THE JOURNAL OF SYMBOLIC LOGIC
Volume 22, Number 1, March 1957

LANGUAGES IN WHICH SELF REFERENCE IS POSSIBLE[1]

RAYMOND M. SMULLYAN

1. Introduction. This paper treats of semantical systems $\boldsymbol{S}$ of sufficient strength so that for any set W definable in $\boldsymbol{S}$ (in a sense which will be made precise), there must exist a sentence X which is true in $\boldsymbol{S}$ if and only if it is an element of W.[2] We call such an X a *Tarski* sentence for W. It is the sentence which (in a purely extensional sense) says of itself that it is in W.[3] If W is the set of all expressions not provable in some syntactical system C, then X is the Gödel sentence which is true (in $\boldsymbol{S}$) if and only if it is not provable (in C). We provide a novel method for the construction of these sentences, which yields sentences particularly simple in structure. The method is applicable to a variety of systems, including a form of elementary arithmetic, and some systems of protosyntax self applied.[4] In application to the former, we obtain an extremely simple and direct proof of a theorem, which is essentially Tarski's theorem that the truth set of elementary arithmetic is not arithmetically definable.

The crux of our method is in the use of a certain function, the 'norm' function, which replaces the classical use of the *diagonal* function. To give a heuristic idea of the norm function, let us define the norm of an expression E (of informal English) as E followed by its own quotation. Now, given a set W (of expressions), to construct a sentence X which says of itself that it is in W, we do so as follows:

W contains the norm of 'W contains the norm of.' This sentence X says that the norm of the expression 'W contains the norm of' is in W. However,

---

Received May 3, 1956.

[1] I wish to express my deepest thanks to Professor Rudolph Carnap, of the University of California at Los Angeles, and to Professor John Kemeny and Dr. Edward J. Cogan, of Dartmouth College, for some valuable suggestions. I also wish to thank the referee for some very helpful revisions.

[2] By a semantical system $\boldsymbol{S}$, we mean a set E of *expressions* (strings of signs), together with a subset S of expressions called *sentences* of $\boldsymbol{S}$ (determined by a set of rules of formation), together with a subset T of S, of elements called *true* sentences of $\boldsymbol{S}$ (determined by a set of 'rules of truth' for $\boldsymbol{S}$).

[3] (At the referee's suggestion) — In an extensional system, the only way we can translate the meta-linguistic phrase '$\boldsymbol{X}$ says that $\boldsymbol{X} \in W$' is by the phrase '$\boldsymbol{X}$ is true if and only if $\boldsymbol{X} \in W$'. Thus the requirement for $\boldsymbol{X}$ to be a Tarski sentence for W, is exceedingly weak; any sentence $\boldsymbol{X}$ which is either both true and in W, or false and not in W, will serve. However, this is as much as we need of a Tarski sentence (for undecidability results). If we were considering an *intensional* system, then we would define a Tarski sentence for W as a sentence $\boldsymbol{X}$ which not only is true if and only if $\boldsymbol{X} \in W$, but which actually expresses the proposition that $\boldsymbol{X} \in W$.

[4] The former will be carried out in this paper and the latter in a forthcoming paper, *Systems of protosyntax self applied.*

the norm of this expression is X itself. Hence X is true if and only if $X \in W$.[5]

This construction is much like one due to Quine.[6] We carry it out for some formalized languages. In section 2, which is essentially expository, we construct a very precise, though quite trivial semantical system $\boldsymbol{S}_P$, which takes quotation and the norm function as primitive. The study of this system will have a good deal of heuristic value, inasmuch as $\boldsymbol{S}_P$, despite its triviality, embodies the crucial ideas behind undecidability results for deeper non-trivial systems. We then consider, in section 3, the general use of the norm function, and we finally apply the results, in section 4, to a system $\boldsymbol{S}_A$, which is a formal variant of elementary arithmetic. This variant consists of taking the lower functional calculus with class abstractors, rather than quantifiers, as primitive. This alteration, though in no way affecting the strength of the system, nevertheless makes possible the particularly simple proof of Tarski's or Gödel's theorem, since the arithmetization of substitution can thereby be circumvented quite simply.

By the norm of an expression E (of $\boldsymbol{S}_A$) we mean E followed by its own Gödel *numeral* (i.e., the numeral designating its Gödel number). Now, given any set W of expressions whose set of Gödel numbers is arithmetically definable, we show quite easily the existence of an expression H of class abstraction, such that for any expression E, H followed by the Gödel *numeral* of E is a true sentence if and only if the *norm* of E is in W. Then, if we follow H by its own Gödel numeral h, the resulting sentence Hh (which is the norm of H) is true if and only if it is in W. This is a rough sketch of our procedure.

