On Sets of Relations Definable by Addition

James F. Lynch


Stable URL: http://links.jstor.org/sici?sici=0022-4812%28198209%2947%3A3%5C659%5C3AOSORDB%5C3E2.0.CO%5C3B2-N

The Journal of Symbolic Logic is currently published by Association for Symbolic Logic.

Your use of the JSTOR archive indicates your acceptance of JSTOR’s Terms and Conditions of Use, available at http://uk.jstor.org/about/terms.html. JSTOR’s Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Please contact the publisher regarding any further use of this work. Publisher contact information may be obtained at http://uk.jstor.org/journals/asl.html.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

JSTOR is an independent not-for-profit organization dedicated to creating and preserving a digital archive of scholarly journals. For more information regarding JSTOR, please contact support@jstor.org.
ON SETS OF RELATIONS DEFINABLE BY ADDITION

JAMES F. LYNCH

Abstract. For every \( k \in \omega \), there is an infinite set \( A_k \subseteq \omega \) and a \( d(k) \in \omega \) such that for all \( Q_0, Q_1 \subseteq A_k \) where \( |Q_0| = |Q_1| \) or \( d(k) < |Q_0|, |Q_1| < \aleph_0 \), the structures \( \langle \omega, +, Q_0 \rangle \) and \( \langle \omega, +, Q_1 \rangle \) are indistinguishable by first-order sentences of quantifier depth \( k \) whose atomic formulas are of the form \( u = v \), \( u + v = w \), and \( Q_0 \), where \( u, v, \) and \( w \) are variables.

§1. Introduction. The results presented here are concerned with sets of relations defined in the following way. Let \( \sigma \) be a first-order sentence appropriate to structures of the form \( \langle \omega, +, Q \rangle \), where \( Q \) is a \( q \)-ary relation on \( \omega \), i.e. \( Q \subseteq \omega^q \). Then \( \{ Q \subseteq \omega^q : \langle \omega, +, Q \rangle \models \sigma \} \) is the set of relations defined by \( \sigma \) in \( \langle \omega, + \rangle \). This kind of definability was first studied by J. Mycielski [15], who proved that the set of relations \( C = \{ Q \subseteq \omega^q : Q \text{ is finite and connected} \} \) is not definable by any first-order sentence in \( \langle \omega, + \rangle \). (By connected we mean that \( \omega \) is regarded as the set of points in the cartesian plane whose coordinates are nonnegative integers, and a chess king can visit all points in \( Q \) without leaving \( Q \).)

Mycielski's argument runs as follows. He proves that if \( C \) is definable in \( \langle \omega, + \rangle \), then the relation \( D = \{ (x, y) \in \omega^2 : x \text{ divides } y \} \) is definable in \( \langle \omega, + \rangle \). By well-known results of K. Gödel [16, Chapter 15] and J. Robinson [19], the theory of \( \langle \omega, +, D \rangle \) is undecidable, and by a result of M. Presburger [16, Chapter 13], the theory of \( \langle \omega, + \rangle \) is decidable. Hence \( C \) is not definable in \( \langle \omega, + \rangle \).

Mycielski then asked if \( E = \{ Q \subseteq \omega : |Q| \text{ is finite and even} \} \) is definable in \( \langle \omega, + \rangle \), since his method could not solve this apparently simpler problem. Our method is indeed quite different. It is based on Ehrenfeucht games and is related to our former work [12].

We show that for every \( k \in \omega \) there is an integer \( d(k) \) and an infinite set \( A_k \subseteq \omega \) (in fact uncountably many such sets) which satisfy the following. For every \( Q_0, Q_1 \subseteq A_k \), if \( |Q_0| = |Q_1| \) or \( d(k) < |Q_0|, |Q_1| < \aleph_0 \), then the Ehrenfeucht game of length \( k \) on \( \langle \omega, +, Q_0 \rangle \) and \( \langle \omega, +, Q_1 \rangle \) is a win for player II.

Several further results follow from our construction of the sets \( A_k \). By analogy with the notion of indiscernibles [3], we show that \( A_k \) is a set of "\( k \)-indiscernibles". That is, let \( \sigma(v_1, \ldots, v_n) \) be a relational formula (i.e. its atomic formulas are of the form \( u = v \) and \( u + v = w \), where \( u, v, \) and \( w \) are variables) of quantifier depth \( k \) whose free variables are \( v_1, \ldots, v_m \), and let \( (a_1, \ldots, a_n) \) and \( (b_1, \ldots, b_n) \) be sequences in \( A_k \) such that \( a_i \leq a_j \iff b_i \leq b_j \) for \( 1 \leq i, j \leq n \). Then

Received April 25, 1980; revised November 16, 1980.

Research partially supported by NSF Grant No. MSC 78-01832.

© 1982, Association for Symbolic Logic
\[ \langle \omega, + \rangle \models \sigma(a_1, \ldots, a_n) \leftrightarrow \sigma(b_1, \ldots, b_n). \] We also show that there is a set \( A = \{ q_i : k \geq 1 \} \) (in fact uncountably many sets \( A \)) such that for each \( k \in \omega, \{ q_i : i \geq 1 \} \) is an \( A_k \) satisfying the above conditions.

Lastly, we consider the problem of definability of finite sets of finite relations. Of course such a set may be defined by enumerating its members, but one may still ask for the shortest sentence that defines it. We show that for sufficiently large \( n \in \omega \), any relational sentence that defines \( E_n = \{ Q \subseteq n : |Q| \text{ is even} \} \) in \( \langle \omega, +, n \rangle \) has quantifier depth greater than \( \frac{1}{2} \log_2 \log_2 \log_2 n. \)

The concluding section describes some related problems, involving the definability of sets of relations in languages more expressive than the first-order language of \( + \) but without the full power of \( + \) and \( \cdot \).

