THE JOURNAL OF SYMBOLIC LOGIC
Volume 34, Number 2, June 1969

DEFINABILITY IN THE MONADIC SECOND-ORDER THEORY OF SUCCESSOR[1]

J. RICHARD BÜCHI and LAWRENCE H. LANDWEBER

§1. Introduction. Let $\mathscr{D} = \langle D, P_1, P_2, \cdots \rangle$ be a relational system whereby D is a nonempty set and P_i is an m_i-ary relation on D. With $\mathscr{D}$ we associate the (weak) monadic second-order theory $(W)MT[\mathscr{D}]$ consisting of the first-order predicate calculus with individual variables ranging over D; monadic predicate variables ranging over (finite) subsets of D; monadic predicate quantifiers; and constants corresponding to $P_1, P_2, \cdots$. We will often use $(W)MT[\mathscr{D}]$ ambiguously to mean also the set of true sentences of $(W)MT[\mathscr{D}]$.

In this note we study variants of the structure $\langle N, ' \rangle$ where N is the set of natural numbers and $'$ is the successor function on N. Our results are a consequence of McNaughton's [7] work on the ω-behavior of finite automata and the decision procedure for $MT[N, ']$ given in [1]. The former is essential as we have been unable to obtain proofs which utilize only [1]'s characterization of ω-behavior. In [2] we discuss related results.

§2 studies definability in $MT[N, ']$. For every formula $C(X)$ of $MT[N, ']$ where X is a vector of unary predicate variables, the relation $C(X)$ is arithmetic and, in fact, is in the Boolean algebra over Π_2. In §3, we investigate the existence of decision procedures for $(W)MT[N, ', Q]$ where Q is a subset of N. Such theories were previously studied by Elgot and Rabin [4]. For any recursive Q, the decision problem for $MT[N, ', Q]$ is in $\Sigma_3 \cap \Pi_3$. We also define a recursive Q for which $(W)MT[N, ', Q]$ is undecidable. This provides a rather natural example of an undecidable theory which is still arithmetic.

§2. Definability in $MT[N, ']$. In this section we study definability in $MT[N, ']$ with respect to the arithmetic and classical Borel hierarchies. In particular we are interested in those relations definable by formulas $C(X)$, X a vector of free monadic predicate variables, of $MT[N, ']$. The main result is that every such relation is in the Boolean algebra over Π_2 of the arithmetic hierarchy. In fact, Lemma 1 below also gives this result for a wider class of $C(X)$ than are definable in $MT[N, ']$. In the following $x, y, z, \cdots$ are individual variables ranging over N.

Let Π_0 be the class of recursive relations on $N^n \times P(N)^k$ where $P(N)$ is the power set of N. Π_1 (Π_2) is the class of relations presentable in the form $(\forall y)C(y, x_1, \cdots, x_n, X_1, \cdots, X_k)$ $((\exists z)(\forall y)C(z, y, x_1, \cdots, x_n, X_1, \cdots, X_k))$ where C denotes a recursive relation. Relations in $\Pi_3, \Pi_4, \cdots$ are obtained by prefixing additional alternating quantifiers to relations in Π_2. The classes

Received October 10, 1967; revised July 22, 1968.

[1] This research was supported by the National Science Foundation (Contract 4730-50-395).

$\Pi_0, \Pi_1, \cdots$ comprise the *arithmetic hierarchy*. It is well known that $\Pi_{i+1} - \Pi_i \neq \varnothing$ for all i. Moreover, if Σ_i is the class of relations whose complements are in Π_i, then for all i, $\Pi_i \subset \Pi_{i+1} \cap \Sigma_{i+1}$. We refer the reader to Kleene [6] and Rogers [9, Chapters 14–15] for a complete discussion of the properties of the arithmetic hierarchy.

