ON A THEOREM OF COBHAM CONCERNING UNDECIDABLE THEORIES

ROBERT L. VAUGHT

University of California, Berkeley, California, U.S.A.

1. Introduction

A general method for establishing the undecidability of theories was developed in [13]. Its basic tools are the following two facts (unexplained terminology throughout is that of [13] [1]):

1.1. *There is a finitely axiomatizable and essentially undecidable theory Q which is a fragment of the arithmetic of natural numbers.*

1.2. *If a theory T is compatible with a finitely axiomatizable and essentially undecidable theory Σ, then T is undecidable.*

A still weaker fragment of number theory, the theory R, was also introduced in [13]. R is not finitely axiomatizable. Moreover,

1.3. *R is essentially undecidable,*

and

1.4. *Any member of R has a finite model.*

The simple argument which establishes 1.2 depends on the finite axiomatizability of Σ. It has, in fact, been proved (cf. [2], [8]) that 1.2 does not always hold if Σ is only assumed to be axiomatizable (i.e., to have a recursive axiom set).

A. Cobham established the important result that, nevertheless, 1.2 does hold for R and, indeed, for a somewhat weaker theory R_0 (to be described in § 3):

1.5. COBHAM'S THEOREM.[2] *If a theory T is compatible with R_0, then T is undecidable.*

Cobham has employed 1.5 in proving that the theory G of finite groups is hereditarily undecidable (and, hence, not axiomatizable).[3] Since the hypoth-

---

This work was supported by the National Science Foundation under grant G–8934.

[1]However, we shall identify a theory with its set of valid sentences. Moreover, a theory T_1 will only be said to be an extension of, or compatible with, a theory T_2 when T_1 and T_2 have the same constants.

[2]This result, as yet unpublished, was communicated orally by Cobham to the author and others in 1957. The full proof and various consequences were described by Cobham in correspondence with A. Tarski and the author in 1958.

[3]This solution of a problem posed in [13, p. 85], also not yet published, was stated and proved in a letter to Tarski in 1958. As Cobham has noted, for such applications it is better to replace 1.5 by the following statement, easily inferred from it: If R_0 is relatively weakly interpretable in a theory T, then T is hereditarily undecidable. (Cf. the discussion in [13, pp. 20–30] of analogous modifications of 1.2.)

esis of 1.2 obviously fails for G, this result could not be obtained by applying 1.2. Thus, it affords an excellent illustration of the additional power possessed by 1.5 as compared with 1.1 and 1.2.

We shall present here a new proof of Cobham's Theorem. Indeed, we shall show that, by a method which may be called existential interpretation, Cobham's Theorem can be derived from a theorem of Trahténbrot. Before stating the latter in 1.6, below, we need the following terminology:

L is the set of logically valid sentences. F is the set of finitely satisfiable sentences. For any set K, NgK is the set of sentences whose negations belong to K. A set Z is said to *separate* a set X from a set Y if Z includes X and is disjoint from Y. Two sets are *recursively inseparable* if no recursive set separates them and they are disjoint.

1.6. TRAHTÉNBROT'S THEOREM.[4] *L and NgF are recursively inseparable.*

In [18], the author showed that:

1.7. *If A is any axiomatizable theory separating L from NgF, then A is compatible with some finitely axiomatizable and essentially undecidable theory.*

In other words, any *axiomatizable* theory whose undecidability follows from Trahténbrot's Theorem also fulfills the hypothesis of Tarski's condition 1.2. We will show (in 5.3) that a similar conclusion applies to axiomatizable theories whose undecidability follows from Cobham's Theorem.

We shall also discuss several other ways in which Cobham's Theorem can be improved or generalized. In the final section, we mention some related problems.

2. Existential Interpretability

We deviate from [13] by considering theories in which the identity symbol may not be present; all notions and results of [13] extend to such theories in an obvious way. On the other hand, in order to avoid bothersome details, we assume that the non-logical constants of any theory are relation symbols, finite in number. (Hence, in further references to the theory R, we shall mean the equivalent theory in our sense.)

A number of principles are known which have, roughly speaking, this form: If a theory T_2 can be 'interpreted' in a theory T_1 in this or that sense, then a certain property (e.g., undecidability, recursive inseparability of T_2 and NgT_2, etc.) is preserved in passing from T_2 to T_1. (Cf. [13, 4, and 12].) In 2.1 below, we shall state some principles of the same sort concerning various properties related to Cobham's Theorem and concerning a new notion called existential interpretability. The latter turns out to be useful not only when applied to theories, but also when applied to other sets of sentences.