2. The preliminary system $\boldsymbol{S}_0$ and the semantical system $\boldsymbol{S}_P$. In this section, we formalize the ideas behind the preceding heuristic account of the norm function. For convenience, we first construct a preliminary system $\boldsymbol{S}_0$, whose expressions will be built from the three signs 'φ', '*', and 'N'. The second sign will serve as our formal quotation mark, since we reserve ordinary quotation marks for meta-linguistic use. The sign 'N' will be endowed with the same meaning as 'the norm of.' The sign 'φ' will be an undefined predicate constant. For any property (set) P of expressions of $\boldsymbol{S}_0$, we then define the semantical system $\boldsymbol{S}_P$ by giving a rule of truth for $\boldsymbol{S}_P$. For any P, 'φ' will be interpreted in $\boldsymbol{S}_P$ as designating P.

---

[5] In contrast with this construction, let us define the *diagonalization* of E as the result of substituting the quotation of E for all occurrences of the variable 'x' in E. Then the following Tarski sentence for W (when formalized) is the classical construction: W contains the diagonalization of 'W contains the diagonalization of x'. This latter construction involves *substitution* (inherent in diagonalization), whereas the norm function involves *concatenation* (the norm of E being E *followed by* its quotation), which is far easier to formalize (cf. Section 5).

[6] 'Yields falsehood when appended to its quotation' yields falsehood when appended to its quotation. This is Quine's version of the famous semantical paradox.

Signs of S_0: φ, *, N.

Preliminary definitions: (1) By an *expression* (of S_0) we mean any string built from the three signs of S_0. (2) By the (formal) quotation of an expression, we mean the expression surrounded by stars. (3) By the *norm* of an expression, we mean the expression followed by its own (formal) quotation.

Formation rules for (individual) designators

(1) The quotation of any expression is a designator.

(2) If E is a designator, so is ⌜NE⌝ (i.e., 'N' followed by E).

Alternative definition

(1)′ A designator is an expression which is either a quotation (of some other expression) or a quotation preceded by one or more 'N's.

Rules of designation in S_0

R1. The quotation of an expression E designates E.

R2. If E_1 designates E_2, then ⌜NE_1⌝ designates the *norm* of E_2.

Definition of a sentence of S_0

(1) A sentence of S_0 is an expression consisting of 'φ' followed by a designator.

The semantical system S_P

For any property P, we define the semantical system S_P as follows:

(1) The rules for designators, designation and sentence formation in S_P are the same as in S_0.

(2) The rule of truth for S_P is the following:

R3. For any designator E, ⌜φE⌝ is true in $S_P \underset{\text{df}}{=}$ the expression designated by E (in S_P) has the property P.

THEOREM 2.1. There exists an expression of S_0, which designates itself.

PROOF. '*N*' designates 'N' [By Rule 1].

Hence 'N*N*' designates the norm of 'N' [By Rule 2] which is 'N*N*'.

Thus 'N*N*' designates itself.

THEOREM 2.2. There exists a sentence G of S_0, such that for any property P, G is true in $S_P \iff G$ has the property P.

PROOF. 'N*φN*' designates 'φN*φN*' [By R1 and R2].

Thus G, viz., 'φN*φN*' is our desired sentence.

REMARK. G is, of course, the formalized version of 'W contains the norm of 'W contains the norm of'.' 'φ' is but an abbreviation of 'W contains,' and 'N' abbreviates 'the norm of.'[7]

---

[7] If we wished to construct a miniature system L_P which formalizes the diagonal function in the same way as S_P does the norm function, we take four signs, viz., 'φ', '*', 'D', 'x', and the rules R_1, (same as S_P), R_2: If E_1 designates E_2, then ⌜DE_1⌝ designates the diagonalization of E_2 (i.e., the result of replacing each occurrence of 'x' in E_2 by the quotation of E_2), R_3: If E_1 designates E_2, then ⌜φE_1⌝ is a sentence and is true in L_P if and only if E_2 has the property P. Then the expression of Theorem 2.1, which designates itself, is 'D*Dx*', and the Tarski sentence (of Theorem 2.2) for P is 'φD*φDx*'.

COROLLARY 2.3. P cannot be coextensive with the set of all false (non-true) sentences of $\boldsymbol{S}_P$, nor is P coextensive with the set of all expressions of $\boldsymbol{S}_P$ which are not true sentences of $\boldsymbol{S}_P$.

2.4. A STRONGER FORM OF THEOREM 2.2. By a *predicate* we mean either 'φ' or 'φ' followed by one or more 'N's'.

We say that an expression E *satisfies* a predicate H (in $\boldsymbol{S}_P$) if H followed by the quotation $\ulcorner$*E*$\urcorner$ of E, is true in $\boldsymbol{S}_P$. Lastly, we say that a set W of expressions of $\boldsymbol{S}_0$ is *definable* (in $\boldsymbol{S}_P$) if there exists a predicate H which is satisfied by all and only those expressions which are in W.

It is worth noting at this point, that if E_1 designates E, then E satisfies H if and only if $\ulcorner HE_1\urcorner$ is true. This follows from R3 by induction on the number of N's occuring in H.

For any set W, we let $\eta(W) \underset{\text{df}}{=}$ set of all expressions whose norm is in W.

