativizations of the axioms of Q to N.

We then know that for every r.e. set S, if B(x) defines S in RE, then $B_N(x)$ is such that (1) $B_N(x)$ defines S on the natural numbers (or the copy of S in the structure defined by N) and (2) for all n, Q_N fi $B_N(\mathbf{0}^{(n)})$ iff $n \in S$. First, $B_N(x)$ clearly defines S (or the copy of S in the structure defined by N). Now, suppose that $n \in S$. Then for the usual reasons, fi $Q \supset B(\mathbf{0}^{(n)})$; it is easy enough to show that fi $Q_N \supset B_N(\mathbf{0}^{(n)})$, and therefore that Q_N fi $B_N(\mathbf{0}^{(n)})$. Now suppose that Q_N fi $B_N(\mathbf{0}^{(n)})$. Then $B_N(\mathbf{0}^{(n)})$ is true in the natural numbers (or in the structure defined by N), and so $n \in S$.

Now consider the theory $\Gamma+D+Q_N$, the set of consequences in the language $L' \cup \{N,\mathbf{0},',A,M\}$ of Γ plus the finite set D of definitional axioms, plus Q_N. Then for every r.e. set S, if B(x) defines S in RE, then $B_N(x)$ defines S (or the copy of S in the structure defined by N), and for all n, (i) if $n \in S$ then $\Gamma+D+Q_N$ fi $B_N(\mathbf{0}^{(n)})$ (by the same reasoning as in the preceding paragraph) and (ii) if $\Gamma+D+Q_N$ fi $B_N(\mathbf{0}^{(n)})$ then $n \in S$, for suppose $n \notin S$; B(x) defines S, so $B(\mathbf{0}^{(n)})$ is false, and so $B_N(\mathbf{0}^{(n)})$ is false in the structure defined by N, and hence in I; but $\Gamma+D+Q_N$ are true in I, so not $\Gamma+D+Q_N$ fi $B_N(\mathbf{0}^{(n)})$. (i) and (ii) establish, in other words, that $B_N(x)$ weakly represents S (or its copy) in $\Gamma+D+Q_N$.

Then, by the deduction theorem, for all n, $n \in S$ iff $\Gamma+D$ fi $Q_N \supset B_N(\mathbf{0}^{(n)})$. This indicates how to prove, using the familiar arguments employing the recursiveness of substitution, that S is 1-reducible to the set of theorems of $\Gamma+D$, i.e. that there is a 1-1 recursive function ψ such that $n \in S$ iff $\psi(n)$ is a Gödel number of a theorem of $\Gamma+D$; $\psi(n)$ will be a Gödel number of a sentence of the form $(x)(x=\mathbf{0}^{(n)} \supset (Q_N \supset B_N(x))$. This shows that the set of theorems of $\Gamma+D$ is 1-complete if it is r.e.

But we have not shown yet that every r.e. set is 1-reducible to the set of theorems of Γ (in the language L). Let us first see how the proof of this will go if we suppose that L and L' are the same, i.e., that no extra constants are aded to L, so that the definitional axioms only contain symbols from L and $\Gamma+D$ is a theory in $L \cup \{N,\mathbf{0},',A,M\}$. Intuitively, the addition of the new non-logical symbols by means of definitions does not add expressive power to L. More precisely, if B is a theorem of $\Gamma+D$ then there is a translation B* of B into L, obtained by replacing "definienda" by "definientes" throughout, such that B* is a theorem of Γ (the converse trivially obtains). In other words, there is a function ϕ such that if m is a Gödel number of a sentence of the language $L \cup \{N,\mathbf{0},',A,M\}$, $\phi(m)$ is a Gödel number of its translation into L. If we could show that we may require ϕ to be recursive and 1-1, then we

would have shown that every r.e. set is 1-reducible to the set of theorems of Γ (in the language L), because the composition of ϕ and ψ would be 1-1 and recursive, and would reduce S to the set of theorems of Γ.

