

MONADIC GENERALIZED SPECTRA

by RONALD FAGIN in Yorktown Heights, New York (U.S.A.)¹⁾

1. Introduction

Let \mathcal{A} be the class of finite models of a second-order existential sentence $\exists P_1 \dots \exists P_m \sigma$, where σ is an arbitrary first-order sentence (with equality). Thus, \mathcal{A} is a PC class in the sense of TARSKI [8], where we restrict our attention to the class of finite structures. If P_1, \dots, P_m are the only nonlogical symbols appearing in σ , then \mathcal{A} can be identified with the set of cardinalities of finite models of σ . H. SCHOLZ [7] called this set the *spectrum* of σ . Hence, in the general case, we call \mathcal{A} a *generalized spectrum*. If P_1, \dots, P_m are each unary predicate symbols, then we call \mathcal{A} a *monadic generalized spectrum*. In this paper, we show, by using FRAÏSSÉ-type games, that the class of monadic generalized spectra is not closed under complement.

If \mathcal{S} is a *similarity type*, that is, a finite set of predicate and constant symbols, then by an \mathcal{S} -*structure*, we mean a relational structure appropriate for \mathcal{S} . We will show that the class of non-connected, finite $\{P\}$ -structures (where P is a binary predicate symbol) is a monadic generalized spectrum, but that the class of connected, finite $\{P\}$ -structures is not (although the latter class is a generalized spectrum with just one existentialized binary predicate symbol, as we will see).

Assume throughout this paper that P is a binary predicate symbol and that U_1, U_2, \dots are unary predicate symbols. Define a *cycle (of length n)* to be a $\{P\}$ -structure $\mathfrak{A} = \langle A; Q \rangle$, where for some n distinct elements a_1, \dots, a_n ,

$$A = \{a_1, \dots, a_n\}, \quad Q = \{\langle a_i, a_{i+1} \rangle : 1 \leq i < n\} \cup \{\langle a_n, a_1 \rangle\}.$$

Write $\text{card}(\mathfrak{A}) = n$. If $\mathfrak{A} = \langle A; Q \rangle$ and $\mathfrak{B} = \langle B; R \rangle$ are cycles, and $A \cap B = \emptyset$, then by the *cardinal sum* $\mathfrak{A} \oplus \mathfrak{B}$, we mean the $\{P\}$ -structure $\langle A \cup B; Q \cup R \rangle$.

We will show that if τ is $\exists U_1 \dots \exists U_d \sigma$, where σ is a first-order $\{P, U_1, \dots, U_d\}$ -sentence (that is, its nonlogical symbols are a subset of $\{P, U_1, \dots, U_d\}$), then there is a constant N such that for each cycle \mathfrak{A} with $\mathfrak{A} \models \tau$ and $\text{card}(\mathfrak{A}) \geq N$, there is a cycle \mathfrak{B} such that $\mathfrak{A} \oplus \mathfrak{B} \models \tau$. It easily follows that the class of connected, finite $\{P\}$ -structures is not a monadic generalized spectrum. This result is related to monadic second-order decidability results in BÜCHI [2] and RABIN [6], but it does not seem to be directly derivable from them. In any case, this result can be derived very directly by the use of FRAÏSSÉ-type games [5], which is the approach we will use.

G. ASSER [1] posed the question of whether the complement of every spectrum is a spectrum. We remark that the author showed in [3] and [4] that there is a particular monadic generalized spectrum \mathcal{A} (namely, the class of all finite models of $\exists U \forall x \exists! y (Pxy \wedge Uy)$, where U is unary, P is binary, and $\exists! y$ is read "There is exactly

¹⁾ This paper is based on a part of the author's doctoral dissertation [3] in the Department of Mathematics at the University of California, Berkeley. Part of this work was carried out while the author was a National Science Foundation Graduate Fellow; also, part of this work was supported by NSF Grant No. GP-24532.

The author is grateful to ROBERT VAUGHET and WILLIAM CRAIG for useful suggestions which improved readability.

one y'') such that the complement of every generalized spectrum is a generalized spectrum (and thus the complement of every spectrum is a spectrum) iff the complement $\tilde{\mathcal{A}}$ of \mathcal{A} is a generalized spectrum.

2. Definitions

Denote the set of *natural numbers* $\{0, 1, 2, \dots\}$ by \mathbf{N} .

If \mathcal{S} is a similarity type and \mathfrak{A} is an \mathcal{S} -structure (both defined earlier), then we denote the universe of \mathfrak{A} by $|\mathfrak{A}|$, and the interpretation (in \mathfrak{A}) of S in \mathcal{S} by $S^{\mathfrak{A}}$.

If \mathfrak{A} and \mathfrak{B} are isomorphic \mathcal{S} -structures via the isomorphism g , then we write $g: \mathfrak{A} \cong \mathfrak{B}$.

Assume that \mathcal{S} and \mathcal{T} are disjoint similarity types, that \mathfrak{A} is an $\mathcal{S} \cup \mathcal{T}$ -structure, and that \mathfrak{B} is an \mathcal{S} -structure. Then \mathfrak{A} is an *expansion* of \mathfrak{B} (to $\mathcal{S} \cup \mathcal{T}$), written $\mathfrak{B} = \mathfrak{A} \upharpoonright \mathcal{S}$, if $|\mathfrak{A}| = |\mathfrak{B}|$, and $S^{\mathfrak{A}} = S^{\mathfrak{B}}$ for each S in \mathcal{S} .