[4]Cf. [15]. 1.6 was an improvement of Trahténbrot's earlier result [14] that the set of finitely valid sentences is not recursive.

If K is any set of sentences, we denote by ExK the set of all expressions involving sentential connectives, quantifiers, variables, and only those relation symbols which occur in members of K. We denote by $L(K)$ or $F(K)$ the set of sentences $\sigma \in ExK$ such that σ is logically valid or finitely satisfiable, respectively. K will be called a *pseudo-theory* if K is a non-empty set of consistent sentences, only finitely many relation symbols occur in members of K, and any sentence $\sigma \in ExK$ logically equivalent to a member of K belongs to K. [5]

Suppose now that K_1 and K_2 are pseudo-theories. To simplify the notation, we confine ourselves, in the following definition of existential interpretation and in the proof of 2.1, below, to the case when only one relation symbol, P, having two places, occurs in members of K_2. (From this case the situation for an arbitrary K_2 will be obvious.) Suppose that natural numbers p, q, r and formulas $\epsilon(v_0, \cdots, v_{r+p})$ and $\rho(v_0, \cdots, v_{2r+1+q})$, belonging to ExK_1, are specified, the formulas having exactly the free variables indicated. Let u_n^m, y_j, z_k $(m, n = 0, 1, \cdots; j < p; k < q)$ be distinct variables. For each sentence $\sigma \in ExK_2$, let σ^* be obtained from σ by replacing each atomic formula $Pv_m v_n$ by

$$\rho(u_o^m, \cdots, u_r^m, u_o^n, \cdots, u_r^n, z_0, \cdots, z_{q-1})$$

and then replacing all subformulas $\bigwedge v_k \gamma$ or $\bigvee v_k \gamma$ by $\bigwedge u_o^k \cdots \bigwedge u_r^k [\epsilon(u_o^k, \cdots, u_r^k, y_0, \cdots, y_{p-1}) \rightarrow \gamma]$ or $\bigvee u_o^k \cdots \bigvee u_r^k [\epsilon(u_o^k, \cdots, u_r^k, y_0, \cdots, y_{p-1}) \wedge \gamma]$, respectively $(k, m, n = 0, 1, \cdots)$. [6] We say that an *existential interpretation* of K_2 in K_1 has been specified if for each σ belonging to K_2, the sentence $\bigvee y_0 \cdots \bigvee y_{p-1} \bigvee z_0 \cdots \bigvee z_{q-1} \sigma^*$ belongs to K_1. [7]

Henceforth, we shall say that an (ordinary) *interpretation* of K_2 in K_1 is specified when, in the above, $p = q = 0$. It should be noted that we are thus understanding the notion 'interpretable' somewhat more broadly than even the notion 'relatively interpretable' of [13]. Roughly speaking, the latter has been extended by allowing the identity symbol to be interpreted like any other relation symbol, and by allowing the elements of models of K_2 to be interpreted by r-tuples of elements of models of K_1. [8] This fact plays a serious role only in 6.4 and 8.3; most of the time, what matters is only the difference between existential and ordinary interpretability.

2.1. *Suppose that K_1 and K_2 are pseudo-theories and that K_2 is existentially interpretable in K_1. Then*:

---

[5] With some slight changes, the last condition could have been omitted in most of what follows.

[6] If ψ is a formula and $w_0, \ldots, w_{n-1}$ are variables, then $\psi(w_0, \ldots, w_{n-1})$ is the formula obtained by simultaneous substitution of $w_0, \ldots, w_{n-1}$ for the free occurrences of $v_0, \ldots, v_{n-1}$ in ψ, bound variables being changed (in some fixed, effective way) when necessary to avoid collisions.

[7] If the identity symbol occurs in K_2, it is to be 'interpreted' in the same way as any other relation symbol.

[8] Cf. [10], and [13, footnote 17, p. 22].

