Decidable Theories of the Ordering of Natural Numbers with Unary Predicates*

Dedicated to Boris A. Trakhtenbrot on the occasion of his 85th birthday

Alexander Rabinovich[1] and Wolfgang Thomas[2]

[1] Tel Aviv University, Department of Computer Science
rabino@math.tau.ac.il
[2] RWTH Aachen, Lehrstuhl Informatik 7, 52056 Aachen, Germany
thomas@informatik.rwth-aachen.de

Abstract. Expansions of the natural number ordering by unary predicates are studied, using logics which in expressive power are located between first-order and monadic second-order logic. Building on the model-theoretic composition method of Shelah, we give two characterizations of the decidable theories of this form, in terms of effectiveness conditions on two types of "homogeneous sets". We discuss the significance of these characterizations, show that the first-order theory of successor with extra predicates is not covered by this approach, and indicate how analogous results are obtained in the semigroup theoretic and the automata theoretic framework.

1 Introduction

In [1], Büchi showed that the monadic theory of the ordering $(\mathbb{N}, <)$ of the natural numbers is decidable. Many authors studied the question for which expansions of $(\mathbb{N}, <)$ this decidability result can be preserved. For most examples of natural functions or binary relations it turned out that the corresponding monadic theory is undecidable, usually shown via an interpretion of first-order arithmetic. This applies, for instance, to the double function $\lambda x.2x$ ([9,20]).

For the expansion of $(\mathbb{N}, <)$ by unary predicates, the situation is different: Many examples P of such predicates are known such that the monadic theory of $(\mathbb{N}, <, P)$ is decidable, among them – as shown by Elgot and Rabin [5] – the set of factorial numbers, the set of powers of k and the set of k-th powers (for fixed k). A larger class of such predicates was presented in [3,4]; another comprehensive study is [11]. Contrary to the case of functions, no "natural" recursive predicate P is known such that the monadic theory of $(\mathbb{N}, <, P)$ is undecidable. Moreover, it is known that in the cases where undecidability holds, the undecidability proof cannot be done via an interpretation of first-order arithmetic (see [2,16]).

* This paper was written during a visit of the first author in Aachen in March 2006, funded by the European Science Foundation ESF in the Research Networking Programme AutoMathA (Automata: From Mathematics to Applications).

Z. Ésik (Ed.): CSL 2006, LNCS 4207, pp. 562–574, 2006.

The approach introduced by Elgot and Rabin [5] for showing decidability of the monadic theory of a structure $(\mathbb{N}, <, P)$ is built on a method to decompose this structure in a "periodic" way, together with the translation of monadic formulas to Büchi automata. By this translation, the monadic theory of $(\mathbb{N}, < , P)$ is decidable iff the following decision problem $\mathrm{Acc}_{u_\mathbf{P}}$ is decidable for the characteristic ω-word u_P associated with P (where $u_P(i) = 1$ if $i \in P$ and otherwise $u_P(i) = 0$):

$$(\mathrm{Acc}_{u_P})\text{: Given a Büchi automaton } \mathcal{A}\text{, does } \mathcal{A} \text{ accept } u_P?$$

Considering the predicate F of factorial numbers as an example, Elgot and Rabin defined for a given Büchi automaton $\mathcal{A}$ a "contraction" $c(u_F)$ of u_F which is accepted by $\mathcal{A}$ iff u_F is accepted by $\mathcal{A}$. The contraction $c_\mathcal{A}(u_F)$ is obtained from u_P by applying a pumping argument to the 0-segments between successive letters 1. The word $c_\mathcal{A}(u_F)$ has 0-segments of bounded length and turns out to be ultimately periodic; so one can decide whether $\mathcal{A}$ accepts $c_\mathcal{A}(u_F)$ and hence u_F. Also the method of [3,4] follows this pattern: Decidability of Acc_{u_P} is reduced to the question whether, given a finite semigroup (replacing the Büchi automaton as used by Elgot and Rabin), one can compute a representation of an ultimately periodic word which can replace u_P for answering the question about u_P. An abstract version of this "effective reduction to ultimately periodic predicates" is given in our main theorem below. As a key tool we use Ramsey's Theorem on the existence of homogeneous sets over colored orderings of order type ω (as already Büchi did in [1]).