LEMMA 2.5. If W is definable in $\boldsymbol{S}_P$, then so is $\eta(W)$.

PROOF. Let H be the predicate which defines W (i.e., which is satisfied by just those elements which are in W). Then H followed by 'N' will be satisfied by precisely those elements which are in $\eta(W)$. Thus $\eta(W)$ is definable (in $\boldsymbol{S}_P$).

We can now state the following theorem, of which Theorem 2.2 is a special case.

THEOREM 2.6. For *any* set W definable in $\boldsymbol{S}_P$, there is a sentence X which is true in $\boldsymbol{S}_P$ if and only if $X \in W$.

PROOF. Assume W is definable. Then so is $\eta(W)$ [by Lemma]. Hence there exists a predicate H such that for any expression E, $\ulcorner$H*E*$\urcorner$ is true (in $\boldsymbol{S}_P$) $\iff$ E$\in\eta(W)$

$$\iff \ulcorner \text{E*E*}\urcorner \in W.$$

Taking E = H, $\ulcorner$H*H*$\urcorner$ is true $\iff \ulcorner$H*H*$\urcorner \in W$.

Thus X, viz., $\ulcorner$H*H*$\urcorner$, is our desired sentence.

REMARK. Theorem 2.6 says (in view of the truth functionality of the biconditional) no more nor less than this: each set definable in $\boldsymbol{S}_P$ either contains some truths of $\boldsymbol{S}_P$ or lacks some falsehoods.

COROLLARY 2.7. The set of false sentences of $\boldsymbol{S}_P$ is not definable in $\boldsymbol{S}_P$, nor is the complement (relative to the set of all *expressions* of $\boldsymbol{S}_P$) of the set of true sentences of $\boldsymbol{S}_P$ definable in $\boldsymbol{S}_P$.

COROLLARY 2.8. Suppose we extend $\boldsymbol{S}_P$ to the enlarged semantical system $\boldsymbol{S}'_P$ by adding the new sign '$\sim$', and adding the following two rules:
R4. If X is a sentence, so is $\ulcorner\sim X\urcorner$.
R5. $\ulcorner\sim X\urcorner$ is true in $\boldsymbol{S}'_P \iff X$ is not true in $\boldsymbol{S}'_P$.
Then in this system $\boldsymbol{S}'_P$, the truth set of $\boldsymbol{S}'_P$ is not definable.

PROOF. For $\boldsymbol{S}'_P$ has the property that the complement of any set definable in $\boldsymbol{S}'_P$ is again definable in $\boldsymbol{S}'_P$, since if H defines W, then $\ulcorner\sim$H$\urcorner$ defines

the complement of W. Hence the truth set is not definable, since its complement is not definable by Corollary 2.7.

REMARK. S'_P is about as simple a system as can be constructed which has the interesting property that the truth set of the system is not definable within the system and that, moreover, any possible extension of S'_P will retain this feature. By an extension, we mean any system constructed from S'_P by possibly adding additional signs, and rules, but retaining the old rules in which, however, the word 'expression' is re-interpreted to mean an expression of the enlarged system. Likewise, if we take any extension of S_P, then, although we may greatly enlarge the collection of definable sets, none of them can possibly be co-extensive with the set of false sentences of the extension.

2.8. EXTENSION OF S_P TO A SEMANTICO-SYNTACTICAL SYSTEM S^C_P.

Suppose now that we select an arbitrary set of sentences of S_0 and call them *axioms*, and select a set of rules for inferring sentences from other sentences (or finite sets of sentences). The axioms, together with the rules of inference, form a so-called syntactical system, or calculus C. Let S^C_P be the ordered pair (S_P, C). Thus S^C_P is a mathematical system, or interpreted calculus. We let T be the set of true sentences of S_P (also called true sentences of S^C_P) and Th, the set of sentences provable in C (also called provable sentences, or theorems, of S^C_P). We already know that the complement $\overline{\mathrm{T}}$ of T (relative to the set of expressions) is not semantically definable in S^C_P (i.e., not definable in S_P); however $\overline{\mathrm{Th}}$ may well happen to be. If it is, however, then we have, as an immediate corollary of 2.6, the following miniature version of Gödel's theorem:

THEOREM 2.9. If the set $\overline{\mathrm{Th}}$ is semantically definable in S^C_P, then either some sentence true in S^C_P cannot be proved in S^C_P or some false sentence can be proved.

This situation is sometimes described by saying that S^C_P is either semantically incomplete or semantically inconsistent.