(a) *If no recursive theory separates $L(K_2)$ from NgK_2, then the same holds for K_1.*[9]

(b) *If $L(K_2)$ and NgK_2 are recursively inseparable, then so are $L(K_1)$ and NgK_1.*

(c) *If $L(K_2)$ and NgK_2 are effectively recursively inseparable,*[10] *then so are $L(K_1)$ and NgK_1.*

(d) *If every axiomatizable theory $L(K_2)$ from NgK_2 is compatible with some finitely axiomatizable, essentially undecidable theory, the the same holds for K_1.*

PROOF. For each sentence $\sigma \varepsilon Ex K_3$, let $f(\sigma)$ be the sentence

$$\bigwedge y_0 \cdots \bigwedge y_{p-1} \bigwedge z_0 \cdots \bigwedge z_{q-1}\{\bigvee u_0^0 \cdots \bigvee u_r^0 \in (u_0^0, \cdots u_r^0, y_0, \cdots, y_{p-1}) \rightarrow \sigma^*\}.$$

Then, obviously:

$$(1) \qquad \begin{cases} \textit{if } \sigma \in L(K_2) \textit{ then } f(\sigma) \in L(K_1); \\ \textit{if } \sigma \in NgK_2 \textit{ then } f(\sigma) \in NgK_1. \end{cases}$$

(Indeed, the first holds by the substitution theorem of predicate logic, and the second is just what the existential interpretation has insured.) Note also that, for any sentences $\sigma_1, \sigma_2 \in ExK_2$,

$$(2) \qquad f(\sigma_1 \rightarrow \sigma_2) \rightarrow [f(\sigma_1) \rightarrow f(\sigma_2)]$$

is logically valid.

Since the function f is recursive, we have by (1) what may be called a *reduction* of the pair $L(K_2)$, NgK_2 to the pair $L(K_1)$, NgK_1. It is well known and easily proved that such a reduction allows us to infer the implications (b) and (c). The implication (a) can be inferred similarly in view of the additional condition (2). To make the ideas clear, we shall outline the details of the proof or (d).

Assume the hypothesis of (d), and suppose A_1 is an axiomatizable theory separating $L(K_1)$ from NgK_1. Let $A_2 = \{\sigma / f(\sigma) \in A_1\}$. By (1), A_2 separates $L(K_2)$ from NgK_2. By (2) and the recursiveness of f, A_2 is, moreover, an axiomatizable theory. Hence there is a sentence δ consistent with A_2 such that $Th(\delta)$[11] is essentially undecidable. Hence $\sim \delta \notin A_2$, so $f(\sim \delta) \notin A_1$ and $\sim f(\sim \delta)$ is consistent with A_1. $Th(\sim f(\sim \delta))$ is obviously essentially undecidable (indeed, this is a special case of (a).[12]

---

[9]If K_2 is a theory, the hypothesis of (a) obviously is equivalent to the statement that any theory compatible with K_2 is undecidable.

[10]For the definition of this notion, see [12] or [17].

[11]$Th(\delta)$ is the theory whose only axiom is δ.

[12]Just this special case of (a) is stated in [13, Part I, Theorem 8]. It may be mentioned that the notion in [13] of inessential extension obviously has the following relation to existential interpretability: T_2 is weakly existentially interpretable in T_1 if and only if T_2 is weakly interpretable in an inessential extension of T_1.

The relation of existential interpretability among pseudo-theories is easily seen to be transitive.

It can be shown by means of an example that an essentially undecidable theory is sometimes existentially interpretable in a consistent, decidable theory.

3. The Theory R_0

The relation symbols of R_0 are Δ_0, Sc, Sm, Pr, and $\leqq$, of 1, 2, 3, 3, and 2 places, respectively. Let the formulas $\Delta_{n+1}(x)$ be defined inductively by requiring that Δ_{n+1} is $\bigvee y[\Delta_n(y) \wedge Sc(y, x)]$. [13] The axioms of R_0 are:

(I) (a) $\bigvee x \Delta_m(x)$ (b) $\Delta_m(x) \rightarrow \sim \Delta_n(x)$ (*for* $m \neq n$)

(II) $\Delta_m(x) \wedge \Delta_n(y) \rightarrow [Sm(x, y, z) \rightarrow \Delta_{m+n}(z)]$

(III) $\Delta_m(x) \wedge \Delta_n(y) \rightarrow [Pr(x, y, z) \rightarrow \Delta_{m \cdot n}(z)]$

(IV) $\Delta_n(y) \rightarrow [x \leqq y \leftrightarrow \Delta_0(x) \vee \cdots \vee \Delta_n(x)]$.

Unlike R_0, R has the identity symbol, and has constant terms for the Δ_n. Moreover, R has an axiom schema $(V)(x \leqq \Delta_n \vee \Delta_n \leqq x)$ having no analogue in R_0. Schema (V) can, in fact, be removed in R itself without changing any of the essential properties of R.

R_0 shares with R the properties 1.3 and 1.4 and the property:

3.1. *Every recursive function or relation is definable in* R_0.