§2. Preliminaries. We use the standard notation of first-order logic (see e.g. Monk [16]). Formulas are constructed from atomic formulas by using the connectives \( \neg, \vee, \wedge \) (not, or, and, respectively) and the quantifiers \( \exists \) and \( \forall \) (there exists, for all, respectively). \( \sigma(v_1, \ldots, v_n) \) will denote a formula whose free variables are \( v_1, \ldots, v_n \). We will write \( \sigma \) if \( v_1, \ldots, v_n \) are understood. By a relational formula we mean one whose atomic formulas are of the form \( u = v, F(v_1, \ldots, v_m) = u \) where \( F \) is an \( m \)-ary function symbol, and \( P(v_1, \ldots, v_m) \) where \( P \) is an \( m \)-ary relation symbol; in our case we have \( u = v, u + v = w, \) and \( Q(u) \). We do not allow terms such as \( u_1 + \cdots + u_m \), although it is obvious that any formula with such terms is equivalent to a relational formula.

The depth of a formula is the maximum nesting of quantification. Inductively, if \( \sigma \) is atomic, then its depth is 0. If the depths of \( \sigma_1 \) and \( \sigma_2 \) are \( d_1 \) and \( d_2 \) respectively, then the depth of \( \neg \sigma \) is \( d_2 \), the depth of \( \sigma_1 \vee \sigma_2 \) and \( \sigma_1 \wedge \sigma_2 \) is \( \max(d_1, d_2) \), and the depth of \( \exists \forall \sigma_1 \) and \( \forall \exists \sigma_1 \) is \( d_1 + 1 \). A sentence is a formula with no free variables. Thus, for a given finite type and every \( k \in \omega \), there are only finitely many (up to equivalence) relational sentences of depth \( k \).

\( \omega \) is the set of nonnegative integers and \( \mathbb{Z} \) is the set of integers; every \( n \in \omega \) is identified with the set \( \{ 0, 1, \ldots, n - 1 \} \); and for any set \( A \) and \( n \in \omega, n^A \) is the set of \( n \)-tuples of elements in \( A \). \( |A| \) is the cardinality of \( A \).

A relational structure, or model, is a nonempty tuple \( \langle U, P_i \rangle_{i \in I} \), where \( U \) is a set (the universe of the structure), \( I \) is a set, and each \( P_i \) is a relation on \( U \). For any formula \( \sigma(v_1, \ldots, v_n) \) appropriate to a structure \( \mathfrak{A} \) with universe \( U \) and any \( a_1, \ldots, a_n \in U \), we put \( \mathfrak{A} \models \sigma(a_1, \ldots, a_n) \) if \( \sigma \) is true in \( \mathfrak{A} \) with \( a_i \) assigned to \( v_i, i = 1, \ldots, n \).

2.1 Definition. For \( j = 0, 1 \) let \( \mathfrak{A}_j = \langle U_j, P_{ji} \rangle_{i \in I} \) be two relational structures of the same type, i.e. for each \( i \in I \) there is a \( p_i \in \omega \) such that \( P_{ji} \subseteq (p_i)^{-1} U_j, j = 0, 1 \). For any \( k \in \omega \) and any sequences \( (a_1, \ldots, a_k) \in ^k U_0 \) and \( (b_1, \ldots, b_k) \in ^k U_1 \) we say that \( (a_1, \ldots, a_k) \) is isomorphic to \( (b_1, \ldots, b_k) \) if the structures \( \langle a_1, \ldots, a_k \rangle \) and \( \langle b_1, \ldots, b_k \rangle \) are isomorphic.

2.2 Definition. Let \( \mathfrak{A}_0, \mathfrak{A}_1 \) be as in 2.1 and \( k \in \omega \). The Ehrenfeucht game \( \Gamma_k (\mathfrak{A}_0, \mathfrak{A}_1) \) of length \( k \) is the following game of perfect information. It starts with player I choosing some element in one of the structures, say \( \mathfrak{A}_{j_1} \). Then player II chooses an element in \( \mathfrak{A}_{1-j_1} \). The game continues with the players alternately...
choosing elements, where the $i$th choice of player I is in either of the structures, say $\mathfrak{U}_i$, and the $i$th choice of player II is in $\mathfrak{V}_i$, until each player has chosen $k$ elements. Let $a_i$ be the $i$th element chosen in $\mathfrak{U}_0$ and $b_i$ be the $i$th element chosen in $\mathfrak{V}_0$. We will refer to the choosing of $a_i$ and $b_i$ as step $i$ of the game and the initial conditions as step 0. Player II wins if $(a_1, \ldots, a_k)$ and $(b_1, \ldots, b_k)$ are isomorphic.

The following theorem is central to our results.

2.3 Theorem (Ehrenfeucht [4]). Let $\mathfrak{U}_0$ and $\mathfrak{V}_1$ be two relational structures of the same type, and consider the following two conditions:

(i) $\mathfrak{U}_0$ and $\mathfrak{V}_1$ cannot be distinguished by any first-order sentence of depth $k$.

(ii) The game $^*\Gamma_4(\mathfrak{U}_0, \mathfrak{V}_1)$ is a win for player II.

Then (ii) implies (i). If $\mathfrak{U}_0$ and $\mathfrak{U}_1$ have a finite number of relations, then (i) and (ii) are equivalent.

As with most applications of this theorem, we need only the part that states (ii) implies (i). When we describe an Ehrenfeucht game on structures of the form $\langle \omega, +, Q \rangle$, where $Q \subseteq \omega$, it is understood that we are considering relational structures, i.e. we treat the $+$ operator as a ternary relation.

§3. Results. Let

(3.1) $d(0) = 5$, $d(i + 1) = (2^{i+3} + 1)d(i)$,

(3.2) $f(0) = 1$, $f(i + 1) = 2f(i)^4$,

(3.3) $g(0) = 0$, $g(i + 1) = 2f(i)^2 g(i) + f(i)^4$,

(3.4) $h(0) = 1$, $h(i + 1) = 2f(i)^2 h(i)^2 f(i + 1)^2 + 3^3$.