In [8] this "non-uniform" procedure of reducing u_P to ultimately periodic sequences, depending on the monadic formula, the Büchi automaton, or the semigroup under consideration, was replaced by a "uniform" periodicity condition on P, thus settling a conjecture raised in [4]. The main result of [8] states that the monadic theory of $(\mathbb{N}, <, P)$ is decidable iff a recursive predicate P' exists which is "homogeneous for P". This predicate captures, in some sense, all the ultimately periodic structures that arise from the non-uniform approach mentioned above.

The purpose of the present paper is to give a streamlined proof of the result of [8], clarifying the connection to the "non-uniform" method, and at the same time generalizing it from monadic logic to a class of logics between first-order logic and monadic logic. We also discuss the case where the successor relation S is considered instead of the ordering $<$ (a modification which is irrelevant when monadic logic is considered). As in [8], we present the proofs in a logical framework, avoiding the use of automata or semigroups, and building on composition theorems in the spirit of Shelah [12] (see also [6,18]). As explained in Section 4.3, however, the arguments do not depend on this framework and can be transferred easily to the domains of automata, respectively semigroups.

The present work represents a merge of ideas of the two independently written and so far unpublished papers [8,15] (of 2005 and 1975, respectively).

In the subsequent section we introduce the background theory and state the main result. Section 3 is devoted to the proof. In Section 4, we discuss several aspects of the main theorem: its significance, its failure for the first-order theory of successor, and the alternative frameworks of semigroups and automata in place of logic. A summary and outlook conclude the paper.

2 Logical Background and Main Result

The structures considered in this paper are of the form $M = (\mathbb{N}, <, P_1, \dots, P_m)$ with $P_i \subseteq \mathbb{N}$. We call them m-labelled ω-chains. These structures are in one-to-one correspondence with ω-words over the alphabet $\{0,1\}^m$. The ω-word $u_{\overline{P}} = u_{\overline{P}}(0)u_{\overline{P}}(1)\dots$ corresponding to $\overline{P} = (P_1, \dots, P_m)$ has value 1 in the j-th component of $u_P(i)$ iff $i \in P_j$. By an m-labelled chain we mean a linear ordering $(A, <, P_1, \dots, P_m)$ with finite A or $A = \mathbb{N}$.

Let us recall some standard logical systems; here we assume that the signature is chosen according to the type of structure above (and in our notation we do not distinguish, for example, between the relation $<$ and the relation symbol denoting it). The system of first-order logic FO[<] has, besides equality, the mentioned relation symbols $<, P_1, \dots, P_m$. The atomic formulas are of the form $x = y, x < y, P_i(x)$ with first order variables x, y. Formulas are built up using boolean connectives and the first-order quantifiers $\exists, \forall$. In the first-order logic FO[S], the successor relation S is used in place of $<$.

It is known that over labelled chains one can increase the expressive power of first-order logic by adjoining "modular counting quantifiers" $\exists^{r,q}$ (with $0 \leq r < q$), where $\exists^{r,q} x \varphi(x)$ means that the number of elements x satisfying φ is finite and equal to r modulo q. We denote this logic by FO[<]+MOD. A detailed introduction is given in [13].

Still more expressive are the logical systems MSO of monadic second-order logic and WMSO of weak monadic second-order logic. They arise from FO[<] by adding unary second-order variables $X, Y, \dots$ and corresponding atomic formulas (written $X(y)$ etc.). In MSO, quantification over set variables ranges over the subsets of $\mathbb{N}$, in WMSO only over the finite subsets of $\mathbb{N}$. Over labelled ω-orderings, WMSO and MSO have the same expressive power, which however exceeds that of FO[<]+MOD (cf. [17,13]).

In the sequel the letter L stands for any of the logics introduced above. The L-theory of $(\mathbb{N}, <, \overline{P})$ is the set of L-sentences which are true in $(\mathbb{N}, <, \overline{P})$.

For the analysis of the L-theory of $(\mathbb{N}, <, \overline{P})$ we use the composition method which was developed by Shelah [12] for monadic second-order logic. We recall the facts underlying the composition method.

Two m-labelled chains M, M' are called k-equivalent for L (written: $M \equiv_k^L M'$) if $M \models \varphi \Leftrightarrow M' \models \varphi$ for every L-sentence φ of quantifier depth k. This is an equivalence relation between labelled chains; its equivalence classes are called k-types for L (and for the given signature with $<$ and m unary predicate symbols). Let us list some fundamental and well-known properties of k-types for any of the logics L above; here we suppress the reference to L for simplicity of notation.