In 3.1 we are now to regard an $(n+1)$-ary numerical relation W (or function f) as *definable in* R_0 if, for some formula $\phi(v_0, \cdots, v_n)$ (resp., $\psi(v_0, \cdots, v_{n+1})$, with free variables as indicated,

$\Delta_{k_0}(v_0) \wedge \cdots \wedge \Delta_{k_n}(v_n) \rightarrow \phi$ is valid in R_0, whenever $Wk_0 \cdots k_n$, and

$\Delta_{k_0}(v_0) \wedge \cdots \wedge \Delta_{k_n}(v_n) \rightarrow \phi$ is valid in R_0, otherwise.

(resp., $\Delta_{k_0}(v_0) \wedge \cdots \wedge \Delta_{k_n}(v_n) \rightarrow [\psi \leftrightarrow \Delta_{f(k_0, \ldots k_n)}, (v_{n+1})]$ is valid in R_0).

That 1.4 holds for R_0 is obvious. Theorem 3.1 may be proved in essentially the usual way (cf. [13, Part II, Theorem 6] or [7, §49 and §59]) except that a device is needed where axiom schema (V) was formerly used. Roughly speaking, this device consists in relativizing some quantifiers to the formula $N(x)$:

$$\bigvee y[\Delta_0(y) \wedge y \leqq x] \wedge \bigwedge y \bigwedge z[y \leqq x \wedge S(y, z) \rightarrow x \leqq z \vee z \leqq x].$$

With these indications we shall leave the details of the proof of 3.1 to the reader. Finally, 1.3 for R_0 follows from 3.1 by an obvious modification of the old argument (cf. [13, Part II, Theorem 1]).

The possibility of modifying R and the definition of 'definable' in these ways and still establishing 3.1, and inferring from it 1.3, was observed independently by Cobham and (later) the author, about three years ago.

---

[13] For readability, we agree henceforth that x, y, z are v_0, v_1, v_2, respectively, and that x, y, z, u, w, $x_0, \ldots, x'$, etc., are distinct variables.

It may be mentioned that Cobham's proof of 1.5 involves an ingenious modification, of a much more fundamental nature than those mentioned above, in the statements and proofs of Theorems 1 and 6 of [13, Part II].

4. Trahténbrot's Theorem

The sets L and F (and hence the meanings of 1.6 and 1.7) are not determined until we specify just what relation symbols are to occur in the sentences of L and F. Trahténbrot [15] showed that for a certain, finite list of relation symbols, 1.6 is true; we shall use the notations L' and F' for this case. The notations L and F will be used for the case when only one binary relation symbol P is allowed. In [18], it was remarked that (by considering one or another of the known ways for reducing the decision problem for L' to that for L) one can infer 1.6 for L and F from 1.6 for L' and F', and the same applies to 1.7. In fact, in the light of § 2, we can say that what was remarked in [18] was, in essence, the fact that

4.1. *F' is existentially interpretable in F,*

plus the appropriate instances of 2.1, (b) and (d).[14] (Note that F and F' are obviously pseudo-theories.)

Trahténbrot's proof of 1.6 (for L' and F') showed that in fact L' and NgF' are effectively recursively inseparable. (One way to see this is to note that Trahténbrot showed there is a reduction of Kleene's pair of recursively enumerable, recursively inseparable sets to the pair L', NgF'. Since the former two sets are effectively recursively inseparable (cf., e.g., [7, § 61]), it follows that so are L' and NgF'.) Hence, by 4.1 and 2.1(c):

4.2. *L and NgF are effectively recursively inseparable.*

5. Cobham's Theorem

5.1. *F is existentially interpretable in R_0.*

PROOF. It is well known that there is a recursive ternary relation W such that for any k and any binary relation Q among the elements of $\{0, \cdots, k\}$, there is a natural number j such that for any m, n, $Wmnj$ if and only if Qmn.

Now we specify p, q, r, $\in$, and ρ so as to obtain the desired existential interpretation (cf. § 2). By 3.1, we may take for $\rho(v_0, v_1, v_2)$ a formula such that

$$(3) \qquad \rho(v_0, v_1, v_3) \textit{ defines } W \textit{ in } R_0.$$

Put $p = q = r = 1$, and for $\varepsilon(v_0, v_1)$ take the formula $v_0 \leqq v_1$.