Although not needed in our proofs, the growth rates of these functions are given by the following easily verified formulas:

\[
\begin{align*}
\log_2 d(i) &= \frac{1}{2} (i^2 + 5i + 4 + b(i)) \text{ where } 0 < b(i) < \log_2 e, \\
\log_2 f(i) &= (4^i - 1)/3, \\
g(i) &= (1 + c(i))(2^{(4^i - 1)/3})! \text{ where } 0 < c(i) \text{ and } c(i) \to 0, \\
\log_2 h(i) &= [2^{i+3}(4^i - 1)/3 - i2^{i+2} + 4^i - 1]/3.
\end{align*}
\]

Choose $k \in \omega$, and let $\langle p_i : i \in \omega \rangle$ be any sequence in $\omega$ such that $p_0 = 0$ and

(3.5) $p_{i+1} \geq 2^{k+3} f(k)^3 p_i + 2f(k)^2 g(k)$ for $i \in \omega$, and $p_i \equiv p_j$ (mod $f(k)!$) for $1 \leq i, j \in \omega$.

For example, we could take

(3.6) $p_i = f(k)! \sum_{j=0}^{k-1} (2^{k+3} f(k)^3)j$ for $k \geq 2$.

Let $A_k = \{p_i : i \geq 1\}$.

3.7 Theorem. Let $Q_0$ and $Q_1$ be any subsets of $A_k$ such that $|Q_0| = |Q_1|$ or $d(k) < |Q_0|$, $|Q_1| < \aleph_0$. Then $^*\Gamma_4(\langle \omega, +, Q_0 \rangle, \langle \omega, +, Q_1 \rangle)$ is a win for player II.

The proof of this theorem is given in the next section. This theorem, together with Theorem 2.3, immediately yields the following.

3.8 Corollary. For all $Q_0, Q_1 \subseteq A_k$ such that $|Q_0| = |Q_1|$ or $d(k) < |Q_0|$, $|Q_1| < \aleph_0$, $\langle \omega, +, Q_0 \rangle$ and $\langle \omega, +, Q_1 \rangle$ are indistinguishable by relational sentences of depth $k$.

The next results are proven by making slight modifications to the proof of 3.7.

3.9 Theorem. $A_k$ is a set of $k$-indiscernibles for $\langle \omega, + \rangle$.

3.10 Theorem. For $m \in \omega$, let $A_{km} = \{p_i : 1 \leq i < m\}$. If $n \geq p_m$, then for any $Q_0, Q_1 \subseteq A_{km}$ such that $|Q_0| = |Q_1|$ or $|Q_0|, |Q_1| > d(k)$, $^*\Gamma_4(\langle \omega, +, n, Q_0 \rangle, \langle \omega, +, n, Q_1 \rangle)$ is a win for player II.
3.11 Theorem. Let $E_n = \{Q \subseteq n: |Q| \text{ is even}\}$. Then for sufficiently large $n$, any relational sentence that defines $E_n$ in $\langle \omega, +, n \rangle$ has depth greater than $\frac{1}{2} \log \log \log n$.

Lastly, let $A = \{q_k: \ k \geq 1\}$, where $q_0 = 0$, and $q_{k+1} \geq 2^{k+3}f(k)^3q_k + 2f(k+1)^2g(k+1)$ and $q_k \equiv 0 \pmod{f(k)!}$ for $k \in \omega$. Then for each $k \geq 1$, $\{q_i: i \geq k\}$ satisfies (3.5), and the following holds.

3.12 Corollary. For every $k \geq 1$, the set $\{q_i: i \geq k\}$ satisfies the above theorems.

§4. Proof of Theorem 3.7.

4.1 Definitions. For $j = 0, 1$ let $\mathcal{A}_j = \langle \omega, +, Q_j \rangle$, and for $x, y \in Q_j$ let $\delta_j(x, y) = |\{z \in Q_j: x \leq z \leq y \text{ or } y \leq z \leq x\}| - 1$. For $0 \leq i \leq k$, an $i$-vector in $\mathcal{A}_0$ is a sequence $s$ of the form $(x_1, \ldots, x_n, \alpha_1, \ldots, \alpha_n, \rho)$ where

(i) $\rho = u/v$, $u, v \in \mathbb{Z}$, $|\rho| \leq g(-i)$, and $|v| \leq h(k - i)$,

(ii) $n \leq 2^{k-i+1}$,

(iii) for each $j = 1, \ldots, n$, $\alpha_j = u_j/v_j$, where $u_j, v_j \in \mathbb{Z}$ and $|u_j|, |v_j| \leq f(k - i)$,

(iv) for each $j = 1, \ldots, n$, either $x_j \in \{a_1, \ldots, a_n\}$, where $a_m$ is the element in $\mathcal{A}_0$ chosen at step $m$ of $\Gamma_k(\mathcal{B}, \mathcal{A}_0)$, or $x_j \in Q_0$ and $x_j \neq a_m$ for $1 \leq j < m \leq n$.

An $i$-vector in $\mathcal{A}_1$ is defined similarly. For $i < k$, a minor $i$-vector is an $i$-vector where, instead of (i) above, we have

(i') $\rho = u/v$, $u, v \in \mathbb{Z}$,

$|\rho| \leq 2f(k - i - 1)^2g(k - i - 1)$, and $|v| \leq 2f(k - i - 1)^2h(k - i - 1)^2$.

By (3.3) and (3.4), a minor $i$-vector is an $i$-vector.

The elements $x_1, \ldots, x_n$ will be referred to as the terms of the $i$-vector $s$. We put $\delta = \sum_{j=1}^n \alpha_j x_j + \rho$.

The choices made by the players at each step $i$ in the game $\Gamma_k(\mathcal{B}, \mathcal{A}_0)$ determine two sets $B_{ji} \subseteq Q_j (j = 0, 1)$. If $|\mathcal{B}_0| = |\mathcal{Q}_0|$ then $B_{ji} = Q_j$. If $d(k) < |\mathcal{B}_0|, |\mathcal{Q}_0| < \mathcal{B}_0$ then $B_{ji} = \{\min(Q_j), \max(Q_j)\}$, and in the description of player II's strategy below, $B_{ji+1}$ is defined inductively from $B_{ji}$ in such a way that $B_{ji} \subseteq B_{ji+1}$. We put $C_{ji} = C_{0ji} \cup \{a_1, \ldots, a_i\}$ and $C_{ji+1} = C_{1ji} \cup \{b_1, \ldots, b_i\}$.