A formula $C(x_1, \cdots, x_n, X_1, \cdots, X_k)$ of $MT[N, ']$ is in $\Pi_k(\Sigma_k)$ if the corresponding relation is in $\Pi_k(\Sigma_k)$. To simplify the notation we do not distinguish between formulas and the relations they define. X is always used as an abbreviation for a vector of unary predicate variables. We implicitly use the obvious correspondence between ω-sequences on $\{T, F\}^k$, k-tuples of unary predicates on N and k-tuples of subsets of N. Let $I_n = \{T, F\}^n$. $I_n{}^*$ is the set of finite sequences on I_n. To simplify the notation we omit the subscript on I_n.

A recursive operator (RO) $Z = \mathscr{A}(X)$ is an operator mapping ω-sequences over the finite set $I = \{T, F\}^n$ into ω-sequences over a finite set S which can be presented in the form

$$Zt = \Phi(\bar{X}\phi(t)) \tag{1}$$

whereby $\bar{X}t = X0 \cdots Xt$ and Φ and ϕ are recursive functions from I^* into S and from N into N respectively. Sup Z is the set of members of S appearing infinitely often in the ω-sequence $Z = Z0, Z1, \cdots$.

LEMMA 1. *Let $Z = \mathscr{A}(X)$ be a RO and $U \subseteq 2^S$. Then the relation $F(X)$ given by*

$$(\exists Z)[Z = \mathscr{A}(X) \wedge \sup Z \in U] \tag{2}$$

is in the Boolean algebra over Π_2 of the arithmetic hierarchy.

PROOF. $F(X)$ can be written as

$$\bigvee_{B \in U} \centerdot (\exists x)(\forall y)[y \geq x \supset \Phi(\bar{X}\phi(y)) \in B] \wedge \bigwedge_{s \in B} (\forall x)(\exists y)[y \geq x \wedge \Phi(\bar{X}\phi(y)) = s].$$

The relations given by $[y \geq x \wedge \Phi(\bar{X}\phi(y)) = s]$ and $[y \geq x \supset \Phi(\bar{X}\phi(y)) \in B]$ are recursive because Φ and ϕ are recursive. Hence $F(X)$ is a Boolean combination of formulas of the form $(\forall y)(\exists x)M(X, x, y)$ where M is recursive so $F(X)$ is in the Boolean algebra over Π_2. Q.E.D.

A finite automata operator (FAO) is a RO $Z = \mathscr{A}(X)$ which can be presented in the form

$$Z0 = c, \qquad Zt' = H[Xt, Zt] \tag{3}$$

whereby $H: I \times S \to S$ and $c \in S$. Let $C(X)$ be a formula of $MT[N, ']$. The main definability results of [1] and [7] (see [2] for more details) state that from C we can effectively construct a presentation of a FAO $Z = \mathscr{E}(X)$ as in (3) (i.e., obtain H, S, and c) and a $U \subseteq 2^S$ such that

$$C(X) \centerdot\equiv\centerdot (\exists Z)[Z = \mathscr{E}(X) \wedge \sup Z \in U].$$

Hence by Lemma 1 we have

THEOREM 1. *Every relation between subsets of N which is definable in $MT[N, ']$ is arithmetical, and in fact occurs in the Boolean algebra over Π_2. Furthermore, given a formula $C(X_1, \cdots, X_n)$ of $MT[N, ']$ one can construct an index of the relation C in the Boolean algebra over Π_2.*

In contrast, all relations $R(y_1, \cdots, y_m, X_1, \cdots, X_n)$ appearing in the function-quantifier hierarchy over recursive relations are definable in $MT[N, ', 2x]$ (see [8]).