Suppose that $\sigma \in F$ and that σ^* is obtained from σ as in § 2. Then, from (3) and axiom schema IV, one easily infers that, as desired, the formula

[14] Cf. § 4 and [18, footnotes 4 and 20]. (In the italicized statement in footnote 20, 'σ' and '$\sigma^\wedge$' should be interchanged.)

(4) $$\bigvee y_0 \bigvee z_0 \sigma^*$$

is valid in R_0.

Theorems 2.1 (c), 4.2, and 5.1, together, provide a new proof of Cobham's Theorem 1.5 and, indeed, of the stronger statement:

5.2. *$L(R_0)$ and NgR_0 are effectively recursively inseparable.*[15]

Similarly, from 2.1(d), 1.7, and 5.1, we infer at once a second result (also implying 1.5).

5.3. *If A is any axiomatizable theory compatible with R_0, then A is compatible with some finitely axiomatizable and essentially undecidable theory.*

Of course, in practice, one must use a fixed theory, such as Q, for Σ in applying 1.2. Thus 5.3 means only that, in theory, 1.5 yields no axiomatizable and undecidable theories not already obtainable from 1.2. Moreover, even in theory, 5.3 does not apply to non-axiomatizable theories, such as G.

6. Set Theories

One can form various set theories, concatenation theories, etc., which resemble the theories R and R_0.

One such theory, a very weak Zermelo-type set theory, is the following theory, Z_0. Z_0 has the identity symbol and the binary relation symbol ε. For any set X, let $P(X)$ be the power set of X; and let $M = 0 \cup P(0) \cup PP(0) \cup \cdots$. For $r = \{s_0, \cdots, s_{n-1}\} \in M$, the formula $D_r(x)$ is defined (inductively) to be:

$$\bigvee y_0 \cdots \bigvee y_{n-1}\Big[\prod_{i<n}(D_{s_i}(y_i) \wedge y_i \in x) \wedge \bigwedge z(z \in x \rightarrow \sum_{i<n} z = y_i)\Big].$$

The axioms of Z_0 are all sentences

$$\prod_{i<n} D_{s_i}(x_i) \rightarrow \bigvee y \bigwedge z(z \in y \leftrightarrow \sum_{i<n} z = x_i) \quad (\textit{for } s_0, \cdots, s_{n-1} \in M).$$

Take $\rho(x, y, z)$ to be the (Kuratowski ordered-pair) formula:

$$\bigvee w \bigvee w' \bigwedge u[(u \in z \leftrightarrow u = w \vee u = w') \wedge (u \in w \leftrightarrow u = x) \wedge (u \in w' \leftrightarrow u = x \vee u = y)].$$

One easily verifies that (4) is valid in Z_0 whenever $\sigma \in F$. Hence,

6.1. *F is existentially interpretable in Z_0.*

---

[15] It was noted in [6, Theorem 2.5B] that condition 3.1 (for R) implies the statement (A): R and NgR are recursively inseparable. In [4], Feferman showed that (B): every axiomatizable consistent extension of R is creative. Smullyan [12] established a result implying both (A) and (B), namely, (C): R and NgR are effectively recursively inseparable. Theorem 5.2, in turn, implies both (C) and 1.5. Cobham's proof of 1.5 may easily be modified by imitating the changes in the proof of 1.3 needed to prove (A), (B), and (C) so as to yield a proof of 5.2.

A theory Z_{00} still weaker than Z_0 might be formed having only the axioms $\bigvee x D_r(x) (r \in M)$. I have been unable to determine whether or not Z_{00} is essentially undecidable.

Next we consider a second-order set theory, T_0. T_0 was described in [18] and differs only slightly from a theory involved in [15]; it will be discussed further in § 7. T_0 has the relation symbols I, ε, and ρ of 1, 2, and 3 places, respectively. (Ix is read 'x is an individual'.) Its axioms are the sentence α:

$$\bigwedge x \bigwedge y \bigwedge u[x \in u \vee \rho(x, y, u) \vee \rho(y, x, u) \rightarrow Ix] \\ \wedge \bigvee u \bigwedge x x \notin u \wedge \bigvee u \bigwedge x \bigwedge y \sim \rho(x, y, u) \\ \wedge \bigwedge u \bigwedge y \bigwedge y' \{Iy \wedge Iy' \rightarrow \bigvee u' \bigvee u'' \bigwedge x \bigwedge x' [[x \in u' \leftrightarrow x \in u \vee x = y] \\ \wedge [\rho(x, x', u'') \leftrightarrow \rho(x, x', u) \vee (x = y \wedge x' = y')]]\},$$

and all the sentences

$$\bigvee x_0 \cdots \bigvee x_{n-1} \Big(\prod_{i>n} Ix_i \wedge \prod_{i>j<n} x_i \neq x_j \Big).$$

Again the sentence (4) is obviously valid in T_0 whenever $\sigma \in F$. Hence,

6.2. *F is existentially interpretable in T_0.*

(T_0 can be considerably weakened without losing this property.)