Assume that \mathfrak{A} is an \mathcal{S} -structure, and that $a \in |\mathfrak{A}|$. We denote by (\mathfrak{A}, a) the $\mathcal{S} \cup \{c\}$ -structure \mathfrak{B} , such that $\mathfrak{B} \upharpoonright \mathcal{S} = \mathfrak{A}$ and $c^{\mathfrak{B}} = a$, where c is a new constant symbol, chosen by some fixed rule.

If \mathfrak{A} is an \mathcal{S} -structure and $B \subseteq |\mathfrak{A}|$, then by $\mathfrak{A} \upharpoonright B$ we mean the substructure of \mathfrak{A} with universe B .

If φ is a formula with distinct free individual variables x_1, \dots, x_k , and if $a_1, \dots, a_k \in |\mathfrak{A}|$, then by $\mathfrak{A} \models \varphi_{a_1 \dots a_k}^{x_1 \dots x_k}$, we mean that φ is satisfied in \mathfrak{A} when x_i is interpreted by a_i ($1 \leq i \leq k$).

An *atomic \mathcal{S} -formula* is a formula $t_1 = t_2$ or $S t_1 \dots t_k$, where each t_i is a constant symbol in \mathcal{S} or an individual variable, and where S is a k -ary predicate symbol in \mathcal{S} . A *negation-atomic \mathcal{S} -formula* is the negation of an atomic \mathcal{S} -formula.

A first-order formula φ is in *prenex normal form* if it is of the form $Q_1 x_1 \dots Q_m x_m \psi$, where each Q_i is \forall or \exists , each x_i is an individual variable, and ψ is quantifier-free. We say that φ *starts with m quantifiers*.

If φ is a first-order formula, then by $\exists \mathcal{S} \varphi$, we mean the (existential second-order) formula $\exists S_1 \dots \exists S_m \varphi$, where $\mathcal{S} = \{S_1, \dots, S_m\}$; similarly for $\forall \mathcal{S} \varphi$. If Γ is a set of formulas, then by $\bigwedge \{\varphi : \varphi \in \Gamma\}$, we mean the conjunction of the formulas in Γ ; similarly for $\bigvee \{\varphi : \varphi \in \Gamma\}$.

3. Fraïssé games

In this section, we will describe some games, of the type first introduced by R. FRAÏSSÉ.

Let \mathcal{S} be a similarity type, let \mathfrak{A} and \mathfrak{B} be \mathcal{S} -structures with $|\mathfrak{A}| \cap |\mathfrak{B}| = \emptyset$, and let r be a natural number. Then we can informally describe a game as follows: Player I moves first, and picks a point in either $|\mathfrak{A}|$ or $|\mathfrak{B}|$. Then player II picks a point in (the universe of) the opposite structure. Let a_1 be the point picked in $|\mathfrak{A}|$, and b_1 the point picked in $|\mathfrak{B}|$. On player I's second move, he again picks a point in either $|\mathfrak{A}|$ or $|\mathfrak{B}|$, and player II then picks a point in the opposite structure. Let a_2 be the point picked in $|\mathfrak{A}|$ on either player I's or player II's second move, and b_2 the point picked in $|\mathfrak{B}|$. Continue until player I and player II have each taken r moves (i.e., until r rounds of the game have been played). Let $\{c_i : 1 \leq i \leq k\}$ be the set of constant symbols in \mathcal{S} ($k = 0$ is possible). Let $a_{r+i}(b_{r+i})$ be $c_i^{\mathfrak{A}}(c_i^{\mathfrak{B}})$, $1 \leq i \leq k$. Then player II wins iff the following two conditions hold:

1. $\{(a_i, b_i): 1 \leq i \leq r+k\}$ is a one-one function, say g . That is, $a_i = a_j$ iff $b_i = b_j$ ($1 \leq i \leq r+k, 1 \leq j \leq r+k$).
2. $g: \mathfrak{A} \setminus \{a_1, \dots, a_{r+k}\} \cong \mathfrak{B} \setminus \{b_1, \dots, b_{r+k}\}$.

We will now inductively define a notion $\mathfrak{A} \sim_r \mathfrak{B}$, which corresponds to the intuitive notion of player II having a winning strategy in the game just informally described. We say $\mathfrak{A} \sim_0 \mathfrak{B}$ if for every quantifier-free \mathcal{S} -sentence σ , we have $\mathfrak{A} \models \sigma$ iff $\mathfrak{B} \models \sigma$. (If \mathcal{S} contains no constant symbols, then there are no quantifier-free \mathcal{S} -sentences.) For each natural number r , we say $\mathfrak{A} \sim_{r+1} \mathfrak{B}$ if

1. For each a in $|\mathfrak{A}|$ there is b in $|\mathfrak{B}|$ such that $(\mathfrak{A}, a) \sim_r (\mathfrak{B}, b)$.
2. For each b in $|\mathfrak{B}|$ there is a in $|\mathfrak{A}|$ such that $(\mathfrak{A}, a) \sim_r (\mathfrak{B}, b)$.