2.10. We can easily construct a system S^C_P obeying the hypothesis of theorem 2.9 as follows: Before we choose a property P, we *first* construct a completely arbitrary calculus C. Then we simply define P to be the set of all expressions not provable in C. Then 'φ' itself will be the predicate which semantically defines $\overline{\mathrm{Th}}$ in S^C_P, and the sentence G, viz., '$\varphi\mathrm{N}^*\varphi\mathrm{N}^*$', of Theorem 2.2 will be our Gödel sentence for S^C_P, which is true if and only if not provable in the system. In fact, for purposes of illustration, let us consider a calculus C with only a finite number of axioms, and no rules of inference. Thus the theorems of C are the axioms of C. Now, if G was included as one of the axioms, it is automatically false (in this system), whereas if G was left out, then it is true, by very virtue ofbeing left out. Thus, this system is, with dramatic clarity, obviously inconsistent or incomplete.

REMARK. Suppose that we take P to be the set of sentences which *are* provable in C. Then G becomes the Henkin sentence for the system $\boldsymbol{S}_P^C$, which is true in this system, if and only if G *is* provable in $\boldsymbol{S}_P^C$. Is G true in $\boldsymbol{S}_P^C$? This obviously depends on C. If, for example, we take C such that its set of axioms is null, then G is certainly both false and non-provable. An example of a choice of C (other than an obvious one in which G itself is an axiom) for which G is true is the following: We take for our single axiom A_1, the expression 'φ*φN*φN**'. We take a single rule R: If two designators E_1 and E_2 have the same designatum in $\boldsymbol{S}_0$, then $\ulcorner\varphi E_2\urcorner$ is directly derivable from $\ulcorner\varphi E_1\urcorner$. This rule is 'reasonable' in the sense that it does preserve truth in $\boldsymbol{S}_P$.

Now, is G, viz., 'φN*φN*', true in this $\boldsymbol{S}_P^C$ or not? It is true, providing it is provable. Now since 'N*φN*' and '*φN*φN**' both have the same designatum 'φN*φN*', then 'φN*φN*' is immediately derivable from 'φ*φN*φN**' by R, i.e., G is immediately derivable from A_1, hence G is provable, and hence also true.

We now consider whether or not this system $\boldsymbol{S}_P^C$ is semantically consistent. We have already observed that rule R does preserve truth, so the question reduces to whether or not A_1 is true. Well, by R3 of $\boldsymbol{S}_P$, A_1 is true precisely in case 'φN*φN*' has the property P, i.e., precisely in case G is provable, which it is.

3. Semantical Systems with predicates and individual constants. In this section, we consider any semantical system $\boldsymbol{S}$, of which certain expressions called *predicates* and certain expressions called (individual) *constants* are so related that any predicate followed by any constant is a sentence of $\boldsymbol{S}$. We also wish to have something of the nature of a *Gödel correspondence* g which will assign a unique constant $g(E)$ to each expression E so that $g(E)$ will be peculiar to E — i.e., we consider a 1 — 1 correspondence g whose domain is the set of all expressions, and whose range is a subset (proper or otherwise) of the individual constants. We shall often write '$\mathring{E}$' for '$g(E)$'. By the *norm* of E, we mean $E\mathring{E}$ (i.e., E followed by $g(E)$).[8] [9] For a set W of expressions of $\boldsymbol{S}$, by $\eta(W)$ we mean the set of all E whose norm is in W.

---

[8] The word 'norm' was suggested by the following usage: In Mathematics, when we have a function f(x, y) of two arguments, the entity f(x, x) is sometimes referred to as the norm of x — e.g., in Algebra the norm of a vector ϵ is $f(\epsilon, \epsilon)$, where f is the function which assigns to any pair of vectors the square root of their inner product. In this paper, the crucial function f (which allows us to introduce a notion of representability) is the function which assigns, to each pair (E_1, E_2) of expressions, the expression $E_1\mathring{E}_2$. Thus for this f, f(E, E) is our norm of E.

[9] Professor Quine has kindly suggested that I remark that the norm of a predicate is a sentence which says in effect that the predicate is *autological*, in the sense of Grelling (i.e., that the predicate applies to itself.)

A predicate H is said to define the set W (relative to g, understood) if W consists of all and only those expressions E such that HE is true. The following theorem, though quite simple, is basic.

THEOREM 3.1. For any set W of expressions of $\boldsymbol{S}$, a sufficient condition for the existence of a Tarski sentence for W is that $\eta(W)$ be definable.

PROOF. Suppose $\eta(W)$ is definable. Then for some predicate H and for any E $\overset{\circ}{\text{HE}}$ is true $\iff \text{E} \epsilon \eta(W)$

$$\iff \overset{\circ}{\text{E}}\overset{\circ}{\text{E}} \epsilon W.$$

Hence $\overset{\circ}{\text{H}}\overset{\circ}{\text{H}}$ is true $\iff \overset{\circ}{\text{H}}\overset{\circ}{\text{H}} \epsilon W$.

COROLLARY 3.2. Letting F be the set of non-true sentences of $\boldsymbol{S}$, and $\overline{\text{T}}$ the set of all expressions which are not true sentences, then neither $\eta(\overline{\text{T}})$ nor $\eta(\text{F})$ are definable in $\boldsymbol{S}$.