Finally, we consider a theory, S, which is closely related to the pseudo-theory F. S has no identity symbol, and has the relation symbols ε and ρ. Its axioms are simply all sentences (4) such that $\sigma \in F$. Obviously,

6.3. *F is existentially interpretable in S.*

Indeed, as one readily shows, S has the peculiar property:

6.4. *For any pseudo-theory K, F is existentially interpretable in K if and only if S is interpretable in K.*

By 6.1, 6.2, 6.3, 2.1, 4.1, and 4.2, we see that 5.2 and 5.3 hold for each of Z_0, T_0, and S_0, in place of R_0.

From the essential undecidability of Z_0 it follows that the stronger theory Z_1, whose only axioms are $\bigvee u \bigwedge z (z \notin u)$ and $\bigwedge x \bigwedge y \bigvee u \bigwedge z (z \in u \leftrightarrow z \in x \vee z = y)$, is also essentially undecidable.[16] This answers a question raised by Szmielew and Tarski, who established (cf. [13, p. 34]) the essential undecidability of the theory having, in addition to these two, the axiom of extensionality.

Cobham observed that Trahténbrot's Theorem 1.6, for the sets $L(G)$ and $F(G)$ (or the sets $L(R_0)$ and $F(R_0)$), is a consequence of his results concerning G (or R_0), because $G \subseteq F(G) (R_0 \subseteq F(R_0)$, by 1.4). By noting that 4.1 was shown to be valid without any specific assumptions about what relation symbols occur in members of F', we may, in fact, conclude from 4.1 that

[16]The same result can also be obtained by interpreting in Z_1 the theory T_0. Regarding the essential undecidability of the latter, cf. [18, pp. 12–13], where reference is made to a related result in [16].

6.5. *Any pseudo-theory K, such that $K \subseteq F(K)$, is existentially interpretable in F.*

Now, it is obvious that $Z_0 \subseteq F(Z_0)$, $T_0 \subseteq F(T_0)$, and $S \subseteq F(S)$. As a result, we infer from 6.5, 6.1, 6.2, 6.3, and 5.1 that in a certain sense all of the pseudo-theories F, Z_0, T_0, S, R_0, and R have the same strength. Indeed:

6.6. *Each of F, Z_0, T_0, S, R_0, and R is existentially interpretable in each of the others.*

Among the theories S, Z_0, T_0, R_0, and R one may ask: which have (ordinary) interpretations in which of the others? It is not difficult to see that Z_0 is interpretable in R_0 and that R_0 (and, hence, Z_0) is interpretable in R and also in T_0. Among Z_0, T_0, R_0, and R these may well be the only cases where the answer is positive, but we have no proof of any of these negative assertions. In general, to establish non-interpretability seems to be a rather awkward matter.

We have succeeded in showing that:

6.7. *None of Z_0, T_0, R_0, and R is interpretable in S.*

The (rather tedious) proof makes use of results of [5] concerning the elementary theory of a relational system which is the cardinal sum of other systems (cf. [5, Theorem 3.2, § 4.7, and Theorem 6.1.2]).

7. An Improvement of 5.3

J. C. Shepherdson raised the question whether one could strengthen 5.3 as follows:

7.1. *If A is an axiomatizable theory compatible with R, then A is compatible with some finitely axiomatizable extension of R.*

I have proved 7.1, and the corresponding statement for R_0. The proofs are obtained by making a slight change at the end of Cobham's proof of 1.5. It appears to be essential to employ in the proof of 7.1 (for R or R_0) the ideas involved in Cobham's proof of 1.5. One may also ask whether 7.1 holds for Z_0, T_0, or S in place of R. It turns out that the answer is affirmative for Z_0 and T_0, but negative for S. Most of the properties we have considered have been shown rather easily to pass from any one of R_0, Z_0, T_0, or S to the others. But in the case of 7.1 one apparently must make a detailed argument for each theory in question.