In fact, we will show how to define directly, for each r.e. set S, a function β whose value for n is (a Gödel number of) a translation of $(x)(x=\mathbf{0}^{(n)} \supset (Q_N \supset B_N(x))$ (where $B_N(x)$ is as before). It is clear that the parts Q_N and $B_N(x)$ of one such formula are (recursively) translatable into appropriate formulae of L (a fixed translation Q* of the conjunction of the axioms of Q and a fixed formula B*(x) defining S (or a copy of it) in L). The part $x=\mathbf{0}^{(n)}$ is the only one that depends on n. Recall that L need not contain symbols for successor and zero. Now, clearly there is some formula $D_n(x)$ of L, obtained by repeated applications of the definitions for **0** and ' and Russell's trick, and such that the sentence $(x)(x=\mathbf{0}^{(n)} \equiv D_n(x))$ is a theorem of Γ+D. To obtain $D_n(x)$ in this way we would need a cumbersome application of the generated sets theorem. But we can obtain an appropriate formula $E_n(x)$ in a simpler fashion using the uniformization theorem. Notice that there must be a formula $E_n(x)$ of L such that $(x)(x=\mathbf{0}^{(n)} \equiv E_n(x))$ is a theorem of D alone (intuitively, we only need the definitions to prove an appropriate equivalence). But D is finite, so its set of theorems is r.e. Therefore the relation R={(n,m): m is a Gödel number of a formula E(x) and E(x) is in L and $(x)(x=\mathbf{0}^{(n)} \equiv E(x))$ is a theorem of D} is an r.e. relation, for the familiar reasons. Clearly for all n there is an m such that R(n,m). So R can be uniformized to a recursive function α such that $\alpha(n)$ is a Gödel number of a formula E_n such that $(x)(x=\mathbf{0}^{(n)} \equiv E_n(x))$ is a theorem of Γ+D (in fact, of D alone); α is clearly 1-1, because otherwise $(x)(x=\mathbf{0}^{(p)} \equiv x=\mathbf{0}^{(q)})$ for some p, q, $p \neq q$ would be a theorem of Γ+D, which is impossible, since that sentence must be true in a model isomorphic to the natural numbers, and any such model makes that sentence false.

Finally, $\beta(n)$ will be definable in RE using concatenation as e.g. the least Gödel number of $(x)(E_n(x) \supset (Q^* \supset B^*(x))$, where $E_n(x)$ is cashed out in the definition of β in RE by means of α. β is thus clearly recursive and 1-1 (since α is). β 1-reduces S to the set of theorems of Γ, since for all n, $n \in S$ iff $\beta(n)$ is a Gödel number of a theorem of Γ.

But we will have proved the Tarski-Mostowski-Robinson theorem only when we prove the same result without assuming that L' is equal to L. So far our proof only establishes (or can be minimally modified to establish) that every r.e. set is 1-reducible to the set of theorems of Γ in L', not in L. But we can easily show how to obtain recursively and in a 1-1 fashion, for a formula of the form $(x)(E_n(x) \supset (Q^* \supset B^*(x))$ possibly containing extra constants, a formula C of L (thus without extra constants) such that Γ fi $(x)((E_n(x) \supset (Q^* \supset B^*(x)) \equiv C)$. Since Γ is a theory in L, any property of the extra constants provable from Γ must be provable in Γ for arbitrary objects; thus, if $F(a_1,...,a_n)$ is provable from Γ, $(y_1)...(y_n)F(y_1,...,y_n)$ (where $y_1,...,y_n$ are the first variables that do not occur in $F(a_1,...,a_n)$) must be provable from Γ.

This concludes our proof of the Tarski-Mostowski-Robinson theorem.

(It may be remarked that we could have proved a weaker result which does not mention

extra constants at all. We will see how the addition of extra constants can be profitably applied in an exercise.)

Both Bernays and Myhill stated a theorem whose statement is closely related to the one we have given, although Myhill (and perhaps also Bernays) did not have an appropriate justification for it. The theorem they stated says that if a theory has a model with a definable submodel which is a model of Q, then the theory is 1-1 complete. This theorem is true (see the exercises) but it is harder to prove than our theorem. What Myhill and Bernays proved, essentially, was this theorem under the hypothesis that the theory is ω-consistent.