It is clear that \sim_r is an equivalence relation, for each r .

In our proofs, we may talk of players I and II, of player I's first move, and so on. It will be clear how to make the arguments formal.

We will now consider another game. Let \mathcal{S} be as before, and let \mathcal{T} be a finite set of predicate symbols with $\mathcal{S} \cap \mathcal{T} = \emptyset$. Let \mathfrak{A} and \mathfrak{B} be \mathcal{S} -structures, and let r be a natural number. On player I's first move, he selects an $\mathcal{S} \cup \mathcal{T}$ -structure \mathfrak{A}' such that $\mathfrak{A}' \upharpoonright \mathcal{S} = \mathfrak{A}$. Then player II selects an $\mathcal{S} \cup \mathcal{T}$ -structure \mathfrak{B}' such that $\mathfrak{B}' \upharpoonright \mathcal{S} = \mathfrak{B}$. Player II wins iff $\mathfrak{A}' \sim_r \mathfrak{B}'$. Formally, we say $\mathfrak{A} \rightarrow_r^{\mathcal{T}} \mathfrak{B}$ if for each expansion \mathfrak{A}' of \mathfrak{A} to $\mathcal{S} \cup \mathcal{T}$, there is an expansion \mathfrak{B}' of \mathfrak{B} to $\mathcal{S} \cup \mathcal{T}$ such that $\mathfrak{A}' \sim_r \mathfrak{B}'$. If \mathcal{T} is a set of d distinct unary predicate symbols, then write $\mathfrak{A} \rightarrow_r^d \mathfrak{B}$ for $\mathfrak{A} \rightarrow_r^{\mathcal{T}} \mathfrak{B}$.

It is easy to see that $\rightarrow_r^{\mathcal{T}}$ is transitive and reflexive, but as we shall see, it is not necessarily symmetric. The corresponding symmetric notion would be $\mathfrak{A} \leftrightarrow_r^{\mathcal{T}} \mathfrak{B}$, which holds if $\mathfrak{A} \rightarrow_r^{\mathcal{T}} \mathfrak{B}$ and $\mathfrak{B} \rightarrow_r^{\mathcal{T}} \mathfrak{A}$.

We will prove the next theorem (which is essentially due to FRAÏSSÉ [5]) in more generality than we will need.

If \mathcal{T} is a finite set of unary predicate symbols, then call \mathcal{T} *monadic*. Let \mathcal{K} be a class of \mathcal{S} -structures. Following TARSKI [8], we say that a class \mathcal{A} of \mathcal{S} -structures is in $PC(\mathcal{K})$ ($PC_1(\mathcal{K})$) if $\mathcal{A} = \{\mathfrak{A} \in \mathcal{K} : \mathfrak{A} \models \exists \mathcal{T} \sigma\}$ for some (monadic) \mathcal{T} and some first-order $\mathcal{S} \cup \mathcal{T}$ -sentence σ . We are interested in the case when \mathcal{K} is the class of finite $\{P\}$ -structures and \mathcal{T} is monadic.

Theorem 1. *Assume $\mathcal{A} \subseteq \mathcal{K}$. Then $\mathcal{A} \in PC(\mathcal{K})$ ($PC_1(\mathcal{K})$) iff there is some (monadic) \mathcal{T} and some natural number r such that whenever $\mathfrak{A} \in \mathcal{A}$, $\mathfrak{B} \in \mathcal{K}$, and $\mathfrak{A} \rightarrow_r^{\mathcal{T}} \mathfrak{B}$, then $\mathfrak{B} \in \mathcal{A}$.*

Proof " \Rightarrow ". Let $\mathcal{A} = \{\mathfrak{A} \in \mathcal{K} : \mathfrak{A} \models \exists \mathcal{T} \sigma\}$. We can assume without loss of generality that \mathcal{T} contains only predicate symbols and that σ is in prenex normal form. Say σ starts with r quantifiers. Assume that $\mathfrak{A} \in \mathcal{A}$, that $\mathfrak{B} \in \mathcal{K}$, and that $\mathfrak{A} \rightarrow_r^{\mathcal{T}} \mathfrak{B}$. We will show that $\mathfrak{B} \in \mathcal{A}$. We will prove this in the special case when σ is $\forall x \exists y M$, where M is quantifier-free. The general case is very similar.

Assume that $\mathfrak{B} \notin \mathcal{A}$. Then $\mathfrak{B} \models \forall \mathcal{T} \exists x \forall y \sim M$. Find an $\mathcal{S} \cup \mathcal{T}$ -structure \mathfrak{A}' such that $\mathfrak{A}' \upharpoonright \mathcal{S} = \mathfrak{A}$ and $\mathfrak{A}' \models \forall x \exists y M$. Let \mathfrak{B}' be an arbitrary $\mathcal{S} \cup \mathcal{T}$ -structure with $\mathfrak{B}' \upharpoonright \mathcal{S} = \mathfrak{B}$. We will show that not $\mathfrak{A}' \sim_2 \mathfrak{B}'$. On player I's first move, he picks b_1 in $|\mathfrak{B}'|$ such that $\mathfrak{B}' \models \forall y \sim M \begin{bmatrix} x \\ b_1 \end{bmatrix}$. Let a_1 in $|\mathfrak{A}'|$ be player II's response. Then $\mathfrak{A}' \models \exists y M \begin{bmatrix} x \\ a_1 \end{bmatrix}$. On player I's next move, he picks a_2 in $|\mathfrak{A}'|$ such that $\mathfrak{A}' \models M \begin{bmatrix} x & y \\ a_1 & a_2 \end{bmatrix}$. Let b_2 in $|\mathfrak{B}'|$ be player II's response. Then $\mathfrak{B}' \models \sim M \begin{bmatrix} x & y \\ b_1 & b_2 \end{bmatrix}$. So player II has clearly lost.