COROLLARY 3.3. If we extend $\boldsymbol{S}$ to a semantico-syntactical system $\boldsymbol{S}^C$, and if $\eta(\overline{\text{Th}})$ are definable in $\boldsymbol{S}$, then $\boldsymbol{S}^C$ is semantically incomplete or inconsistent.

Actually, to apply Corollary 3.3 to concrete situations, one would most likely show that $\eta(\overline{\text{Th}})$ is definable by showing (1) $\overline{\text{Th}}$ is definable and (2) For and set W, if W is definable, so is $\eta(W)$. Semantical systems strong enough to enjoy property (2) (which is purely a property of $\boldsymbol{S}$, rather than of C) are of particular importance. We shall henceforth refer to such systems as semantically *normal* (or, more briefly, 'normal'). Thus $\boldsymbol{S}$ is normal if, whenever W is definable in $\boldsymbol{S}$, so is $\eta(W)$. Semantical normality is, of course, relative to the Gödel correspondence g.

COROLLARY 3.4. If $\boldsymbol{S}$ is semantically normal, then

(1) There is a Tarski sentence X for each definable set.

(2) F is not definable in $\boldsymbol{S}$, nor is $\overline{\text{T}}$.

(3) If non-theoremhood of C is definable in $\boldsymbol{S}$, then $\boldsymbol{S}^C$ is semantically incomplete or inconsistent.

REMARK. The trivial systems $\boldsymbol{S}_P$ of Section 2 are semantically normal, relative to the correspondence g mapping each expression onto its quotation (the individual constants of $\boldsymbol{S}_P$ are, of course, the designators). In fact, Lemma 2.5 asserts precisely that. It is by virtue of normality that we showed the non-definability of $\overline{\text{T}}$ in $\boldsymbol{S}_P$. $\boldsymbol{S}_P$ was deliberately constructed with the view of establishing normality as simply as possible.

We now turn to a non-trivial system $\boldsymbol{S}_A$, for which we easily establish semantical normality.

4. Systems of Arithmetic. The first arithmetical system $\boldsymbol{S}_A$ which we consider is much like arithmetic in the first order functional calculus. We have numerals (names of numbers), numerical variables, the logical connectives (all definable from the primitive '$\downarrow$' of joint denial), identity,

and the primitive arithmetical operations of $\cdot$ (multiplication) and $\neg$ (exponentiation). We depart from the lower functional calculus in that, given a (well formed) formula F and a variable, e.g., 'x', we form the (class) abstract $\ulcorner x(F)\urcorner$, read 'the set of x's such that F.' We use abstracts to form new formulas in two ways, viz., (1) For a numeral N, $\ulcorner x(F)N\urcorner$ (read 'N is a member of the set of x's such that F,' or 'the set of x's such that F contains N') and (2) $\ulcorner x(F_1) = x(F_2)\urcorner$ (read 'the set of x's such that F_1 is identical with the set of x's such that F_2.') By (2) we easily define universal quantification thus: $\ulcorner(\forall x)(F)\urcorner \underset{\text{df}}{=} \ulcorner x(F) = x(x = x)\urcorner$.

A formal description of $\boldsymbol{S}_A$ now follows.

Signs of $\boldsymbol{S}_A$: x, ′, (,), $\cdot$, $\neg$, =, $\downarrow$, 1. We call these signs 'S_1', 'S_2', ..., 'S_9' respectively.

Rules of Formation, Designation and Truth

1. A numeral (string of '1's) of length n, designates the positive integer n.
2. 'x' alone, or followed by a string of accents, is a variable.
3. Every numeral and every variable is a term.
4. If t_1 and t_2 are terms, so are $\ulcorner(t_1) \cdot (t_2)\urcorner$ and $\ulcorner(t_1)\neg(t_2)\urcorner$. If t_1 and t_2 contain no variables and respectively designate n_1 and n_2, then the above new terms respectively designate $n_1 \times n_2$ and $n_1^{n_2}$.
5. If t_1 and t_2 are terms, then $\ulcorner t_1 = t_2\urcorner$ is a formula, called an atomic formula. All occurrences of variables are free. If no variables are present, then $\ulcorner t_1 = t_2\urcorner$ is a sentence, and is a true sentence if and only if t_1 and t_2 designate the same number n.
6. If F is a formula, α a variable, then $\ulcorner\alpha(F)\urcorner$ is a (class) abstract. No occurrence of α in $\ulcorner\alpha(F)\urcorner$ is free. If β is a variable distinct from α, then the free occurrences of β in $\ulcorner\alpha(F)\urcorner$ are those in F. If F contains no free variable other than α, then $\ulcorner\alpha(F)\urcorner$ is called a *predicate*, and the *abstraction* of F.
7. If H_1 and H_2 are abstracts, then $\ulcorner H_1 = H_2\urcorner$ is a formula. The free occurrences of any variable α in this formula are those of H_1 and those of H_2.
8. If α and β are variables, F_1 and F_2 formulae, and if $\ulcorner\alpha(F_1) = \beta(F_2)\urcorner$ contains no free variables, it is a sentence. It is true if and only if, for every numeral N, the result $F_1(N)$ of replacing all free occurrences of α in F_1 by N, and the result $F_2(N)$ of replacing all free occurrences of β in F_2 by N, are equivalent in $\boldsymbol{S}_A$ [i.e., are either both true in $\boldsymbol{S}_A$ or neither one true].
9. For any predicate $\ulcorner\alpha(F)\urcorner$ and numeral N, the expression $\ulcorner\alpha(F)N\urcorner$ is a sentence (as well as a formula) and is true if and only if the result F(N) of replacing all free occurrences of α in F by N, is true.
10. If F_1 and F_2 are formulae, so is $\ulcorner(F_1)\downarrow(F_2)\urcorner$. The free occurrences of any variable α are those of F_1 and those of F_2. If F_1 and F_2 are sen-