An $i$-correspondence $c$ is a mapping from $C_{0ji} \cup D_0$ to $C_{1ji} \cup D_1$ where $D_j$ is a subset of $Q_j (j = 0, 1)$, $c$ is one-to-one and order preserving on $C_{0ji} \cup D_0$, $c(B_{0ji}) = B_{1ji}$, $c(D_0) = D_1$, and for $1 \leq m \leq i$, $c(a_m) = b_m$.

If $s = (x_1, \ldots, x_n, \alpha_1, \ldots, \alpha_n, \rho)$ is an $i$-vector whose terms are in the domain of $c$, then we put $c(s) = (c(x_1), \ldots, c(x_n), \alpha_1, \ldots, \alpha_n, \rho)$.

4.2 Player II's strategy. We assume player I has chosen $a_i$ in $\mathcal{A}_0$. The case when player I has chosen $b_i$ in $\mathcal{A}_1$ is symmetric. If $a_i \in Q_0$ we take $s_m = s_M = (a_i, 1, 0)$. If $a_i \notin Q_0$ but there is some $(i - 1)$-vector $s$ such that $s = a_i$, we take $s_m = s_M = s$.

If there is no $(i - 1)$-vector $s$ such that $s = a_i$, let $s_m$ be a minor $(i - 1)$-vector such that $s_m$ is maximal among all minor $(i - 1)$-vectors $s$ such that $s < a_i$, and let $s_M$ be a minor $(i - 1)$-vector such that $s_M$ is minimal among all minor $(i - 1)$-vectors $s$ such that $a_i < s$. If there are no $s$ such that $a_i \leq s$, then $s_M$ is undefined.

Let $x_1, \ldots, x_n$ be the terms of $s_m$ and $s_M$ that are in $Q_0$. Then $B_{0ji} = B_{0ji-1} \cup \{x_1, \ldots, x_n\}$. We will show that there exists an $(i - 1)$-correspondence $c$ such that $x_1, \ldots, x_n$ are in the domain of $c$ and for all $x, y \in B_{0ji}$,

$\delta_0(x, y) = \delta_1(c(x), c(y)) < d(k - i)$,
Then \( B_{1i} = c(B_{0i}) \). We then show that there exists some \( b_i \in \omega \) such that \( a_i \equiv b_i \pmod{f(k - i)!} \) and \( \overline{c(s_m)} \leq b_i \leq \overline{c(s_M)} \) (or \( \overline{c(s_m)} < b_i \) if \( s_M \) is undefined). Player II chooses such a \( b_i \).

The proof that this is a winning strategy for player II consists in showing that the following conditions hold at the end of each step \( i \), if player II follows the strategy.

(4.3) \( a_i \equiv b_i \pmod{f(k - i)!} \).

(4.4) The set of \( i \)-correspondences is nonempty.

(4.5) Let \( x, y \in B_{0i} \) and \( c \) be an \( i \)-correspondence. Then either

\[ \delta_0(x, y) = \delta_1(c(x), c(y)) < d(k - i), \quad \text{or} \]

\[ \delta_1(x, y), \delta_1(c(x), c(y)) \geq d(k - i). \]

(4.6) Let \( s_1 \) and \( s_2 \) be \( i \)-vectors in \( \mathbb{H}_0 \) whose terms are in the domain of an \( i \)-correspondence \( c \). Then \( \bar{s}_1 \leq \bar{s}_2 \) if and only if \( c(s_1) \leq c(s_2) \).

(4.7) \( a_i \in \mathcal{Q}_0 \) if and only if \( b_i \in \mathcal{Q}_1 \).

It is clear that if (4.4), (4.6), and (4.7) hold for all steps 1 through \( k \) then \((a_1, \ldots, a_k)\) is isomorphic to \((b_1, \ldots, b_k)\) and player II has won. The series of lemmas below will show that (4.3) through (4.7) hold for \( i = 0 \), and if they hold for \( i - 1 \) then player II can play according to the strategy of 4.2 and they will hold for \( i \).

4.8 Lemma. (4.3) through (4.7) hold at step 0.

Proof. (4.3) and (4.7) hold vacuously, and (4.4) and (4.5) are immediate from the definition of \( B_{00} \) and \( B_{10} \) in 4.1 above.

To prove (4.6), let \( s_1 = (x_1, \ldots, x_m, a_1, \ldots, a_m, \beta) \) and \( s_2 = (y_1, \ldots, y_p, \gamma_1, \ldots, \gamma_p, \delta) \) be 0-vectors in \( \mathbb{H}_0 \), and let \( c \) be a 0-correspondence whose domain includes \( x_1, \ldots, x_m, y_1, \ldots, y_p \). Suppose \( \bar{s}_1 \leq \bar{s}_2 \). Treating the distinct terms of \( s_1 \) and \( s_2 \) as independent vectors over the field of rational numbers, we get an inequality

\[ (4.9) \ 0 \leq \sum_{j=1}^{q} \varepsilon_j z_j + \delta - \beta \]

where each \( z_j \) is a term of \( s_1 \) or \( s_2 \) and \( z_j \neq z_j \) for \( 1 \leq i < j \leq q \).

If \( \varepsilon_j = 0 \) for \( j = 1, \ldots, q \) then by reversing the steps used to get (4.9) with \( z_j \) replaced by \( c(z_j) \) for \( 1 \leq j \leq q \), we get \( c(s_1) \leq c(s_2) \).

If some \( \varepsilon_j \neq 0 \), let \( z_1 \) be the largest \( z_j \) such that \( \varepsilon_j \neq 0 \), say \( z_1 = p_{r+1} \) as in (3.5). We claim \( \varepsilon_1 > 0 \).