" \Leftarrow ". For each finite set \mathcal{S}' of predicate symbols, we will define the notion of an m -type $_{r,(\mathcal{S}')}$, for $0 \leq m \leq r$, by backwards induction (from $m = r$ to $m = 0$.) An r -type $_{r,(\mathcal{S}')}$ is any formula

$$\bigwedge \left\{ \theta : \mathfrak{A} \models \theta \left[\begin{array}{c} v_1 \dots v_r \\ a_1 \dots a_r \end{array} \right] \text{ and } \theta \text{ is an atomic or negation-atomic } \mathcal{S}\text{-formula} \right\},$$

such that \mathfrak{A} is an \mathcal{S}' -structure and $a_1, \dots, a_r \in |\mathfrak{A}|$. For each set A of $(m+1)$ -types $_{r,(\mathcal{S}')}$, the following is an m -type $_{r,(\mathcal{S}')}$:

$$\bigwedge \{ \exists v_{m+1} \varphi : \varphi \in A \} \wedge \bigwedge \{ \forall v_{m+1} \sim \varphi : \varphi \notin A \}.$$

It is easily proved by induction that for any \mathcal{S}' -structure \mathfrak{A} and any a_1, \dots, a_m in $|\mathfrak{A}|$, we have $\mathfrak{A} \models \varphi \left[\begin{array}{c} v_1 \dots v_m \\ a_1 \dots a_m \end{array} \right]$ for exactly one m -type $_{r,(\mathcal{S}')}$ φ . For each m ($0 \leq m \leq r$), there is only a finite number of distinct m -types $_{r,(\mathcal{S}')}$, and each has free variables v_1, \dots, v_m .

For each \mathcal{S}' -structure \mathfrak{A} , and each natural number r , denote σ , the 0-type $_{r,(\mathcal{S}')}$ such that $\mathfrak{A} \models \sigma$, by $\sigma(\mathfrak{A}, r)$. It is easy to see that if \mathfrak{A} and \mathfrak{B} are \mathcal{S}' -structures, then $\mathfrak{A} \sim_r \mathfrak{B}$ iff $\mathfrak{B} \models \sigma(\mathfrak{A}, r)$: player II's strategy is to make sure that after the m th move, if $a_1, \dots, a_m (b_1, \dots, b_m)$ are the points that have been picked in $|\mathfrak{A}| (|\mathfrak{B}|)$, then $\mathfrak{A} \models \varphi \left[\begin{array}{c} v_1 \dots v_m \\ a_1 \dots a_m \end{array} \right]$ and $\mathfrak{B} \models \varphi \left[\begin{array}{c} v_1 \dots v_m \\ b_1 \dots b_m \end{array} \right]$ for the same m -type $_{r,(\mathcal{S}')}$ φ .

If \mathfrak{A} is an \mathcal{S} -structure, then let $\tau(\mathfrak{A}, \mathcal{F}, r)$ be the (finite) conjunction $\bigwedge \{ \exists \mathcal{F} \sigma(\mathfrak{A}', r) : \mathfrak{A}' \text{ is an } \mathcal{S} \cup \mathcal{F}\text{-structure with } \mathfrak{A}' \upharpoonright \mathcal{S} = \mathfrak{A} \}$. It is easy to see that if \mathfrak{B} is an \mathcal{S} -structure, then $\mathfrak{A} \rightarrow_r^{\mathcal{F}} \mathfrak{B}$ iff $\mathfrak{B} \models \tau(\mathfrak{A}, \mathcal{F}, r)$.

Let $\mathcal{A} \subseteq \mathcal{X}$ have the property that whenever $\mathfrak{A} \in \mathcal{A}$, $\mathfrak{B} \in \mathcal{X}$, and $\mathfrak{A} \rightarrow_r^{\mathcal{F}} \mathfrak{B}$, then $\mathfrak{B} \in \mathcal{A}$. Then

$$\mathcal{A} = \{ \mathfrak{B} \in \mathcal{X} : \mathfrak{B} \models \bigvee \{ \tau(\mathfrak{A}, \mathcal{F}, r) : \mathfrak{A} \in \mathcal{A} \} \}.$$

So $\mathcal{A} \in PC(\mathcal{X})$, because a finite conjunction or disjunction of existential second-order sentences is an existential second-order sentence. Likewise, if \mathcal{F} is monadic, then $\mathcal{A} \in PC_1(\mathcal{X})$.