tences, then $\ulcorner(F_1)\downarrow(F_2)\urcorner$ is a sentence and is true if and only if neither F_1 nor F_2 is true.[10]

Note: The notation 'F(N)' of (8) or (9) will also be used for an arbitrary term t, not necessarily a numeral — i.e., $\ulcorner F(t)\urcorner \underset{\text{df}}{=}$ the result of substituting freely the term t for the free variable of F.

Gödel Numbering. For any expression E, let $\sigma(E)$ be the string of Arabic numerals obtained by replacing S_1 by the Arabic numeral '1', S_2 by '2', ..., S_9 by '9'. This string $\sigma(E)$ designates (in decimal notation) a number, which we will call $g_0(E)$. We shall take for our Gödel number of E (written '$\overset{\circ}{g}(E)$' or '$\overset{\circ}{E}$') the number $g_0(E)+1$.

It will facilitate our exposition if we identify the numbers with the numerals (strings of '1's) which designate them in $\boldsymbol{S_A}$. Then we define the norm of E to be E followed by its own Gödel number.

Arithmetization of the norm function. The following extremely simple definition accomplishes all the arithmetization of syntax which we need:

Def. 1. $n(x) \underset{\text{df}}{=} x \cdot 10^x$.

Explanation. If x is the g.n. (Gödel number) of E, then $n(x)$ is the g.n. of the norm of E. Thus, for example, 37 is the g.n. of S_3S_6. The norm of S_3S_6 is $S_3S_6\underbrace{S_9S_9\ldots\ldots S_9}_{37}$, and its g.n. is

$$\underbrace{3699\ldots.9}_{37}+1 = 3700\ldots0 = \underbrace{37}_{37}\times 10^{37}.$$[11]

Semantical Normality. We say that an expression E satisfies the predicate H (relative to the Gödel correspondence g) if $H\overset{\circ}{E}$ is true. The set W of all expressions satisfying H is precisely the set defined by H, in the sense of Section 3. We also say that E satisfies the *formula* F (when F contains one free variable) if $F(\overset{\circ}{E})$ is true, and we shall also refer to the set W of all E which satisfy F as the set defined by the formula F. Now, the crucial role played by the class abstractors of $\boldsymbol{S_A}$ is that definability by a predicate, and definability by a formula, are thereby equivalent. This is an immediate consequence of Rule 9 of $\boldsymbol{S_A}$ since, if H is the abstraction of F, then E satisfies $H \Longleftrightarrow H\overset{\circ}{E}$ is true $\Longleftrightarrow F(\overset{\circ}{E})$ is true [by Rule 9] $\Longleftrightarrow$ E satisfies F. Thus the sets respectively defined by H and F are the same.

A formula F_N will be called a *normalizer* of formula F if F_N is satisfied by just those expressions E whose norm satisfies F. In the light of the preceding paragraph, the statement that $\boldsymbol{S_A}$ is semantically normal is

[10] We could have used the single primitive '$\subset$' [class inclusion] in place of the joint use of '$\downarrow$' and '$=$' (as occurring between abstracts). We would then have a system formulated in a logic based on inclusion and abstraction (in the sense of Quine). All results of this paper would still go through.

[11] Had we used g_0, rather than g for our Gödel correspondence, then, if x were the g.n. of E, the g.n. of the norm of E would have been $(x+1)10^x - 1$, rather than $x.10^x$.

equivalent to the statement that every formula F (with one free variable) has a normalizer F_N (since F_N defines $\eta(W)$, when F defines W).

THEOREM 4. S_A is semantically normal, relative to g.

PROOF. We must show that every F has a normalizer F_N. Well, take F_N to be the result of replacing the free variable α of F by $\alpha \cdot 10^{\alpha}$ [or rather, by the unabbreviated form $\ulcorner(\alpha) \cdot ((1111111111\neg(\alpha))\urcorner$.]

Then, for any number x, $F_N(x)$ and $F(n(x))$ have the same truth-values. Thus, for any expression E,

$$\text{E satisfies } F_N \iff F_N(\mathring{E}) \text{ is true}$$

$$\iff F(n(\mathring{E})) \text{ is true}$$

$\iff$ the norm of E satisfies F (since $n(\mathring{E})$ is the g.n. of the norm of E!). Hence F_N is satisfied by those E whose norm satisfies F and is thus a normalizer of F.