By 4.1(ii), \( q \leq 2^{k+2} \). By 4.1(iii), \( |\varepsilon_j| \leq 2f(k) \) for \( 1 \leq j \leq q \). Therefore \( \sum_{j=1}^{q} \varepsilon_j z_j \leq 2^{q+3} f(k)p_i \). Also, \( \delta - \beta \leq 2g(k) \) by 4.1(i), and \( |\varepsilon_1| > 1/f(k)^2 \) by 4.1(iii). If we assume \( \varepsilon_1 < 0 \) then \( z_1 < 2^{q+3} f(k)^3 p_i + 2f(k)^2 g(k) \). But this contradicts condition (3.5), and therefore \( \varepsilon_1 > 0 \). This, together with the fact that \( c(z_1) > c(z_j) \) for \( 2 \leq j \leq q \) such that \( \varepsilon_j \neq 0 \), implies

\[ 0 \leq \sum_{j=1}^{q} \varepsilon_j c(z_j) + \delta - \beta \]

by similar reasoning.

Again, reversing the steps used to get (4.9) with \( z_j \) replaced by \( c(z_j) \) for \( 1 \leq j \leq q \), we obtain \( c(s_1) \leq c(s_2) \). The proof that \( c(s_1) \leq c(s_2) \) implies \( \bar{s}_1 \leq \bar{s}_2 \) is similar. Q.E.D.

We now assume (4.3) through (4.7) hold for \( i - 1 \) where \( 1 \leq i \leq k \). Let player I choose \( a_i \) in \( \mathbb{H}_0 \) and let \( s_m, s_M \), and \( B_{0i} \) be as in 4.2.
Lemma. There is an \((i-1)\)-correspondence \(c\) such that for all \(x, y \in B_{0i}\) either

\[
\delta_0(x, y) = \delta_1(c(x), c(y)) < d(k-i), \quad \text{or} \quad 
\delta_0(x, y), \delta_1(c(x), c(y)) \geq d(k-i).
\]

Proof. If \(B_{0i, i-1} = B_{0i}\), then the lemma follows from the induction hypothesis for (4.5) and the fact that \(d(k-i) < d(k-i+1)\).

If \(B_{0i, i-1} \neq B_{0i}\), then \(|B_{0i}| < \kappa_0\), so let \(B_{0i} = \{u_0, \ldots, u_{p-1}\}\) where \(u_j < u_{j+1}\) for \(j < p - 1\). Let \(q \geq 1\) be such that \(u_q \in B_{0, i-1}\) and \(u_j \notin B_{0i, i-1}\) for \(1 \leq j < q\). The existence of \(q\) follows from \(u_{p-1} \in B_{00} \subseteq B_{0, i-1}\) (see 4.1).

By (4.4) there is an \((i-1)\)-correspondence \(c\) with domain \(C_{0, i-1}\). We will show how to extend \(c\) to \(\{u_1, \ldots, u_{q-1}\}\) in such a way that the lemma is satisfied for \(x, y \in \{u_0, \ldots, u_q\}\). Since for \(m = 1, \ldots, i-1, a_m \in Q_0\) implies \(a_m \in B_{0, i-1}\) by 4.2, \(\{u_1, \ldots, u_{q-1}\} \cap C_{0, i-1} = \varnothing\), and the extension will be consistent.

Case I. \(\delta_0(u_0, u_q) < d(k-i+1)\). By (4.5), \(\delta_0(u_0, u_q) = \delta_1(c(u_0), c(u_q))\). Then we can extend \(c\) to \(\{u_1, \ldots, u_{q-1}\}\) so that \(\delta_0(u_j, u_m) = \delta_1(c(u_j), c(u_m))\) for \(0 \leq j, m \leq q\).

Case II. \(\delta_0(u_0, u_q) \geq d(k-i+1)\). Then there exists some \(t < q\) such that \(\delta_0(u_t, u_{t+1}) \geq d(k-1)\). If there were no such \(t\), then since \(q \leq 2^{k+i+3} + 1\) by 4.1(ii), we would have \(\delta_0(u_0, u_q) < (2^{k+i+3} + 1) d(k-i) = d(k-i+1)\) by (3.1).

We now extend \(c\) to \(\{u_1, \ldots, u_{q-1}\}\) as follows. Since \(u_0 \in B_{0, i-1}\), \(c(u_0)\) is already defined. Now assume \(c(u_j)\) has been defined, where \(j < t\). We define \(c(u_{t+1})\) to be that element of \(Q_1\) such that \(c(u_t) < c(u_{t+1})\) and \(\delta_1(c(u_t), c(u_{t+1})) = \min(\delta_0(u_j, u_{t+1}), d(k-i))\). We extend \(c\) to \(\{u_{t+1}, \ldots, u_{q-1}\}\) in a similar fashion, using a decreasing induction on \(j\) from \(q\) to \(t+1\). To show that the lemma is satisfied for \(x, y \in \{u_0, \ldots, u_q\}\), we need only show \(c(u_t) < c(u_{t+1})\) and \(\delta_1(c(u_t), c(u_{t+1})) \geq d(k-1)\). But if this were not the case, then \(\delta_1(c(u_0), c(u_q)) < (2^{k+i+3} + 1) d(k-i) = d(k-i+1)\), which contradicts (4.5) and our assumption that \(\delta_0(u_0, u_q) \geq d(k-i+1)\).

Repeating this procedure on \(\{u_q, \ldots, u_{p-1}\}\), we can extend \(c\) to \(B_{0i}\) in such a way that the lemma is satisfied. Q.E.D.

In Lemmas 4.11 through 4.13 below, \(c\) will be the \((i-1)\)-correspondence whose existence was proven in the preceding lemma. Then \(B_{1i} = c(B_{0i})\), and (4.5) is established for step \(i\).

Lemma. There is no \((i-1)\)-vector \(s'\) in \(\mathfrak{A}_1\) such that \(c(s_m) < s' < c(s_M)\) (or \(c(s_m) < s'\) if \(s_M\) is undefined).

Proof. Suppose there were such a \(s'\). Assuming \(|Q_0| < \kappa_0\), let \(B_{0i} = \{u_0, \ldots, u_{p-1}\}\) as in the proof of Lemma 4.10, \(u_j = c(u_j)\) for \(j < p\), and \(S_j = \{z' \in Q_1: u_j < z' < u_{j+1}\}\) and \(z'\) is a term of \(s'\) for \(j < p - 1\). Now for each \(j < p - 1\), \(|S_j| \leq 2^{k-i+2} \leq 2^{k-i+3} < d(k-i)\) by (3.1).