4. Nonclosure under complement

In this section, we will show that for each pair d, r of natural numbers, there are structures \mathfrak{A} and \mathfrak{B} such that \mathfrak{A} is a cycle, \mathfrak{B} is the cardinal sum of two cycles, and $\mathfrak{A} \rightarrow_r^d \mathfrak{B}$. (In fact, $\mathfrak{B} = \mathfrak{A} \oplus \mathfrak{C}$ for some cycle \mathfrak{C} .) It then follows easily from Theorem 1, that the class of connected, finite $\{P\}$ -structures is not a monadic $\{P\}$ -spectrum (a $\{P\}$ -structure \mathfrak{A} is *connected* if for each a, b in $|\mathfrak{A}|$ there is a finite sequence a_1, \dots, a_n of points in $|\mathfrak{A}|$ such that $a_1 = a$, $a_n = b$, and either $P^{\mathfrak{A}} a_i a_{i+1}$ or $P^{\mathfrak{A}} a_{i+1} a_i$, for $1 \leq i < n$). However, as we will see, the class of nonconnected, finite $\{P\}$ -structures is a monadic $\{P\}$ -spectrum.

Let $\mathcal{S} = \{P, U_1, \dots, U_d\}$ as before. Let \mathfrak{A}' be an \mathcal{S} -structure, with $\mathfrak{A}' \upharpoonright \{P\}$ the cardinal sum of cycles. If $a \in |\mathfrak{A}'|$, then define the *weak marking* m on a to be the subset $m \subseteq \{U_1, \dots, U_d\}$; where $U_i \in m$ iff $U_i^{\mathfrak{A}'} a$. Assume that

1. $a_1, \dots, a_t \in |\mathfrak{A}'|$.
2. m_i is the weak marking on a_i ($1 \leq i \leq t$).
3. $P^{\mathfrak{A}'} a_i a_{i+1}$ ($1 \leq i < t$).

Then $\langle m_1, \dots, m_t \rangle$ is a *weak sequence* (of length t) in \mathfrak{A}' . A weak sequence $\langle m_1, \dots, m_t \rangle$ occurs at least n times in \mathfrak{A}' if there are at least n different t -tuples $\langle a_1, \dots, a_t \rangle$ such that the three conditions above hold.

Define $v: \mathbb{N} \rightarrow \mathbb{N}$ by

$$v(0) = 1, \quad v(r + 1) = 2v(r) + 1.$$

Let $n(r) = rv(r)$ for each r .

The next lemma is the main tool in proving our result.

Lemma 2. *Let r be a natural number, and let \mathfrak{A} and \mathfrak{C} be \mathcal{S} -structures, with \mathcal{S} as above, such that $\mathfrak{A} \uparrow \{P\}$ and $\mathfrak{C} \uparrow \{P\}$ are each cycles of length at least $v(r + 1)$. Assume that every weak sequence of length $v(r)$ in \mathfrak{C} occurs at least $n(r)$ times in \mathfrak{A} . Then $\mathfrak{A} \sim_r \mathfrak{A} \oplus \mathfrak{C}$.*

Proof. Let $\mathfrak{B} = \overline{\mathfrak{A}} \oplus \mathfrak{C}$, where $f: \mathfrak{A} \cong \overline{\mathfrak{A}}$. We write $f(a)$ as \bar{a} , for each a in $|\mathfrak{A}|$. We assume that $|\mathfrak{A}| \cap |\mathfrak{B}| = \emptyset$.

Assume that each player has made k selections of points. Then the *strong marking* $m = m(k, a)$ on a point a in $|\mathfrak{D}|$, where \mathfrak{D} is \mathfrak{A} or \mathfrak{B} , is the subset m of $\{U_1, \dots, U_a\} \cup \{1, \dots, k\}$, where $U_i \in m$ iff $U_i^a a$, and $i \in m$ iff the point a was selected by either player in the i th round. For each natural number n , denote by $S(a, n)$ ($S'(a, n)$) the $(2n + 1)$ -tuple

$$\langle m_{-n}, m_{-n+1}, \dots, m_0, \dots, m_n \rangle,$$

where for some $(2n + 1)$ -tuple $\langle a_{-n}, \dots, a_n \rangle$ of members of $|\mathfrak{D}|$, we have $P^{\mathfrak{D}} a_i a_{i+1}$ for $-n \leq i < n$, with $a_0 = a$, and where m_i is the weak (strong) marking on a_i . Call $S'(a, n)$ *clean* if $S'(a, n) = S(a, n)$. Let $C(a, n) = \{a_{-n}, \dots, a_n\}$.

We will show that player II has a strategy such that if there are p moves remaining for each player (that is, $r - p$ rounds have been played), and if the point picked in $|\mathfrak{A}|$ ($|\mathfrak{B}|$) on the i th round was a_i (b_i), for $1 \leq i \leq r - p$, then

1. $\{(a_i, b_i) : 1 \leq i \leq r - p\}$ is a one-one function, say g .
2. $g: \mathfrak{A} \uparrow \{a_1, \dots, a_{r-p}\} \cong \mathfrak{B} \uparrow \{b_1, \dots, b_{r-p}\}$.
3. $S'(a_i, v(p)) = S'(b_i, v(p))$, $1 \leq i \leq r - p$.
4. If $b_i \neq \bar{a}_i$, then $S(a_i, v(p))$ occurs at least $n(r)$ times in \mathfrak{A} .