The Tarski-Mostowski-Robinson theorem can be applied to show that several algebraic theories are undecidable. Among them, the elementary theories of rings, commutative rings, integral domains, ordered rings, ordered commutative rings (all with or without unit), the elementary theory of fields, etc. The proof for the theory of rings is given as an exercise.

Despite its simplicity, the Tarski-Mostowski-Robinson theorem is a very striking result, since it states that for a theory to be undecidable, it is enough that it have just one model in which the natural numbers are definable as a submodel. Part of the reason it is so striking is that it is commonly applied to theories (like the theory of rings) for which there is no single standard interpretation. However, it is really no different in principle from the result that Q is undecidable. Q also has many different interpretations, but we tend to think of one particular interpretation as "standard" or "intended", so we are less surprised when that interpretation is used to show that Q is undecidable; nonetheless, mathematically speaking, using the standard interpretation of Q to show that it is undecidable is no different from using the fact that the integers form a ring to show that the theory of rings is undecidable.

If we have already shown that a given theory is decidable and that I is a model of that theory, it will follow that the set of natural numbers is not definable in I. For example, consider the model in the language of arithmetic whose domain is the real numbers. It is a famous theorem of Tarski that the first-order theory of this model (i.e. the set of sentences true in this model) is decidable; it follows from the Tarski-Mostowski-Robinson theorem that the set of natural numbers is not definable in this model. Similar remarks apply to the complex numbers. This also illustrates the fact that, for the theorem to apply, the formula that picks our the natural numbers must be a formula of the object language, since in the metalanguage we can certainly pick out the natural numbers from the real numbers.

Note also that the theorem relates the undecidability of Γ in L to the existence of a certain kind of model of Γ in a possibly larger language L'. It is important to notice that L' is only allowed to differ from L' by the addition of finitely many constants; the theorem does not hold if we allow L' to have additional predicates or function symbols as well. To see this, recall that the first-order theory Γ of the reals in the language L of arithmetic is decidable. However, letting $L' = L \cup \{N\}$ (where N is any unary predicate), we see that Γ is undecidable *in L'*: simply let I be the model for L whose domain is the set of reals, etc., let I' be the expansion of I to L' in which N is interpreted as applying to the natural numbers, and apply the Tarski-Mostowski-Robinson theorem.

Exercises

1. A classical theorem of elementary number theory says that every positive integer is the sum of four squares. Use this to prove that the elementary theory of rings is 1-1 complete. (Remark: For those who know about such things, the same argument can be used to prove that the elementary theories of commutative rings with or without unit, of integral domains, of ordered rings and ordered integral domains, etc. are 1-1 complete. It is more difficult to prove the 1-1 completeness of the elementary theory of fields, which uses a similar but more difficult method.)

2. (a) Show that the theorem that every maximal enumeration is a recursion enumeration can be proved using the method employed in the lectures to prove the self-reference lemma (with the recursion theorem for W as a special case). Remember that a maximal enumeration is one with the substitution property.

(b) Formulate an appropriate version of the recursion property with parameters, and prove that the diagonalization of any maximal subnumeration has the recursion property with parameters.

3. Recall the recursively inseparable sets S_1 and S_2 from the lectures.

(a) Let C be an r.e. set containing S_1 and disjoint from S_2. Prove that C is completely creative. Hint: Let A(x, y, z) be the r.e. relation $(y \in C \wedge z = \mathbf{0}') \vee (W(x, y) \wedge z = \mathbf{0})$. Let $\psi(x, y)$ be a uniformization of A(x, y, z). Prove that there is a recursive function χ such that $\psi(x, y) = \Phi(\chi(x), y)$, for all x, y. Prove that χ is a completely creative function for C.

(b) Give an example of a formula A(x) in the language L of arithmetic such that if Γ is any consistent r.e. extension of Q in L, then A(x) weakly represents a completely creative set in Γ. (A(x) need not represent the same completely creative set in all these systems.)