We can also consider $C(X)$ as defining a subset of the Cantor space of ω-sequences over I, namely, the set of ω-sequences over I which satisfy C. Those sets that are both open and closed in the usual totally disconnected topology on this space are of the form $U_{w_1} \cup \cdots \cup U_{w_n}$ whereby $w_i \in I^*$ and $U_w = \{X \mid (\exists t)[\bar{X}t = w]\}$. A set is open if it is a denumerable union of sets which are both open and closed. $G_\delta(F_\sigma)$ is the class of sets which are denumerable intersections (unions) of open (closed) sets. $G_{\delta\sigma}, G_{\delta\sigma\delta}, \cdots$ and $F_{\sigma\delta}, F_{\sigma\delta\sigma}, \cdots$ sets are defined in the obvious manner. The *Borel hierarchy* is the increasing sequence of classes $G, G_\delta, G_{\delta\sigma}, \cdots$ (see [9, Chapter 15] for a comparison of the Borel and arithmetic hierarchies).

If C is recursive, there is an effective procedure which decides whether $C(X)$ or $\sim C(X)$ is true after being given some finite portion $\bar{X}t = X0 \cdots Xt$ of X. Hence, if X_0 is such that $\bar{X}_0 t = \bar{X}t$, then $C(X) \equiv C(X_0)$. This implies that every recursive set of X's is open and closed. But every $C(X)$ of $MT[N, ']$ is a Boolean combination of expressions of the form $(\forall x)(\exists y)M(x, y, X)$ where for fixed x and y $\hat{X}M(x, y, X)$ is open and closed (since M is recursive). Thus by Theorem 1 we obtain

COROLLARY 1. *If $C(X)$ is a formula of $MT[N, ']$, then the relation $C(X)$ is in the Boolean algebra over G_δ of the Borel hierarchy.*

We conclude this section with an example of a $C(X)$ of $MT[N, ']$ which is neither a G_δ nor an F_σ (and therefore neither a Σ_2 nor a Π_2). The following remark is observed in [3].

(1) A set $C(X)$ is a G_δ, if and only if, there is a set W of words over I such that $C(X)$ holds if and only if $w < X$ for infinitely many $w \in W$.

Here $w < X$ (w is initial segment of X) stands for $(\exists t)\bar{X}t = w$. Now define $C(X)$ by,

(2) $$[X0 \wedge (\forall x)(\exists y)[x \leq y \wedge Xy]] \vee [\sim X0 \wedge (\exists x)(\forall y)[x \leq y \supset \sim Xy]].$$

Suppose C is a G_δ. Then, by (1), there exists a $W \subseteq I^*$ such that

(3) $$C(X) .\!\equiv\!. W \cap \{w \mid w < X\} \text{ is infinite.}$$

Define the sequence $w_0, w_1, w_2, \cdots$ by

(4) $$\begin{aligned} w_0 &= \text{shortest } v,\ v \in W \wedge v \text{ of form } FF^k, \\ w_{n+1} &= \text{shortest } v,\ v \in W \wedge v \text{ of form } w_n TFF^k. \end{aligned}$$

By (2) F^ω belongs to C, therefore by (3) w_0 exists and $F \leq w_0$. Assume inductively that w_n exists and $F \leq w_n$. Then by (2) $w_n TF^\omega$ belongs to C, therefore by (3) w_{n+1} exists and $F \leq w_{n+1}$. Thus (4) really defines a sequence of words, and clearly $w_i \in W$, $F \leq w_0 < w_1 < w_2 \cdots$. Thus, by (3) and (2), the sequence Y having all w_i's as initial segments belong to C. But this is contradictory, as Y starts with F and has infinitely many T's. Thus $C \notin G_\delta$, and similarly one shows $\sim C \notin G_\delta$. But $x \leq y$ is definable in $MT[N, ']$, and therefore C is. Consequently, (2) provides an example of a set C, definable in $MT[N, ']$, but neither in G_δ nor F_σ.

§3. Decision problems for extensions of $MT[N, ']$. Elgot and Rabin [4] have studied the existence of decision procedures for extensions of $MT[N, ']$. In particular they have shown that $MT[N, ', Q]$ is decidable if Q is either of $\{x^k \mid x \in N\}$, $\{k^x \mid x \in N\}$ or $\{x! \mid x \in N\}$ where k is a fixed natural number. The results are obtained by reducing the decision problem for $MT[N, ', Q]$ to that for $MT[N, ']$ and then applying the procedure given in [1]. If $Q = \{(x, 2x) \mid x \in N\}$, then the corresponding weak monadic theory is undecidable [8].