When $p = r$, these are trivially true. Assume that these are true for $p = s + 1$; we will show that player II can play so that they are true when $p = s$.

On his $(r - s)$ th move, player I can pick a point in either $|\mathfrak{A}|$ or $|\mathfrak{B}|$. Assume first that he picks $a = a_{r-s}$ in $|\mathfrak{A}|$. There are now three cases.

Case I. $S'(a, v(s))$ is not clean. Intuitively speaking, player I has selected a point a which is near another point a' that has already been selected. If b' is the point in $|\mathfrak{B}|$ which was selected in the same round as a' , then player II's strategy is to pick a point b in $|\mathfrak{B}|$ such that b relates to b' (with respect to distance and direction) as a relates to a' .

Formally, we know that $a_i \in C(a, v(s))$, for some i with $1 \leq i \leq r - s - 1$. By assumption, $S'(a_i, v(s + 1)) = S'(b_i, v(s + 1))$. For some $j \leq v(s)$, there are x_1, \dots, x_j in $|\mathfrak{A}|$, with $P^{\mathfrak{A}} x_k x_{k+1}$, for $1 \leq k < j$, such that either $x_1 = a$ and $x_j = a_i$, or $x_1 = a_i$ and $x_j = a$. If the former is true, then find y_1, \dots, y_j in $|\mathfrak{B}|$, with $P^{\mathfrak{B}} y_k y_{k+1}$, for $1 \leq k < j$, and with $y_j = b_i$. Set $b_{r-s} = y_1$. The latter case is similar, with $y_1 = b_i$ and $y_j = b_{r-s}$.

Now $C(a, v(s)) \subseteq C(a_i, v(s+1))$, and so it is easy to check that the four conditions hold with $p = s$. For example, $S'(a_{r-s}, v(s)) = S'(b_{r-s}, v(s))$ since $S'(a_i, v(s+1)) = S'(b_i, v(s+1))$.

Case 2. $S'(a, v(s))$ is clean and $S'(\bar{a}, v(s))$ is clean. Player I has selected a point a which is not near any point that has been selected before; also, \bar{a} is not near any point which has been selected before. Let $b_{r-s} = \bar{a}$: that is, player II selects \bar{a} . The four conditions hold for $p = s$.

Case 3. $S'(a, v(s))$ is clean and $S'(\bar{a}, v(s))$ is not clean. Player I has selected a point a which is not near any point that has been selected before, but \bar{a} is near a point that has been selected before. So player II cannot pick \bar{a} ; he must instead pick a point in $|\mathfrak{C}|$ whose immediate neighborhood looks like the immediate neighborhood of a .

We know that $b_i \in C(\bar{a}, v(s))$ for some i with $1 \leq i \leq r-s-1$. Then $b_i \neq \bar{a}_i$, because if $b_i = \bar{a}_i$, then $a_i \in C(a, v(s))$, and so $S'(a, v(s))$ would not be clean. By conditions 3 and 4, we therefore know that $S(a_i, v(s+1)) = S(b_i, v(s+1))$ occurs at least $n(r)$ times in \mathfrak{A} . Now $C(\bar{a}, v(s)) \subseteq C(b_i, v(s+1))$, and so $S(a, v(s)) = S(\bar{a}, v(s))$ occurs at least $n(r)$ times in \mathfrak{A} (and \mathfrak{A}). Now $\bigcup_{k=1}^{r-s-1} C(b_k, v(s))$ contains at most $(r-s-1)v(s+1) \leq (r-1)v(r) < n(r)$ points. So we can find d in $|\mathfrak{A}|$ such that $S(d, v(s)) = S(a, v(s))$, and with d not in $\bigcup_{k=1}^{r-s-1} C(b_k, v(s))$. Hence $S'(d, v(s))$ is clean. Let $b_{r-s} = d$. The four conditions now hold for $p = s$.

Now say player I picks $b = b_{r-s}$ in $|\mathfrak{B}|$. There are two cases: $b \in |\mathfrak{A}|$ or $b \in |\mathfrak{C}|$.

Case 1'. $b \in |\mathfrak{A}|$. For some a , we have $b = \bar{a}$. There are three subcases.

Case 1'a. $S'(\bar{a}, v(s))$ is not clean. This is dealt with exactly like Case 1.

Case 1'b. $S'(\bar{a}, v(s))$ and $S'(a, v(s))$ are both clean. Let $a_{r-s} = a$.

Case 1'c. $S'(\bar{a}, v(s))$ is clean, and $S'(a, v(s))$ is not clean. Then as in Case 3, we can find d in $|\mathfrak{A}|$ such that $S(d, v(s)) = S(\bar{a}, v(s))$, such that $S'(d, v(s))$ is clean, and such that $S(d, v(s))$ occurs at least $n(r)$ times in \mathfrak{A} . Let $a_{r-s} = d$.

Case 2'. $b \in |\mathfrak{C}|$. There are two subcases.

Case 2'a. $S'(b, v(s))$ is not clean. This is dealt with like Case 1 (and Case 1'a).