We shall not prove 7.1 here, but we shall establish the analogous result for T_0 (which we call 7.2). Indeed, the theory T_0 was formed (essentially, by Trahténbrot) in just such a way as to make possible short proofs (in [15] and [18]) of 1.6 and 1.7. Hence, the argument for 7.2, which differs only slightly from those proofs, is also short.

PROOF OF 7.2. Let A be an axiomatizable theory compatible with T_0. For each sentence $\sigma \in ExF$, construct the formula $\sigma^* \in Ex(T_0)$ as in § 2 (with $p = q = r = 1$). Then the sentence $\bar{\sigma}$

$$\alpha \wedge \bigvee y_0 \bigvee z_0[\bigvee w(w \in y_0) \wedge \sigma^*]$$

is obviously consistent with A when $\sigma \in F$ (since then $\bar{\sigma} \in T_0$). Now, if $\bar{\sigma}$ were consistent with A *only* when $\sigma \in F$, we could infer that $\sim \bar{\sigma} \in A$ if and only if $\sigma \in F$; and, hence, that F has a recursively enumerable complement. This would contradict Trahténbrot's result of 1950 [14]. Thus A is consistent with some $\bar{\sigma}$ such that σ has no finite model. By looking at the axioms of T_0 one now easily verifies that $Th(\bar{\sigma})$ is an extension of T_0, completing the proof.

Since 7.2 implies that the hypothesis of 2.1(d) holds for T_0, one obtains (noting 6.6) a proof of 1.7.

As regards Z_0, the situation is similar to that for R. To show that 7.1 holds for Z_0, one must first develop certain detailed facts about Z_0.

It can be shown that 7.1 fails for S; that, in fact:

7.3. *There is a consistent, finite extension of S having no consistent, finitely axiomatizable extension which is an extension of S.*

7.3 may be established rather easily by relying on a result of [5, Theorem 6.10—6.7.1] concerning cardinal sums.

Ehrenfeucht and Feferman [3] have shown that in any axiomatizable theory which is a consistent extension of R, every recursively enumerable set is representable.[17] Shepherdson obtained a more direct proof of the same theorem, and then, by employing the results involved in Cobham's proof of 1.5, was able to establish the stronger theorem:

7.4. (SHEPHERDSON) *In any axiomatizable theory A compatible with R, every recursively enumerable set is representable.*

Shepherdson was then motivated to propose 7.1 by observing that by means of 7.1 one could obtain 7.4 at once from the theorem of Ehrenfeucht and Feferman. (Indeed, let A be compatible with $Th(\sigma)$, which extends R; and let $\theta(x)$ represent a set Y in $Th(A \cup \{\sigma\})$. Then $\sigma \rightarrow \theta(x)$ represents Y in A.)

Putnam and Smullyan [9] have improved the Theorem of Ehrenfeucht and Feferman in another direction by showing that: in any axiomatizable theory which is a consistent extension of R, every disjoint pair of recursively enumerable sets can be 'exactly separated' (in a sense described in [9]). By the same argument of Shepherdson, one can again replace 'consistent extension of' by 'compatible with', in view of 7.1.

8. Some Problems

8.1. *Is there a theory which is finitely axiomatizable and undecidable but is not compatible with any finitely axiomatizable, essentially undecidable theory?*[18]

[17]For the meaning here of 'representable', see [6]. In considering representability, we shall discuss R, rather than R_0, in order to avoid considering the various alternatives that appear when one tries to define this notion for R_0.

[18]This problem, which seems to have originated with Tarski, has been discussed in [1, p. 393] and [18, p. 13].

Ehrenfeucht [3] and Putnam [8] showed that 8.1 can be answered affirmatively if the word 'finitely' is deleted. As was remarked by Cobham, a variant of 8.1, whose answer seems not to be known, is obtained by replacing 'finitely axiomatizable and undecidable' by 'axiomatizable and hereditarily undecidable'. (The same applies to 8.2 and 8.3, below.) Cobham's Theorem, 1.5, might seem at first glance to offer a possible method for finding a theory as demanded in 8.1, but 5.3 shows that in fact it does not.

The second problem was stated by Feferman [4]:

8.2. *Is there a theory which is finitely axiomatizable and undecidable but not of the highest degree of undecidability for recursively enumerable sets?*

Feferman [4] obtained an affirmative answer if 'finitely' is deleted; Shoenfield [11] showed that this remains correct if, in addition, 'undecidable' is changed to 'essentially undecidable'. Feferman [4] showed that no theory in which R is interpretable or Q is weakly interpretable can provide an example for Problem 2. By 5.2, the same applies to any theory in which R is weakly existentially interpretable.