Let Q be a subset of N. If $WMT[N, ', Q]$ is undecidable, then so is $MT[N, ', Q]$. This follows from the definability of 'X is a finite set' in $MT[N, ']$, by the formula $(\exists x)(\forall t)[t \geq x \supset \sim Xt]$ where $t \geq x$ is an abbreviation of $(\forall Y) . Yt \wedge (\forall w)[Yw' \supset Yw] \supset Yx$.

If Q is not recursive, then $WMT[N, ', Q]$ is undecidable (e.g., $0'^{\cdots\prime} \in Q$ can not be effectively decided). If Q is recursive, the hierarchy result of §2 can be applied to give an upper bound to the complexity of decision problems for $MT[N, ', Q]$. $\psi(y, Z)$ is a universal predicate for Π_2 if for each $P(Z) \in \Pi_2$, there is an e_p such that for all Z, $\psi(e_p, Z) \equiv P(Z)$.

THEOREM 2. *If Q is recursive, then truth in $MT[N, ', Q]$ is in $\Sigma_3 \cap \Pi_3$.*

PROOF. Let $\Psi(e, Z)$ be a universal predicate for all predicates $P(Z)$ in Π_2, which is itself in Π_2 [6]. By Theorem 1, there is a recursive function B which maps every formula $\Phi(Z)$ of $MT[N, ']$ into a Boolean expression B_Φ, and a recursive function f which maps every formula $\Phi(Z)$ of $MT[N, ']$ into a finite sequence $f_\Phi = \langle f_{\Phi,1}, \cdots, f_{\Phi,n} \rangle$ of numbers, such that for any $Z \subseteq N$,

(1) $\Phi(Z)$ holds in $MT[N, '] . \equiv . B_\Phi[\Psi(f_{\Phi,1}, Z), \cdots, \Psi(f_{\Phi,n}, Z)]$.

Let $\chi(e)$ stand for $\Psi(e, Q)$, and note that because $\Psi \in \Pi_2$ and Q is recursive it follows that $\chi \in \Pi_2$. Furthermore, (1) may be restated as,

(2) $\Phi(Q)$ holds in $MT[N, ', Q] . \equiv . B_\Phi[\chi(f_{\Phi,1}), \cdots, \chi(f_{\Phi,n})]$.

Note that the functions B, f are recursive, and all sentences of $MT[N, ', Q]$ are of form $\Phi(Q)$ where $\Phi(Z)$ is a formula of $MT[N, ']$. It follows that (2) provides for a recursive reduction of $\{\Sigma \mid \Sigma \text{ true in } MT[N, ', Q]\}$ to the set χ (i.e. a Turing machine can be built which, given a sentence Σ of $MT[N, ', Q]$ and an oracle for membership in χ, decides whether or not Σ is true). Thus, truth in $MT[N, ', Q]$ is reducible to some $\chi \in \Pi_2$. It follows, by a well-known result of Post (see [9, p. 314]), that truth in $MT[N, ', Q]$ belongs to $\Sigma_3 \cap \Pi_3$. Q.E.D.

Theorem 2 shows that for no recursive Q is it possible to prove $MT[N, ', Q]$ undecidable by the standard method of showing that all recursive relations are definable.

If Q is the set of primes, then $(\forall x)(\exists y)[y > x \wedge Q(y) \wedge Q(y'')]$ states the twin prime problem in $MT[N, ', Q]$. Indeed, this sentence is in the first order theory of $\langle N, ', <, Q\rangle$. Hence, the problem as to whether $(W)MT[N, ', \text{primes}]$ is decidable, would seem very difficult. Namely, a positive answer would settle the twin prime problem, while on the negative side, the standard methods of proving theories undecidable is not available.