Case 2'b. $S'(b, v(s))$ is clean. Now each weak sequence of length $v(s)$ in \mathfrak{C} occurs at least $n(r)$ times in \mathfrak{A} . As in Case 3, we can find d in $|\mathfrak{A}|$ such that $S(d, v(s)) = S(b, v(s))$, with $S(d, v(s))$ clean. Of course, $S(d, v(s))$ occurs at least $n(r)$ times in \mathfrak{A} . Let $a_{r-s} = d$.

The induction is complete. When $p = 0$, we see from conditions 1 and 2 that player II wins.

Let $p = \langle p_1, \dots, p_m \rangle$ and $q = \langle q_1, \dots, q_n \rangle$ be sequences. Then by $p \frown q$, we mean the concatenated sequence $\langle p_1, \dots, p_m, q_1, \dots, q_n \rangle$. We call p a *consecutive subsequence* of q if for some j , $0 \leq j \leq n-m$, we have $p_i = q_{i+j}$, $1 \leq i \leq m$. The *length* of the sequence $p = \langle p_1, \dots, p_m \rangle$ is m .

Define $f: \mathbb{N} \times \mathbb{N} \rightarrow \mathbb{N}$ by $f(d, r) = e^{v(r)} e^{e^{v(r)}} n(r)$, where $e = 2^d$.

Theorem 3. *Let d and r be natural numbers, and let \mathfrak{A} be a cycle of length at least $f(d, r)$. Then there is a positive integer a such that for each positive integer k and each cycle \mathfrak{C} of length ka , $\mathfrak{A} \rightarrow_r^d \mathfrak{A} \oplus \mathfrak{C}$.*

Proof. Let $\mathcal{S} = \{P, U_1, \dots, U_a\}$. Assume that \mathfrak{A} is a cycle of length at least $f(d, r)$, and that \mathfrak{A}' is an \mathcal{S} -structure, with $\mathfrak{A}' \upharpoonright \{P\} = \mathfrak{A}$. We will find some weak sequence

$s = s(\mathfrak{A}')$ of length at least $v(r)$, with the property that if s' is any consecutive subsequence of length $v(r)$ of $s^\wedge s$, then s' occurs at least $n(r)$ times in \mathfrak{A}' . Then we will show that this is sufficient to prove the theorem.

Let $e = 2^d$. We will first construct a certain weak sequence $u = \langle u_1, \dots, u_{e^{v(r)}+v(r)} \rangle$ of length $e^{v(r)} + v(r)$. The number of possible weak sequences of length $v(r)$ is $e^{v(r)}$. Since $f(d, r) = e^{v(r)}e^{e^{v(r)}}n(r)$, some weak sequence $t = \langle t_1, \dots, t_{v(r)} \rangle$ of length $v(r)$ occurs at least $e^{e^{v(r)}}n(r)$ times in \mathfrak{A}' . Let $u_i = t_i$, $1 \leq i \leq v(r)$. Let t' be the weak sequence $\langle t_2, \dots, t_{v(r)} \rangle$ of length $v(r) - 1$. Since t' occurs at least $e^{e^{v(r)}}n(r)$ times in \mathfrak{A}' , we know that for some weak marking b , the weak sequence $t' \wedge \langle b \rangle$ occurs at least $e^{e^{v(r)}-1}n(r)$ times in \mathfrak{A}' . Let $u_{v(r)+1} = b$. Now we can find c such that $\langle u_3, u_4, \dots, u_{v(r)+1} \rangle \wedge \langle c \rangle$ occurs at least $e^{e^{v(r)}-2}n(r)$ times in \mathfrak{A}' . Let $u_{v(r)+2} = c$. Continue this process $e^{v(r)}$ times. Then each consecutive subsequence of length $v(r)$ of u occurs at least $n(r)$ times in \mathfrak{A}' .

For each i , $1 \leq i \leq e^{v(r)} + 1$, let $q_i = \langle u_i, \dots, u_{i+v(r)-1} \rangle$. There are only $e^{v(r)}$ possible different q_i 's, and so $q_i = q_j$ for some $i < j$. There are now two cases.

Case 1. $i + v(r) - 1 < j$. Let $s = \langle u_i, \dots, u_{j-1} \rangle$. If s' is any consecutive subsequence of length $v(r)$ of $s^\wedge s$, then s' is a consecutive subsequence of length $v(r)$ of u ; hence, s' occurs at least $n(r)$ times in \mathfrak{A}' by construction.

Case 2. $i + v(r) - 1 \geq j$. Let $t = \langle u_i, \dots, u_{j-1} \rangle$, and let $s = t \wedge t \wedge \dots \wedge t$, with the concatenation taken just enough times that the length of s is at least $v(r)$. Once again, each consecutive subsequence s' of length $v(r)$ of $s^\wedge s$ is a consecutive subsequence of length $v(r)$ of u , and so it occurs at least $n(r)$ times in \mathfrak{A}' .

So given \mathfrak{A}' , we have found $s = s(\mathfrak{A}')$ with the desired property.

Let $s = \langle m_1, \dots, m_b \rangle$ be an arbitrary sequence of subsets of $\{U_1, \dots, U_d\}$, and let k be a positive integer. We will now define an \mathcal{S} -structure $\mathfrak{E} = \mathfrak{E}(s, k)$ which corresponds to the intuitive picture of Figure 1, where s is written down k times.