Then by Lemma 4.10, \(|S_j| \geq |S_j'|\), where \(S_j = \{z \in Q_0: u_j < z < u_{j+1}\}\).

We claim \(S_j \cap C_{0i} = \varnothing\). Otherwise, let \(z \in S_j \cap C_{0i}\). Recalling that \(C_{0i} = B_{0i} \cup \{a_i, \ldots, a_i\}\), then \(z \in B_{00}\) by the definition of \(B_{0i}\) in 4.2. But \(S_j \cap B_{00} = \varnothing\), a contradiction. Therefore, letting \(T_j\) be any subset of \(S_j\) of cardinality \(|S_j'|\), we can extend \(c\) to an \((i-1)\)-correspondence \(c_1\) such that \(c_1(T_j) = S_j'\) for \(j < p - 1\), i.e. the range of \(c_1\) includes all of the terms of \(s'\). Then \(s_m < c_1^{-1}(s') < s_M\) by (4.6), which contradicts our definition of \(s_m\) and \(s_M\) (see 4.2). Therefore no such \(s'\) exists.
If \(|Q_0| = \mathbb{N}_0\), then by 4.1, \(B_{1i} = Q_1\), and the range of \(c\) already includes all of the terms of \(s'\). We again conclude \(s'\) cannot exist. Q.E.D.

4.12 Lemma. There exists a \(b_i\) such that \(a_i \equiv b_i \pmod{f(k - i)!}\), if \(\bar{s}_m = a_i\), then \(\bar{c}(s_{m}) = b_i\), and if \(\bar{s}_m < a_i < \bar{s}_M\) then \(c(s_{m}) < b_i < \bar{c}(s_{M})\) (or \(\bar{c}(s_{m}) < b_i\) if \(s_M\) is undefined).

Proof. If \(s_M\) is undefined the result is immediate; thus let \(s_m = (x_1, \ldots, x_n, \alpha_1, \ldots, \alpha_n, \beta)\) and \(s_M = (y_1, \ldots, y_p, \gamma_1, \ldots, \gamma_p, \delta)\). If \(\bar{s}_m = a_i\), we take \(b_i = \bar{c}(s_{m})\). Now, for every \(j = 1, \ldots, n, \alpha_j = u_j/v_j\), where \(|v_j| \leq f(k - i + 1)\) by 4.1(iii). If \(x_j \in Q_0\) then \(c(x_j) \in Q_1\) and \(x_j \equiv c(x_j) \pmod{f(k)!}\) by (3.5), and if \(x_j = a_m\) for some \(m < i\), then \(c(x_j) = b_m\) and \(x_j \equiv c(x_j) \pmod{f(k - m)!}\) by (4.3). Therefore \(x_j/v_j \equiv c(x_j/v_j \pmod{f(k - i)!}, and \(a_i \equiv b_i \pmod{f(k - i)!}\).

If, on the other hand, there is no \((i - 1)\)-vector \(s\) such that \(\bar{s} = a_i\), then \(\bar{s}_m < a_i\) and \(s_m\) is a minor \((i - 1)\)-vector. Let \(\varepsilon = \bar{c}(s_M) - \bar{c}(s_{m})\). We will show that \(\varepsilon > f(k - i)!\). The lemma then follows immediately.

Suppose \(\varepsilon \leq f(k - i)!\). Since \(s_M\) is also a minor \((i - 1)\)-vector, \(\beta + \varepsilon = \sum_{j=1}^{p} \gamma_j c(y_j) - \sum_{j=1}^{n} \alpha_j c(x_j) + \delta = u/v, \) where, by 4.1(i)', (ii), (iii) and (3.4),

\[|v| \leq 2f(k - i)^2 h(k - i)^2 f(k - i + 1)^2(k - i + 3) = h(k - i + 1).\]

Also, by 4.1(i)' and (3.3),

\[|\beta + \varepsilon| \leq 2f(k - i)^2 g(k - i) + f(k - i)! = g(k - i + 1)\]

Therefore \(s = (x_1, \ldots, x_n, \alpha_1, \ldots, \alpha_n, \beta + \varepsilon)\) is an \((i - 1)\)-vector such that \(c(s) = \bar{c}(s_{M})\). By the induction hypothesis for (4.6), \(\bar{s} = \bar{s}_M\), so \(\bar{s}_M - \bar{s}_m = \varepsilon\).

Letting \(\zeta = a_i - \bar{s}_m, \zeta < \varepsilon \leq f(k - i)!, and by similar reasoning \(t = (x_1, \ldots, x_n, \alpha_1, \ldots, \alpha_n, \beta + \zeta)\) is an \((i - 1)\)-vector. But \(i = a_i,\) a contradiction. Therefore \(\varepsilon > f(k - i)!\). Q.E.D.

Therefore player II can choose \(b_i\) so that (4.3) is satisfied. If we extend \(c\) to \(c(a_i) = b_i\), then \(c\) is an \(i\)-correspondence, and (4.4) is satisfied.

4.13 Lemma. \(a_i \in Q_0\) if and only if \(b_i \in Q_1\).

Proof. If \(a_i \in Q_0\), then \(b_i \in Q_1\), as shown in 4.2.

If \(a_i \notin Q_0\) but \(\bar{s}_m = a_i\), then \(b_i = \bar{c}(s_{m})\). Now if \(b_i \in Q_1\), there is some \(u \in Q_0\) such that we can extend \(c\) to an \((i - 1)\)-correspondence \(c_1\) such that \(c_1(u) = b_i\). This follows from Lemma 4.10. By (4.6), \(c_1(s_m) = c_1(u)\) implies \(s_m = u\). But then \(a_i = u \in Q_0\). Therefore \(b_i \notin Q_1\).

If \(\bar{s}_m < a_i\), then \(a_i \notin Q_0\) (otherwise \(s = (a_i, 0)\) would be an \((i - 1)\)-vector such that \(\bar{s} = \bar{a}_i\)). By Lemma 4.11, \(b_i \notin Q_1\). Q.E.D.

Therefore condition (4.7) is satisfied. It remains only to prove (4.6) holds.

