0. INTRODUCTION

Let T be a sufficiently strong theory formalized in the language L_A of (first order) arithmetic. Following Gödel, we want to show that there is a sentence φ of L_A which is true (of the natural numbers) but not provable in T. Gödel's idea was to achieve this by constructing φ in such a way that

(*) φ is true if and only if φ is not provable in T.

Then, assuming (for simplicity) that all theorems of T are true, we are done. For, suppose φ is provable in T. Then, by (*), φ is not true and so, by hypothesis, φ is not provable in T. Thus, φ is not provable in T. But then, by (*), φ is true.

One way to achieve (*) is to find a sentence φ which, in some sense, "says" of itself that it is not provable in T. There are then three major difficulties. First of all, the sentences of L_A deal with natural numbers, they do not deal with syntactical objects such as sentences (of a formal language), proofs, etc. Secondly, even if some of the sentences of L_A can, somehow, be understood as dealing with syntactical objects, it is not clear that it is possible to "say" anything about provability (in T) using only the means of expression available in L_A. And, finally, even if this is possible, there may be no sentence which "says" of *itself* that it isn't provable.

Gödel, however, was able to overcome these difficulties. The first problem is solved by assigning natural numbers to syntactical expressions in a certain systematic way. This is sometimes called a Gödel numbering, and the number assigned to an expression, the Gödel number of that expression. Thus, the numeral of the number assigned to an expression can be regarded as a name of that expression and certain number theoretic statements can be regarded as statements about syntactical objects. (In what follows "φ is a formula", "p is a proof", etc. are short for "φ is the Gödel number of a formula", "p is the Gödel number of a proof", etc.)

To overcome the second difficulty Gödel (re)invented the primitive recursive functions (sets, relations). He showed that a number of crucial properties of (Gödel numbers of) expressions, such as that of being a (well–formed) formula, are primitive recursive. In particular, Gödel showed that, if the set of axioms of T is primitive recursive, this is also true of the relation $PRF_T(\varphi,p)$: p is a proof of the sentence φ in T. φ is provable in T, $PR_T(\varphi)$, if and only if $\exists p PRF_T(\varphi,p)$. This property, however, is not (primitive) recursive.

Gödel then went on to prove that all primitive recursive functions (sets, relations) are definable in L_A. Thus, in particular, there is a formula $Prf_T(x,y)$ of L_A defining $PRF_T(k,m)$. But then $Pr_T(x) := \exists y Prf_T(x,y)$ defines $PR_T(k)$. (In what follows we write $T\vdash \varphi$ for $PR_T(\varphi)$.)

Gödel, however, proved more and this is crucial: for every sentence φ, $T\vdash \varphi$ if and only if $T\vdash Pr_T(\varphi)$. (This is the first time we use the assumption that T is sufficiently strong; but, of course, if T isn't, T is incomplete for that reason.)