COROLLARY 4.2. (1) For every definable set of S_A, there is a Tarski sentence. (2) The complement of the truth set T of S_A is not definable in S_A (relative to g). (3) T itself is not definable in S_A (relative to g). (4) Any proposed axiomatization of S_A such that the set of its theorems is definable in S_A (relative to g) is semantically incomplete or inconsistent.

(1) and (2) immediately follow from the preceding theorem, together with the results of Section 3. In particular, in (1), to construct a Tarski sentence for a set W defined by formula F, we first construct the normalizer F_N of F by the method of the preceding theorem, then take the abstraction H of F_N, and then follow H by its own Gödel number. Thus the Tarski sentence for W is the norm of the abstraction of the normalizer of the formula which defines W. (3) and (4) follow, since S_A contains negation (definable from '$\downarrow$').

REMARK. (4) of Corollary 4.2 can be thought of as one form of Gödel's theorem. Definability in S_A is actually equivalent to definability in protosyntax (in the sense of Quine). Thus any formal system for S_A whose set of theorems is protosyntactically definable will be semantically incomplete or inconsistent. This is essentially similar to Quine's result that protosyntax itself is not protosyntactically completable.

4.3. We have just shown a method for constructing normalizers which works for the particular Gödel correspondence g, which we employed. Actually, it will work for any Gödel correspondence relative to which the norm function (i.e., the function which assigns to each expression its norm) is *strictly* definable, in the following sense:

A function f (from expressions to expressions) will be said to be *strictly* defined by the term t (with one variable α), if, for any two expressions E_1 and E_2, $E_1 = f(E_2)$ if and only if the numeral $\mathring{E}_1$ and the term $t(\mathring{E}_2)$ [viz., the result of substituting $\mathring{E}_2$ for α in t] designate the same number.

This notion of strict definability is quite different from the usual much weaker notion of definability of f, viz., the existence of a formula M with two free variables such that, for any E_1 and E_2, $E_1 = f(E_2)$ if and only if $M(\mathring{E}_1, \mathring{E}_2)$ is true. We can, in an obvious manner, extend both notions of definability to functions of more than one argument.

If now the norm function is *strictly* definable relative to g, then to construct a normalizer for F we simply replace all free occurrences of the free variable α of F by the term $t(\alpha)$ which defines the norm function. Since this process nowhere makes use of quantifiers (or other identity of class abstracts),[12] or logical connectives, or more than one variable, then if we completely stripped $\boldsymbol{S}_A$ of its logical connectives, quantifiers, and all variables but one, the resulting vastly weaker system $\boldsymbol{S}_a$ would still be normal and, moreover, so would any extension of $\boldsymbol{S}_a$. Let us state this more precisely:

By the system $\boldsymbol{S}_a$, we mean the system whose signs are those of $\boldsymbol{S}_A$, except for '$\downarrow$' and ''', and whose rules are those of $\boldsymbol{S}_A$, with the omission of Rules (7) and (10), and with Rule (2) changed to (2'), 'x' is a variable. By an extension of $\boldsymbol{S}_a$, we mean a system constructed from $\boldsymbol{S}_a$ by possibly adding additional signs and rules. Then the following theorem is a considerable strengthening of Theorem 4.1:

THEOREM 4.4. Any extension $\boldsymbol{S}'_a$ of $\boldsymbol{S}_a$ is normal relative to any Gödel correspondence g, relative to which the norm function is strictly definable providing that whenever two terms t_1 and t_2 have the same designata, $F(t_1)$ and $F(t_2)$ have the same truth-values.

4.5. NORMALITY OF $\boldsymbol{S}_A$ RELATIVE TO OTHER GÖDEL CORRESPONDENCES. If the norm function is definable (relative to g) in only the weaker sense, rather than strictly definable, then, although the above method of constructing normalizers is no longer available to us, we still have another method which will work for $\boldsymbol{S}_A$, but *not* for $\boldsymbol{S}_a$ (or any arbitrary extension thereof), since the construction depends on quantification.

Letting $N(\alpha, \beta)$ be the formula which defines the norm function, we let $F_N \underset{\text{df}}{=} \ulcorner(\exists\beta)(N(\beta, \alpha) \,\&\, F(\beta))\urcorner$, where the existential quantifier is defined from the universal quantifier in the usual manner, the latter defined as previously indicated, and '&' is defined from '$\downarrow$' in the usual manner. Then F_N is a normalizer of F. Hence,

THEOREM 4.6. If the norm function is definable in $\boldsymbol{S}_A$ relative to a Gödel correspondence g, then $\boldsymbol{S}_A$ is normal relative to g.