The difficulty of Problems 8.1 and 8.2 seems to be related to another, imprecise question: Is there any way of showing that a finitely axiomatizable theory is undecidable except what is in essence the original method of Gödel and Church? In 8.3 we attempt to present a precise version of this question.

8.3. *Is there a theory which is finitely axiomatizable and undecidable but in which S is not weakly interpretable?*

By 6.4, 6.5, and 6.6, one sees that a number of possible alternatives to Problem 3 are all equivalent.

Added in proof: If ExK_2 contains $=$ (and, say, P), then the first of the assertions (1) in the proof of 2.1 need not be correct. However, 2.1 remains valid because the assertion in question *is* valid for *some* existential interpretation. Indeed it is easy to verify that such an existential interpretation is obtained from any given one by changing it only as follows: In forming the new σ^*, $x = y$ should be replaced by the old ϕ^*, where ϕ is the "built-in equality" formula:

$$\wedge z[(z = x \rightarrow z = y) \wedge (x = z \rightarrow y = z) \wedge (Pxz \rightarrow Pyz) \wedge (Pzx \rightarrow Pzy)].$$

(For this argument it is essential that there are only finitely many relation symbols in ExK_2.)

References

[1] Cobham, A. Effectively decidable theories. *Summaries of talks presented at the Summer Institute for Symbolic Logic, Cornell University*, 1957, second edition, Institute for Defense Analyses, 1960, pp. 391–395.

[2] Ehrenfeucht, A. Two theories with axioms built by means of pleonasms. *Journal of Symbolic Logic*, Vol. 22 (1957), pp. 36–38.

[3] EHRENFEUCHT, A., and S. FEFERMAN. Representability of recursively enumerable sets in formal theories. To appear in *Archiv für Math. Logik u. Grundlagenforschung*.

[4] FEFERMAN, S. Degrees of unsolvability associated with classes of formalized theories. *Journal of Symbolic Logic*, Vol. 22 (1957), pp. 161–175.

[5] FEFERMAN, S., and R. VAUGHT. The first order properties of products of algebraic systems. *Fundamenta Mathematicae*, Vol. 47 (1959), pp. 57–103.

[6] GRZYGORCZYK, A., A. MOSTOWSKI, and C. RYLL-NARDZEWSKI. The classical and the ω-complete arithmetic. *Journal of Symbolic Logic*, Vol. 23 (1958), pp. 188–206.

[7] KLEENE, S. C. *Introduction to Metamathematics*. Amsterdam, North Holland, 1952, x + 550 pp.

[8] PUTNAM, H. Decidability and essential undecidability. *Journal of Symbolic Logic*, Vol. 22 (1957), pp. 39–49.

[9] PUTNAM, H., and R. M. SMULLYAN. Exact separation of recursively enumerable sets within theories. To appear in *Proceedings of the American Mathematical Society*. (Abstract presented at this Congress.)

[10] ROBINSON, A., and A. H. LIGHTSTONE. Syntactical transforms. *Transactions of the American Mathematical Society*, Vol. 86 (1957), pp. 220–245.

[11] SHOENFIELD, J. Degrees of formal systems. *Journal of Symbolic Logic*, Vol. 23 (1958), pp. 389–392.

[12] SMULLYAN, R. M. Theories with effectively inseparable nuclei. *Journal of Symbolic Logic*, Vol. 23 (1958), p. 458.

[13] TARSKI, A., with A. MOSTOWSKI and R. M. ROBINSON. *Undecidable Theories*. Amsterdam, North-Holland, 1953, xi + 98 pp.

[14] TRAHTÉNBROT, B. A. Impossibility of an algorithm for the decision problem in finite classes *Doklady Akadémii Nauk SSSR*, Vol. 70 (1950), pp. 569–572.

[15] TRAHTÉNBROT, B. A. On recursive separability. *Ibid.*, Vol. 88 (1953), pp. 953–956.

[16] TRAHTÉNBROT, B. A. Definition of finite sets and deductive incompleteness of the theory of sets. *Izvestiya Akadémii Nauk SSSR*, ser. mat., Vol. 20 (1956), pp. 569–582.

[17] USPÉNSKI, J. Theorem of Gödel and theory of algorithms. *Doklady Akadémii Nauk SSSR*, Vol. 91 (1953), pp. 737–740.

[18] VAUGHT, R. Sentences true in all constructive models. *Journal of Symbolic Logic*, Vol. 24 (1959), pp. 1–15.