4.14 Lemma. Given that player II has chosen \(b_i\) as above, (4.6) holds.

Proof. Let \(s_1, s_2,\) and \(c\) be as in (4.6), and assume \(s_1 \leq s_2,\) where \(s_1 = (x_1, \ldots, x_n, \alpha_1, \ldots, \alpha_n, \beta)\) and \(s_2 = (y_1, \ldots, y_p, \gamma_1, \ldots, \gamma_p, \delta)\). We may assume \(a_i = x_1 = y_1\) (if \(a_i\) is not a term of \(s_1\) we take \(\alpha_1 = 0,\) and similarly for \(s_2)\). Let \(t_1 = (x_2, \ldots, x_n, \alpha_2, \ldots, \alpha_n, \beta)\) and \(t_2 = (y_2, \ldots, y_p, \gamma_2, \ldots, \gamma_p, \delta)\). If \(a_1 = \gamma_1,\) then since \(t_1\) and \(t_2\) are \((i - 1)\)-vectors and \(i_1 \leq i_2,\) we get \(c(t_1) \leq c(t_2)\) by the induction hypothesis, and therefore \(c(s_1) \leq c(s_2)\).

If \(a_1 \neq \gamma_1,\) say \(a_1 < \gamma_1,\) then \((i_1 - i_2)/(\gamma_1 - a_1) \leq a_i\). Now \((i_1 - i_2)/(\gamma_1 - a_1) = i,\) where \(i\) is a minor \((i - 1)\)-vector. We will prove this by showing that 4.1(i)' through (iv) hold for \(i.\)
\((i') (\beta - \delta)/(\gamma_1 - \alpha_1) = u/v, \) where \(u, v \in \mathbb{Z}, |u/v| \leq 2f(k - i)^2 g(k - i),\) and \(|v| \leq 2f(k - i)^2 h(k - i)^2.\)

(ii) \(n + p \leq 2^{k-i+2}.\)

(iii) Let \(x_j\) be a term of \(t_1\) but not of \(t_2.\) Then \(\alpha_j = u/v, \) where \(u, v \in \mathbb{Z} \) and \(|u|, |v| \leq f(k - i) < f(k - i + 1).\) A similar conclusion applies to any \(y_j\) which is a term of \(t_2\) but not of \(t_1.\) Now let \(x_j = y_m.\) Then \((\alpha_j - \gamma_m)/(\gamma_1 - \alpha_1) = u/v, \) where \(u, v \in \mathbb{Z} \) and \(|u|, |v| \leq 2f(k - i)^4 f(k - i + 1)\) by (3.2).

(iv) is obvious.

By definition (see 4.2), \(i \leq \delta_m.\) The domain of \(c\) includes all the terms of \(\delta_m,\) since they are in \(C_\alpha.\) Therefore by the induction hypothesis \(c(i) \leq c(s_m) \leq b_i,\) and \(c(s_1) \leq c(s_2).\)

The proof that \(c(s_1) \leq c(s_2)\) implies \(\delta_1 \leq \delta_2\) is similar. We get a minor \((i - 1)-\)

vector \(i' \leq b_i.\) Then by Lemma 4.11, \(i' \leq c(s_m),\) and \(\delta_1 \leq \delta_2.\) Q.E.D.

This concludes the proof of Theorem 3.7.

4.15 Proof of 3.9. Let \(c_1, \ldots, c_m, d_1, \ldots, d_n \in A_k\) such that \(c_i \leq l_j \leq d_i\)

for \(1 \leq i, j \leq n.\) We will show that \(\langle \omega, +, c_1, \ldots, c_n \rangle\) and \(\langle \omega, +, d_1, \ldots, d_n \rangle\) cannot be distinguished by any sentence of depth \(k.\) Let \(Q_0 = \{c_1, \ldots, c_n\}\) and \(Q_1 = \{d_1, \ldots, d_n\}.\) We define \(B_{ji} = Q_j\) for \(j = 0, 1\) and \(i = 1, \ldots, k\) as in 4.1, and we use the same strategy as in 4.2. By (4.4) and (4.6), \(a_1, \ldots, a_k\) in \(\langle \omega, +, c_1, \ldots, c_n \rangle\) is isomorphic to \(b_1, \ldots, b_k\) in \(\langle \omega, +, d_1, \ldots, d_n \rangle,\) and player II wins. Q.E.D.

4.16 Proof of 3.10. The proof is very similar to the proof of Theorem 3.7. The only difference is that \(n\) is now a distinguished point, much like \(\max(Q_i), j = 0, 1,\) in the proof of 3.7. Thus, if \(|Q_0| = |Q_1|,\) we put \(B_{ij} = Q_j \cup \{n\}\) for \(i = 1, \ldots, k,\) and if \(|Q_0|, |Q_1| > d(k),\) we put \(B_{j0} = \{\min(Q_j), \max(Q_j), n\}\) for \(j = 0, 1\), and in 4.1(iv), \(n\) can be a term of an \(i\)-vector. Also, if player I chooses \(n\) in one of the structures, then player II must respond by choosing \(n\) in the other structure. The rest of the proof is unchanged. Q.E.D.

4.17 Proof of 3.11. Let \(\sigma\) be a relational sentence of depth \(k\) that defines \(E_n\) in \(\langle \omega, +, n \rangle,\) and let \(m = d(k) + 3.\) Then for any set \(A_{hm}\) as in 3.10, if \(n \geq p_m\) there are \(Q_0, Q_1 \subseteq A_{hm}\) such that \(|Q_0| = d(k) + 1\) and \(|Q_1| = d(k) + 2.\) But then \(\langle \omega, +, n, Q_0 \rangle\) and \(\langle \omega, +, n, Q_1 \rangle\) are indistinguishable by \(\sigma.\)

Therefore \(p_m > n.\) Taking \(p_m\) as defined in (3.6),

\[f(k)! \sum_{j=0}^{d(k)+2} (2^{k+3} f(k)^3 j) > n,\]

\[2^{(4^k-1)/3} > n,\]

\[(2^{(4^k-1)/3})(4^k - 1)/3 + (k + 2 + 4^k)2^{k^2} > \log_2 n,\]

\[2^k > \log_2 n\]

for sufficiently large \(n.\) Therefore \(k > \frac{1}{2} \log_2 \log_2 \log_2 n.\) Q.E.D.