Proposition 1. *1. For every m and k there are only finitely k-types of m-labelled chains. (In the case of FO[<]+MOD, we assume that also a maximal divisor q is fixed in advance.)*

2. For each k-type t there is a sentence (called "characteristic sentence") which defines t (i.e., is satisfied by a labelled m-chain iff it belongs to t). For given k and m, a finite list of characteristic sentences for all the possible k-types can be computed. (We take the characteristic sentences as the canonical representations of k-types. Thus, for example, transforming a type into another type means to transform sentences.)

3. Each sentence φ is equivalent to a (finite) disjunction of characteristic sentences; moreover, this disjunction can be computed from φ.

The proofs of these facts can be found in several sources, we mention [12,18,19] for MSO and FO, and [13] for FO[<]+MOD.

As a simple consequence we note that the L-theory of an m-labelled chain M is decidable iff the function which associates to each k the k-type of M for L is computable.

Given m-labelled chains M_0, M_1 we write $M_0 + M_1$ for their concatenation (ordered sum). In our context, M_0 will always be finite and M_1 finite or of order type ω. Similarly, if for $i \geq 0$ the chain M_i is finite, the model $\Sigma_{i\in\mathbb{N}} M_i$ is obtained by the concatenation of all M_i in the order given by the index structure $(\mathbb{N}, <)$.

We need the following composition theorem on ordered sums:

Theorem 2 (Composition Theorem). *Let L be any of the logics considered above.*

(a) *The k-types of m-labelled chains M_0, M_1 for L determine the k-type of the ordered sum $M_0 + M_1$ for L, which moreover can be computed from the k-types of M_0, M_1.*

(b) *If the m-labelled chains $M_0, M_1, \ldots$ all have the same k-type for L, then this k-type determines the k-type of $\Sigma_{i\in\mathbb{N}} M_i$, which moreover can be computed from the k-type of M_0.*

Part (a) of the theorem justifies the notation $s + t$ for the k-type of an m-chain which is the sum of two m-chains of k-types s and t, respectively. Similarly, we write $t * \omega$ for the k-type of a sum $\Sigma_{i\in\mathbb{N}} M_i$ where all M_i are of k-type t.

Let us call a logic L *compositional* if the Composition Theorem above with parts (a) and (b) holds. All logics L listed above are compositional. For FO[<] and WMSO this goes back to Läuchli, for MSO to Shelah [12], and for FO[S] and FO[<]+MOD one may consult [13].

The fundamental fact which enters all decidability proofs below (and which underlies also Büchi's work [1]) is the following: The two parts (a) and (b) of the Composition Theorem suffice to generate the k-types of arbitrary (even non-periodic) m-labelled chains $M = (\mathbb{N}, <, P_1, \ldots, P_m)$. This is verified by decomposing M into segments such that all of them except possibly the first one have the same k-type. The elements (numbers) that separate the segments of such a decomposition form a "homogeneous set". Given $M = (\mathbb{N}, <, \overline{P})$, let us write $M[i,j)$ for the m-labelled chain with domain $[i,j) = \{i,\ldots,j-1\}$ and the predicates $<$ and $\overline{P}$ restricted to $[i,j)$.

Definition 3 (k-homogeneous set). *A set $H = \{h_0 < h_1 < \ldots\}$ is called* k-homogeneous for $M = (\mathbb{N}, <, \overline{P})$ *with respect to L, if all segment models $M[h_i, h_j)$ for $i < j$ (and hence all segment models $M[h_i, h_{i+1})$ for $i \geq 0$) have the same k-type for L.*

In the main theorem below, a stronger notion of homogeneity [8] enters:

Definition 4 (uniformly homogeneous set). *A set $H = \{h_0 < h_1 < \ldots\}$ is called* uniformly homogeneous for $M = (\mathbb{N}, <, \overline{P})$ *with respect to L if for each k the set $H_k = \{h_k < h_{k+1} < \ldots\}$ is k-homogeneous with respect to L.*

The existence of uniformly homogeneous sets will be shown in the next section, while the existence of k-homogeneous sets is well-known (see e.g. [17]):

Proposition 5 (Ramsey). *Let f be a function from $\mathbb{N}^2$ into a finite set C. Then there is $c \in C$ and an infinite set H such that $f(i,j) = c$ for all $i < j \in H$.*

In particular, if L is a logic satisfying item 1 of Proposition 1, and M an m-labelled ω-chain, there is a k-homogeneous set for M with respect to L.