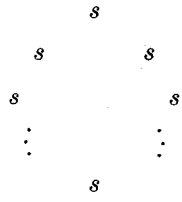


Figure 1

Let $|\mathfrak{E}| = \{1, 2, \dots, kb\}$. Let $P^{\mathfrak{E}} = \{\langle i, i + 1 \rangle : 1 \leq i < kb\} \cup \{\langle kb, 1 \rangle\}$. Let $U_i^{\mathfrak{E}}j$ hold iff $U_i \in m_e$, where $e \equiv j \pmod b$ and $1 \leq e \leq b$. By Lemma 2, for each expansion \mathfrak{A}' of \mathfrak{A} to \mathcal{S} , we have $\mathfrak{A}' \sim_r \mathfrak{A}' \oplus \mathfrak{E}(s(\mathfrak{A}'), k)$.

Let a be the least common multiple of the cardinality of each $s(\mathfrak{A}')$, over all expansions \mathfrak{A}' of \mathfrak{A} to \mathcal{S} . Then a is the number called for in the statement of the theorem. For, let \mathfrak{C} be a cycle of length ka for some positive integer k . We will show that $\mathfrak{A} \xrightarrow{d} \mathfrak{A} \oplus \mathfrak{C}$. Let \mathfrak{A}' be any expansion of \mathfrak{A} to \mathcal{S} . If b is the length of $s(\mathfrak{A}')$, let $f = a/b$. Then $\mathfrak{A}' \sim_r \sim_r \mathfrak{A}' \oplus \mathfrak{E}(s(\mathfrak{A}'), kf)$.

Corollary 4. Let d and r be natural numbers. Then there are structures \mathfrak{A} and \mathfrak{B} such that \mathfrak{A} is a cycle, \mathfrak{B} is the cardinal sum of two cycles, and $\mathfrak{A} \rightarrow_r^d \mathfrak{B}$.

Proof. Immediate from Theorem 3.

Theorem 5. Let \mathcal{A} be the class of nonconnected, finite $\{P\}$ -structures. Then \mathcal{A} is a monadic generalized spectrum, but $\tilde{\mathcal{A}}$ is not. Hence, the class of monadic generalized spectra is not closed under complement.

Proof. \mathcal{A} is a monadic generalized spectrum, via

$$\exists U (\exists x Ux \wedge \exists x \sim Ux \wedge \forall x \forall y (Pxy \rightarrow (Ux \leftrightarrow Uy))).$$

Assume that \mathcal{A} is a monadic generalized spectrum. From Theorem 1, we can find natural numbers d and r such that if $\mathfrak{A} \in \tilde{\mathcal{A}}$, \mathfrak{B} is a finite $\{P\}$ -structure, and $\mathfrak{A} \rightarrow_r^d \mathfrak{B}$, then $\mathfrak{B} \in \tilde{\mathcal{A}}$. But this contradicts Corollary 4.

We close by noting that $\tilde{\mathcal{A}}$ is a generalized spectrum with one existentialized binary predicate symbol $<$. Let σ be the first-order sentence " $<$ is a strict partial order (transitive and irreflexive) with a largest element, and if y is an immediate successor of x , then $Pxy \vee Pyx$." Then $\tilde{\mathcal{A}}$ is the class of finite models of $\exists < \sigma$: Clearly if \mathfrak{A} is finite and $\mathfrak{A} \models \exists < \sigma$, then \mathfrak{A} is connected. Conversely, assume that \mathfrak{A} is finite and connected. Select any a in $|\mathfrak{A}|$; this will be the largest element in the partial order. We will define various "levels" which partition $|\mathfrak{A}|$. The first, or top, level contains only a . The second level contains every point in $|\mathfrak{A}|$ (except a) which connects to a (b connects to a if $P^{\mathfrak{A}}ab$ or $P^{\mathfrak{A}}ba$). The third level contains every point not in the first or second level which connects to a point in the second level, and so on. Define $<_1$ on $|\mathfrak{A}|$ by saying that $x <_1 y$ if x and y connect and if x is one level below y . Let $<_2$ be the transitive closure of $<_1$; then $<_2$ is the desired strict partial order.

Bibliography

- [1] ASSER, G., Das Repräsentantenproblem im Prädikatenkalkül der ersten Stufe mit Identität. This Zeitschr. 1 (1955), 252–263.
- [2] BÜCHI, J. R., Weak second-order arithmetic and finite automata. This Zeitschr. 6 (1960), 66–92.
- [3] FAGIN, R., Contributions to the model theory of finite structures. Doctoral dissertation, Univ. of Calif., Berkeley, 1973.
- [4] FAGIN, R., Generalized first-order spectra and polynomial-time recognizable sets. In: Complexity of Computation (ed. R. KARP), SIAM-AMS Prog. 7 (1974), 43–73.
- [5] FRAÏSSÉ, R., Sur les classifications des systèmes de relations. Publications Sc. d l'Université D'Alger 1, No. I (1954).
- [6] RABIN, M. O., Decidability of second-order theories and automata on finite trees. IBM research report RC-2012 (1968).
- [7] SCHOLZ, H., Problems. J. of Symb. Log. 17 (1952), 160.
- [8] TARSKI, A., Contributions to the theory of models I, II. Indag. Math. 16 (1954), 572–588.

(Eingegangen am 6. Dezember 1973)