(c) Prove that if Γ is as above, every r.e. set is weakly representable in Γ.

(d) Prove that if Γ is as above, the set of all theorems of Γ is one-to-one complete.

Comment: this finally shows that the results we stated before under the hypothesis that Γ extends Q and is ω-consistent, concerning weak representability, 1-completeness, etc. all hold if ω-consistency is weakened to consistency. (Or almost all: this does not show that the result about *nice* weak representability still holds. This can also be proved, but requires another argument.) Rosser's work gave a start for this, but it took several decades to reach the point of this exercise.

(e) Use the results above to show how to prove the Tarski-Mostowski-Robinson results, stated in class under the hypothesis that Γ has a model with a definable submodel

isomorphic to the standard model of L, under the weaker hypothesis that Γ has a model with a definable submodel which is a model of Q.

Comment: Tarski, Mostowski and Robinson, as said in class, used a less model-theoretic formulation. However, their work would have implied the result in (e) with the conclusion of undecidability only. I know of no significant application to a specific theory, however, where the generalization to models of Q is really useful.

4. (a) An r.e. set S is *creative* if there is a recursive function ψ such that whenever W_x is disjoint from S, $\psi(x) \notin S \cup W_x$. Prove that every creative set is many-one complete. Hint: Let S* be any r.e. set. Prove that there is a recursive function χ such that, for all x,

$$W_{\chi(x)} = \begin{cases} \{\psi(\chi(x))\} \text{ if } x \in S^* \\ \emptyset \text{ if } x \notin S^* \end{cases}$$

(b) Conclude from what we have done so far that the concepts creative, completely creative, and many-one complete are equivalent. Also show that the concepts 1-1 complete, 1-1 completely creative, and 1-1 creative (defined in the obvious way) are equivalent. Later on it will turn out that all six concepts are equivalent.

(c) Another equivalent concept: prove that a set S satisfies the effective form of Gödel's theorem, as defined for nonrecursive r.e. sets, iff S is creative.

Comment: all of the concepts $\leq_m$, $\leq_1$, m-complete, 1-complete, creative, and simple are due to Post. Many theorems relating them are also due to Post, as (essentially) is the connection between creativeness and Gödel's theorem (which inspired the term "creative"). Other important properties of these concepts were proved by Myhill.

5. Show that the set of all valid formulae in the first-order language with one two-place predicate letter and no others, is 1-1 complete. Also show that the elementary theory of one irreflexive relation and the elementary theory of one asymmetric relation are 1-1 complete. Sketch of the method: consider a certain structure with set-membership as the only relation between elements of the structure. Set membership is irreflexive and asymmetric. The structure will consist of the natural numbers, the sets of natural numbers, the sets whose elements are natural numbers and sets of natural numbers, and so on, through all finite levels. Kuratowski defined the ordered pair $\langle x, y\rangle$ as $\{\{x\}, \{x, y\}\}$. Prove that this has the property of a pairing function: that is, if $\langle x_1, x_2\rangle = \langle y_1, y_2\rangle$ then $x_1 = y_1$ and $x_2 = y_2$. An ordered triple etc. can be defined in terms of ordered pairs. This pairing function is an important tool in the proof. The proof is much simpler if one realizes that definitions with extra constants are allowed.

Comment: some of you may know that a model of the natural numbers together with definitions of + and · can be defined in set theory. This fact could have been used to do this exercise, but the method given above presupposes much less prior background, and shows that this is not needed.

6. A *reduction class* is a recursive class of formulae in a first-order language such that there is an effective mapping ψ of arbitrary formulae of the full language of first-order logic (i.e. the language of first-order logic with all predicates and constants) into formulae of the class such that a formula A of the full language is valid iff ψ(A) is valid. Reduction classes were an active topic of research even before the development of recursion theory.

(a) A recursive class C of formulae is a reduction class iff the set of all (Gödel numbers of) valid formulae in C is . . . Fill in the dots with a concept already defined in this course, and prove the correctness of your answer.

(b) Give a non-trivial example of a reduction class, using the answer to part (a).