Given a k-homogeneous set $H = \{h_0 < h_1 < \ldots\}$ for $M = (\mathbb{N}, <, \overline{P})$ with respect to L, the Composition Theorem implies that the k-type for L of $M = (\mathbb{N}, <, \overline{P})$ can be computed from the k-types for L of $M[0, h_0)$ and of $M[h_0, h_1)$; note that all the segment models $M[h_i, h_{i+1})$ have the same k-type for L.

For two k-types s, t (for L) of m-labelled chains consider the following condition:

$\mathbf{Hom}^L_{s,t}$:

There is a k-homogeneous set $H = \{h_0 < h_1 < \ldots\}$ with respect to L such that $M[0, h_0)$ has k-type s and $M[h_0, h_1)$ has k-type t for L.

If $\mathrm{Hom}^L_{s,t}$ is true in $M = (\mathbb{N}, <, \overline{P})$, the k-type of $M = (\mathbb{N}, <, \overline{P})$ for L is computable as the type $s + t * \omega$. Thus, Ramsey's Theorem reduces the decision problem for the L-theory of $M = (\mathbb{N}, <, \overline{P})$ to the problem of deciding, for each k and k-types s, t, whether the statement $\mathrm{Hom}^L_{s,t}$ holds in M. Ramsey's Theorem guarantees that for given M and k such a pair (s,t) of k-types exist.

For an m-labelled ω-chain M, let $\mathrm{RecRamsey}^L(M)$ be the following condition:

$\mathbf{RecRamsey}^L(M)$:

There is a recursive function assigning to each k a pair of k-types s and t for L such that $\mathrm{Hom}^L_{s,t}$ holds in M.

We call the logic L *expressive for the existence of homogeneous sets* if for any k-types s, t for L, there is an L-sentence which expresses $\mathrm{Hom}^L_{s,t}$.

We can now state our main result.

Theorem 6. *Let L be a logic which is both compositional and expressive for the existence of homogeneous sets. Then the following are equivalent for any given m-labelled ω-chain $M = (\mathbb{N}, <, \overline{P})$ with recursive sets $\overline{P}$:*

1. *The L-theory of M is decidable.*
2. *RecRamsey$^L(M)$.*
3. *There is a recursive uniformly homogeneous set for M with respect to L.*

Let us verify that the theorem covers all the logics L mentioned above, excepting FO[S]. For this it remains to show that FO[<], FO[<]+MOD, WMSO, MSO are expressive for the existence of homogeneous sets.

This is obvious for MSO; in a straightforward formalization of $\mathrm{Hom}^L_{s,t}$ one uses an existential set quantifier $\exists X$ and relativizes the characteristic sentences for s and t to the segments from 0 to the first element of X, respectively to the segments enclosed by successive X-elements. For the remaining logics it suffices to show the following (see e.g. [17]).

Proposition 7. *FO[<] is expressive for the existence of homogeneous sets.*

Proof. We write $T_k[x,y) = t$ for a formula expressing that the k-type of the segment $[x,y)$ (for FO[<]) is t. The proof covers all logics L considered here which extend FO[<]; in our notation we suppress the reference to FO[<] or to such L. Note that $\mathrm{Hom}_{s,t}$ can only hold for a k-type t with $t = t + t$. We show that $\mathrm{Hom}_{s,t}$ holds iff

$$\exists x(T_k[0,x) = s \ \wedge \ \forall y\, \exists z z' \ (\ y < z < z' \ \wedge \ T_k[x,z) = t \ \wedge T_k[z,z') = t\ )$$

The direction from left to right is trivial; take, e.g., for x the minimal element of the homogeneous set given by the assumption.

For the direction from right to left choose a number x as given by the formula, and apply its latter clause by choosing a sequence of numbers $z_1 < z_1' < z_2 < z_2' < \ldots$ such that $T_k[x,z_i) = T_k[z_i,z_i') = t$ and hence (note that $t+t = t$) $T_k[x,z_i') = t$. We shall find a subsequence $z_1 < z_{i_1} < z_{i_2} \ldots$of $z_1 < z_2 < \ldots$ such that $T_k[x,z_{i_m}) = T_k[z_{i_m}, z_{i_n}) = t$ for all $m < n \in \mathbb{N}$.