We lastly observe that if there is a formula $C(\alpha, \beta, \gamma)$ such that, for any expressions E_1, E_2, E_3, $E_3 = E_1E_2$ if and only if $C(\mathring{E}_1, \mathring{E}_2, \mathring{E}_3)$ is true (which we express by saying that concatenation is definable, relative to g) and if

[12] As indicated at the beginning of Section 4, quantification is defined, using a formula which employs the identity sign between class abstracts.

there is a formula $G(\alpha, \beta)$ such that, for any expressions E_1 and E_2, E_1 is the Gödel numeral of E_2 if and only if $G(\mathring{E}_1, \mathring{E}_2)$ is true (which we express by saying that g itself is definable relative to g), then the formula $\ulcorner \exists(\gamma)(G(\gamma, \beta).\&.C(\beta, \gamma, \alpha))\urcorner$ defines the norm function and $\boldsymbol{S}_A$ is normal. Hence

THEOREM 4.7. A sufficient condition for $\boldsymbol{S}_A$ to be normal, relative to g, is that concatenation and g itself both be definable, relative to g.

REMARK. Gödel correspondences satisfying the hypothesis of Theorem 4.7 include all those that are *effective* (i.e., include all those g such that the function h, which assigns to each number x the Gödel number of (the numeral designating) x, is a recursive function.[13] This, in conjunction with previous results, yields the proposition that, relative to any effective Gödel correspondence g, the truth set of $\boldsymbol{S}_A$ is not definable. This, in essence, is Tarski's Theorem.

5. Concluding Remarks: Diagonalization vs. Normalization. We should like, in conclusion, to compare the norm function, used throughout this paper, with the more familiar diagonal function, used for systems in standard formalization.

Firstly, to sketch a general account of diagonalization,[5] analogous to Section 3 for normalization, we consider now an arbitrary language $\boldsymbol{L}$ which (like $\boldsymbol{S}$ of Section 2) contains expressions, sentences, true sentences, and individual constants. Instead of predicates, however, we now have certain expressions called 'formulas' and others called 'variables,' and certain occurrences of variables in formulas termed 'free occurrences,' subject to the condition that the substitution of individual constants for all free occurrences of variables in a formula always yields a sentence. We again have a Gödel correspondence mapping each expression E onto an individual constant $\mathring{E}$. For any formula F with one free variable α and any expression E, we define F(E) as the result of substituting $\mathring{E}$ for all free occurrences of α in F. The expression F(F) is defined to be the *diagonalization* of F. The set of all E such that F(E) is true, is called the set *defined* by F. For any set W, we define $D(W)$ as the set of all F whose diagonalization is in W. Then the analogue of Theorem 3.1 is 'A sufficient condition for the existence of a Tarski sentence for W is that $D(W)$ be definable.'

[13] E.g., the correspondence $\boldsymbol{g}_0$. To show that, relative to $\boldsymbol{g}_0$, the norm function is weakly definable, we must construct a formula $\ulcorner \varphi(\alpha, \beta)\urcorner$ such that, for any two numbers n and m, $\varphi(n, m)$ is true $\Leftrightarrow m+1=(n+1).10^n$ (cf. (9)). We first define addition as follows: Add $(\alpha, \beta, \gamma) \underset{\text{df}}{=} n^\alpha.n^\beta=n^\gamma$ (where we take n any number $\neq 1$). Then we define $\varphi(\alpha, \beta) \underset{\text{df}}{=} (\exists\gamma)$ [Add $(\alpha, 1, \gamma)$ & Add $(\beta, 1, \gamma.10^\alpha)$] (this construction can be simplified by introducing descriptors). Thus the tricky correspondence $\boldsymbol{g}_0+1$, which we used, was introduced only for purposes of simplicity, and is certainly not necessary for the success of our program.

Hence also, $D(\overline{W})$ is not definable. We would then define normality, for such a language $\boldsymbol{L}$, by the condition that whenever W is definable in $\boldsymbol{L}$, so is $D(W)$. Then all other theorems in Section 3 have their obvious analogues.[14]

To apply these general notions to systems in standard formalization, e.g., elementary arithmetic, we would have, in analogy with the notion 'normalizer,' that of 'diagonalizer,' where a diagonalizer F_D of a formula F would be a formula satisfied by just those expressions whose diagonalization satisfied F. Then, if W is defined by F, and if there exists a diagonalizer F_D for F, then the diagonalization of F_D (which is $\mathrm{F}_D(\mathrm{F}_D)$), is the Tarski sentence for W.

This is essentially the classical construction. The construction of the diagonalizer F_D is considerably more involved than the construction of the normalizer F_N. Again, we might say, this is due to the fact that concatenation is easier to arithmetize than substitution.

DARTMOUTH COLLEGE

[14] We can profitably avoid repetition of analogous arguments for $\boldsymbol{S}$ and $\boldsymbol{L}$ by regarding both as special cases of a more general structure. This approach will be presented in a forthcoming paper, *Abstract structure of unsaturated theories*, in which we study, in considerable generality, the deeper properties of undecidable systems, uncovered by Gödel and Rosser.