THEOREM 3. *There is a recursive Q such that $WMT[N, ', Q]$ is undecidable.*[2]

PROOF. Let R be a recursively enumerable set of primes which is not recursive. Let $r_1, r_2, \cdots$ be a recursive enumeration of R and let $Q_0 = \{r_i^2 p_i \mid i = 1, 2, \cdots\}$,

[2] Michael O. Rabin has obtained a similar result (personal correspondence).

whereby p_i is the ith prime. Q_0 is obviously recursive. To prove that $WMT[N, ', Q_0]$ is undecidable it is sufficient to show that the first order theory (FT) of $\langle N, M_1, M_2, \cdots, Q_0\rangle$ is undecidable whereby M_k stands for the set of multiples of k. Just note that each M_k is definable in $WMT[N, ', Q_0]$ by the formula

$$M_k(w): (\forall X). Xw \wedge (\forall y)[X(y + k) \supset Xy] \supset X0.$$

From the definition of R and Q_0 we obtain

$$(*) \qquad R(k) .\equiv. k \neq 1 \wedge (\exists y)[M_{k^2}(y) \wedge Q_0(y)].$$

Let Σ_k be the sentence $k \neq 1 \wedge (\exists y)[M_{k^2}(y) \wedge Q_0(y)]$. By (*) Σ_k is true in $FT[N, M_1, M_2, \cdots, Q_0]$ if and only if $k \in R$. But R is not recursive so there is no effective procedure for deciding truth in $FT[N, M_1, M_2, \cdots, Q_0]$. Q.E.D.

PROBLEM 1. Is there an 'interesting' recursive Q such that $(W)MT[N, ', Q]$ is undecidable? How about Q = primes?

Although $WMT[N, ', Q_0]$ is undecidable, we have not classified its decision problem in the arithmetic hierarchy. This suggests

PROBLEM 2. Is there a recursive Q such that the decision problem for $(W)MT[N, ', Q]$ is in $\Sigma_3 \cap \Pi_3$ but not in the Boolean algebra over Π_2?

Another interesting question is,

PROBLEM 3. Is there a recursive Q such that $WMT[N, ', Q]$ is decidable but $MT[N, ', Q]$ is undecidable?

A negative answer to Problem 3 should imply the decidability of $MT[N, ']$ as a consequence of the decidability of $WMT[N, ']$ ($Q = \varnothing$). Hence, a negative answer might be quite difficult.

BIBLIOGRAPHY

[1] J. R. BÜCHI, *On a decision procedure in restricted second order arithmetic*, ***Proceedings of the international congress on logic, methodology and the philosophy of science***, Stanford University Press, Stanford, California, 1962.

[2] J. R. BÜCHI and L. H. LANDWEBER, ***Solving sequential conditions by finite state operators***, Purdue Report CSD TR 14.

[3] M. DAVIS, *Infinitary games of perfect information*, ***Advances in game theory***, Princeton University Press, Princeton, New Jersey, 1964, pp. 85–101.

[4] C. C. ELGOT and M. O. RABIN, *Decidability and undecidability of extensions of second (first) order theories of (generalized) successor*, this JOURNAL, vol. 31 (1966), pp. 169–181.

[5] S. C. KLEENE, ***Introduction to metamathematics***, Van Nostrand, New York, Amsterdam and Noordhoff, Groningen, 1952.

[6] S. C. KLEENE, *Hierarchies of number theoretic predicates*, ***Bulletin of the American Mathematical Society***, vol. 61 (1955), pp. 193–213.

[7] R. MCNAUGHTON, *Testing and generating infinite sequences by a finite automaton*, ***Information and control***, vol. 9 (1966), pp. 521–530.

[8] R. M. ROBINSON, *Restricted set theoretical definitions in arithmetic*, ***Proceedings of the American Mathematical Society***, vol. 9 (1958), pp. 238–242.

[9] H. ROGERS, JR., ***Theory of recursive functions and effective computability***, McGraw-Hill, New York, 1967.

PURDUE UNIVERSITY
UNIVERSITY OF WISCONSIN