Define a coloring col from $\mathbb{N}^2$ to the set T_k of all k-types as follows: $\mathrm{col}(i,j) = T_k[z_i', z_j)$. By Ramsey's theorem there is $t_1 \in T_k$ and an infinite set $i_1 < i_2 \ldots$ such that $\mathrm{col}(i_m, i_n) = t_1$ for all $m < n$. Note that for $m < n \in \mathbb{N}$:

$$t = T_k[x, z_{i_n}) = T_k[x, z_{i_m}') + T_k[z_{i_m}', z_{i_n}) = t + t_1$$

Hence,

$$T_k[z_{i_m}, z_{i_n}) = T_k[z_{i_m}, z_{i_m}') + T_k[z_{i_m}', z_{i_n}) = t + \mathrm{col}(i_m, i_n) = t + t_1 = t.$$

□

Let us address the relation of Theorem 6 to the main result of [8]. It is shown there that the following conditions are equivalent:

1. The monadic (second-order) theory of $M = (\mathbb{N}, <, \overline{P})$ is decidable.
2. There is a recursive uniformly homogeneous set for M with respect to the monadic (second-order) logic.

The proof of the implication (1)⇒(2) in [8] relies on the expressive power of MSO-logic and proceeds as follows. For $k = 1, 2, \ldots$ an MSO-formula $H_k(X, Y, \overline{P})$ is constructed (effectively from k and the number of predicates in $\overline{P}$) which defines for any infinite subset Y of $M = (\mathbb{N}, <, \overline{P})$ an infinite set $X \subseteq Y$ which is k-homogeneous for M. (The uniqueness proof for X requires a nontrivial uniformization result, using [7].) Hence, the set Q_1 such that $M \models H_1(Q_1, \mathbb{N}, \overline{P})$ is 1-homogeneous for M, and more generally the set Q_{k+1} such that $M \models H_{k+1}(Q_{k+1}, Q_k, \overline{P})$ is $(k+1)$-homogeneous for M. The sets $Q_1 \supseteq Q_2 \supseteq \ldots$ are definable by formulas and therefore are recursive (in the monadic theory of M). Hence, the set $H = \{a_k \ : \ a_k$ is k-th element of $Q_k\}$ is recursive and uniformly homogeneous for M.

In the present paper, the proof of Theorem 6 relates the conditions of non-uniform and uniform homogeneity in a direct way, covers more logics (between FO[<] and MSO) than MSO, and is somewhat simpler since it does not involve the uniformization result of [7].

3 Proof of Theorem 6

For the conditions

(1) The L-theory of M is decidable
(2) RecRamsey$^L(M)$
(3) There is a recursive uniformly homogeneous set for M with respect to L

we show the implication chain (3) ⇒ (2) ⇒ (1) ⇒ (3).

(3)⇒(2). Assume that $H = \{h_0 < h_1 < \ldots\}$ is recursive and uniformly homogeneous for M with respect to L.

Let k be a natural number. If s is the k-type of $M[0, h_k)$ and t is the k-type of $M[h_k, h_{k+1})$ then $M \models Hom^L_{s,t}$. Let t_i be the k-type of one element chain $M[i, i]$. Note that t_i is computable because M is recursive. The k-type of $M[0, h_k)$ is $s = \sum_{i=0}^{i=h_k-1} t_i$ and the k type of $t = M[h_k, h_{k+1})$ is $\sum_{i=h_k}^{i=h_{k+1}-1} t_i$. These sums are computable from the Composition Theorem.

(2)⇒(1). Let ψ be a sentence. In order to check whether ψ holds in M we proceed as follows:

1. Let k be the quantifier depth of ψ.
2. By RecRamsey$^L(M)$ we can compute k-types s and t for L such that $M \models Hom^L_{s,t}$.
3. Hence, the k-type t_1 of M can be computed as $t_1 = s + t * \omega$.
4. In order to check whether $t_1 \rightarrow \psi$ is valid, we can compute a finite disjunction of k-characteristic sentences which is equivalent to ψ, and note that $t_1 \rightarrow \psi$ holds iff t_1 is one of these disjuncts.
5. Finally, $t_1 \rightarrow \psi$ iff ψ holds in M.