§5. Related problems. Our results show that the first-order language of \(+\) is very limited in the sets of relations that it can define. On the other hand, the language
of $+$ and $\cdot$ is sufficiently powerful to define any recursively enumerable set of finite $q$-ary relations over $\omega$. Also, every such set is representable in the form
\[ \{ Q \subseteq \omega : (\exists P \subseteq \omega) \langle \omega, +, P, Q \rangle \models \sigma \} \]
where $\sigma$ is a first-order sentence. The same applies to $(\forall P \subseteq \omega)$. This is because there is a sentence about $\langle \omega, +, P \rangle$ which secures $P = \{ n^2 : n \in \omega \} \cup \{ n^2 - n : n \in \omega \}$ and multiplication is first-order definable in $\langle \omega, +, P \rangle$ (see [13]). Between these two extremes, however, there are languages about which comparatively little is known. Among these are certain languages that can define those sets of finite sequences of 0's and 1's which are recognizable by time or space bounded Turing machines ([6], [10], [11], [13]). The following is representative of these results.

Let $\{0, 1\}^* = \bigcup_{n \in \omega} \#2^n$ be the set of all finite sequences of 0's and 1's. We identify each $x \in \#2^n$ with the relation $R_x = \{ i < n : x(i) = 1 \}$. We say that $\sigma$ is an existential second-order sentence of degree $d$ if $\sigma$ is of the form $\exists Q_1, \ldots, \exists Q_{\tau} \tau$, where $\tau$ is a first-order sentence, each $Q_i$ is a $q_i$-ary relation symbol, and $q_i \leq d$. If $d = 1$, we say that $\sigma$ is monadic.

5.1 Theorem (Lynch [13]). Let $X \subseteq \{0, 1\}^*$, $T$ be a nondeterministic Turing machine, and $f$ be a function in $\tau$ such that for every $n \in \omega$ and $x \in \omega^2, x \in X$ if and only if $T$ accepts $x$ in time $f(n)$. Then there is a monadic existential second-order sentence $\varphi$ such that for all $n \in \omega$ and $x \in \omega^2, x \in X$ if and only if $\langle f(n), +, R_x \rangle \models \varphi$.

This gives a refinement of the well-known second-order characterization of NP due to N. Jones and A. Selman [11] and R. Fagin [6]. Thus, if one could show that a given $X \subseteq \{0, 1\}^*$ is not definable by an existential second-order sentence of degree $d$, then it would immediately follow that $X$ is not recognizable in time $n^d$. A natural first step would be to characterize sets of relations definable by monadic existential second-order sentences in $+$, i.e. those recognizable in linear time. There are results on the definability of sets of relations by monadic second-order sentences in the language of successor ([2], [5], [7], [17]). Of course, comparable results for the language of $+$ would be much more difficult.

A related, but possibly more tractable, problem is to characterize sets of relations in $n$ which are definable in a primal algebra on $n \langle n, f_1, \ldots, f_c \rangle$ is primal if every function $g : \omega^n \rightarrow n$ is obtainable by composing $f_1, \ldots, f_c$. The Galois fields of prime order are primal.) For example, let $E_n$ be as in 3.11. Is there a natural sequence of primal algebras $\langle n, f_1^{(n)}, \ldots, f_c^{(n)} \rangle$, where $c$ is fixed, and a first-order sentence $\sigma$ such that for each $n \in \omega, E_n = \{ Q \subseteq n : \langle n, f_1^{(n)}, \ldots, f_c^{(n)}, Q \rangle \models \sigma \}$? A similar question may be asked of $C_n = \{ Q \subseteq \omega : Q$ is connected$\}$.

An alternative approach to characterizing definable sets of relations is to study their size. Results in [1], [8], [9], [12], [18] show that for certain structures $\mathfrak{A}$ and certain measures and topologies, any set of relations defined by a first-order sentence in $\mathfrak{A}$ has measure 0 or 1, and is meager or comeager. (See also [14] for related results about sentences in $L_{\omega_1^{\omega}}$.) For $\mathfrak{A} = \langle \mathbb{Z}, +, x + 1 \rangle$, it was shown in [12] that for every first-order sentence $\sigma$ there exists a partition of the space of relations into two clopen sets $P_1$ and $P_2$ such that the subset of $P_1$ which satisfies $\sigma$ is comeager in $P_1$ and its measure equals that of $P_1$ and the subset of $P_2$ which satisfies $\sigma$ is meager and of measure 0. Analogous results hold for the set of finite structures $\langle n, +, (\text{mod } n), x + 1 (\text{mod } n) \rangle, n \in \omega$. Letting $\mu(\sigma, n)$ be the probability that
\langle n, + (\text{mod } n), x + 1 (\text{mod } n), Q \rangle \models \sigma \text{ for a randomly selected } Q \subseteq \mathbb{N}, \text{ it was shown in [12] that for every } \sigma, \text{ there is an } a \in \omega \text{ such that for all } b < a, \lim_{n \to \infty} \mu(\sigma, an + b) \text{ exists. Central to the proofs in [12] was a strategy for player II in the Ehrenfeucht game played on structures of the form } \langle \mathbb{Z}, +, x + 1, Q \rangle \text{ and } \langle n, + (\text{mod } n), x + 1 (\text{mod } n), Q \rangle. \text{ However, the same methods do not apply to } \langle \omega, + \rangle \text{ and } \langle n, + \rangle, \text{ or even to } \langle \omega, \leq \rangle \text{ and } \langle n, \leq \rangle. \text{ M. Benda [1] proved a 0-1 law for sets of unary relations definable in } \langle \mathbb{Z}, \leq \rangle, \text{ and Ehrenfeucht (in [12]) proved a limit law for sets of unary relations definable in } \langle n, \leq \rangle, \text{ but these results do not extend to sets of binary relations. Thus if more general theorems that apply to these structures could be proven, it would be a significant extension of the known probabilistic results in model theory.}

REFERENCES