(1)⇒(3) Assume that the L-theory of M is decidable. We present an algorithm which enumerates in increasing order the numbers of a recursive homogeneous set $H = \{n_1 < n_2 < n_3 < \ldots\}$ for M. We use T_k^L for the (finite) set of k-types of the language L.

Algorithm

Basis

1. Find $t_1, s_1 \in T_1^L$ such that $t_1 = t_1 + t_1$ and $M \models Hom^L_{s_1,t_1}$. Note that such s_1 and t_1 exist by the Ramsey Theorem. Moreover, there is an algorithm to find s_1 and t_1, because of finiteness of T_1^L, the assumption that $Hom^L_{s_1,t_1}$ is expressible in L, and decidability of the L-theory of M.
2. Let n_1 be the minimal n such that

$$s_1 \text{ is the 1-type of } M[0,n) \text{ and} \tag{1}$$

$$M[n,\infty) \models Hom^L_{t_1,t_1} \tag{2}$$

This number n_1 can be computed as follows. Let $\alpha_s(v)$ be a formula which expresses $T_k[0,v) = s$, and let $\beta_t(v)$ be a formula obtained from the sentence $Hom_{t,t}$ by relativizing all quantifiers to the interval $[v,\infty)$. It is clear that n_1 defined above is the unique natural number that satisfies

$$\gamma(v) =_{def} \alpha_{s_1}(v) \wedge \beta_{t_1}(v) \wedge \forall u((0 < u < v) \rightarrow \neg(\alpha_{s_1}(u) \wedge \beta_{t_1}(u))).$$

From the fact that the L theory of M is decidable and that every natural number n is defined by an L formula $\psi_n(v)$ (computable from n) we can compute this n_1 by finding the minimal number n such that $M \models \exists v\big(\psi_n(v) \wedge \gamma(v)\big)$.

Inductive step $k \mapsto k+1$

1. Find $t_{k+1}, s_{k+1} \in T_{k+1}^L$ such that $t_{k+1} \rightarrow t_k$ and $s_{k+1} \rightarrow t_k$ are valid and $t_{k+1} = t_{k+1} + t_{k+1}$ and $M[n_k,\infty) \models Hom^L_{s_{k+1},t_{k+1}}$. The arguments similar to the arguments in the step 1 of the basis show that t_{k+1}, s_{k+1} are computable.
2. Let n_{k+1} be the minimal $n > n_k$ such that

$$s_{k+1} \text{ is the k+1 type of } M[n_k,n) \text{ and} \tag{3}$$

$$M[n,\infty) \models Hom^L_{t_{k+1},t_{k+1}} \tag{4}$$

The arguments similar to the arguments in the step 2 of the basis show that t_{k+1}, s_{k+1} and n_{k+1} are computable.

It is clear that the set $H = \{n_1 < n_2 < \ldots\}$ generated by our algorithm is recursive. We show that it is uniformly homogeneous:

By our construction for every k the k-type of $M[n_k, n_{k+1})$ is t_k.

Since $s_i \rightarrow t_k$ and $t_i \rightarrow t_k$ for $i > k$ we obtain that the k-type of $M[n_i, n_{i+1})$ is also t_k for all $i > k$. Since $t_k + t_k = t_k$, we obtain that the k type of $M[n_i, n_j)$ is also t_k for all $j > i > k$. This proves the uniform homogeneity of H.

4 Discussion

4.1 Comments on Uniform Homogeneity

The main theorem provides two reductions of the decision problem for the L-theory of a structure $M = (\mathbb{N}, <, \overline{P})$: With the first reduction one can transform the question "Is the sentence φ true in M?" to a problem to determine a decomposition of M into a sequence of segments, which depends only on the complexity k of φ. This decomposition gives two L-types s_k and t_k from which one can infer by a an algorithmic procedure whether φ is implied. The decision problem for the L-theory of M is thus reduced to the question whether the function $k \mapsto s_k, t_k$ is recursive.

The second reduction captures this recursiveness by the recursiveness of a *single* decomposition of M into segments. This single decomposition results from an infinite refinement process of the types s_k, t_k mentioned above, and correspondingly it leads to a sequence of decomposition segments which satisfy k-types for larger and larger k.

In a more general formulation on the existence of uniformly homogeneous sets we can cover arbitrary unary predicates $\overline{P}$ rather than just recursive ones. Consider an m-chain $M = (\mathbb{N}, <, \overline{P})$. Note that the Algorithm of Section 3 is effective when given an oracle which can supply the truth value of any condition $\mathrm{Hom}^L_{s,t}$. So we obtain the following result from the proof of the Theorem 6:

Theorem 8. *Let L be compositional and expressive.*

1. *For each structure $M = (\mathbb{N}, <, \overline{P})$ there is a uniformly homogeneous set H which is recursive in the L-theory of M.*
2. *For each structure $M = (\mathbb{N}, <, \overline{P})$ and each uniformly homogeneous set H for M, the L-theory of M is recursive in the recursion theoretic join of $(\overline{P}, H)$.*

We can refine this result by a bound on the recursion theoretic complexity of H relative to $\overline{P}$. By Proposition 7, $\mathrm{Hom}^L_{s,t}$ is a Σ^0_3 statement over the recursion-theoretic join of the predicates in $\overline{P}$, which implies that $\mathrm{Hom}^L_{s,t}$ is recursive in $\overline{P}'''$, the third jump of the recursion-theoretic join of the predicates in $\overline{P}$. (For recursion theoretic terminology see [10].) Thus in the first part of Theorem 8, H can be chosen to be recursive in $\overline{P}'''$. As shown in [2,17], the quantifier structure of the formula that expresses $\mathrm{Hom}^L_{s,t}$ can be simplified even to a boolean combination of Σ^0_2-formulas. So the recursion theoretic bound on H can be sharpened to "truth-table reducible to $\overline{P}''$". By [16] this is optimal in the sense that bounded truth-table reducibility does not suffice.

While our main theorem provides two characterizations of the decidable L-theories of structures $(\mathbb{N}, <, \overline{P})$, it is not easily applicable in order to find interesting predicates P such that, say, the first-order or the monadic second-order theory of $(\mathbb{N}, <, P)$ is decidable. Let us first compare the condition $\mathrm{RecRamsey}^L(M)$ with the classical method of Elgot and Rabin [5], taking the factorial predicate F as an example. Elgot and Rabin proposed a deterministic procedure to transform F into an ultimately periodic predicate, depending on the given formula (or automaton). The condition $\mathrm{RecRamsey}^L(M)$ only involves the existence of a procedure and does not provide one in concrete examples. However, the decomposition ensured by $\mathrm{RecRamsey}^L(M)$ is "stronger" in the sense that it provides an ultimately constant (and not just periodic) sequence of types.

The uniformly homogeneous sets given by the third clause of the theorem also do not settle (immediately) the decision problem for concrete theories of structures $(\mathbb{N}, <, P)$. A prominent example is the predicate $\mathbb{P}$ of the prime numbers. The open twin prime hypothesis is easily expressible already in FO[<] (we use here for simplicity also the successor relation S, which is definable in FO[<]):

$$\varphi_0 := \forall x \exists y_0 \exists y_1 \exists y_2 (x < y_0 \wedge \mathbb{P}(y_0) \wedge S(y_0, y_1) \wedge S(y_1, y_2) \wedge \mathbb{P}(y_2))$$

Now $k = 5$ is the quantifier depth of an FO[<]-sentence which avoids this abbreviation with S. Taking the uniformly homogeneous set H for $\mathbb{P}$ with respect to FO[<], one could decide φ_0 by inspecting the segment from the 5-th to the 6-th element of H: There are infinitely many twin primes iff a pair of primes of distance 2 occurs in this segment; otherwise all twin primes would be included in the initial segment before. It is clear that H encodes this information not only about the twin primes but all other conceivable configurations of primes that are MSO-definable, for instance patterns within segments of some bounded length. Thus H encodes a lot of known and unknown number theory.

4.2 The Successor Theory

In the main result Theorem 6 we excluded the logic FO[S]. For example, the proof of Proposition 7, which shows that FO[<] is expressive for the existence of homogeneous sets, uses the $<$ relation in an essential way. Indeed, we can show that the main theorem fails for FO[S].

It turns out that a recursive uniformly homogeneous set H encodes more information than needed for deciding FO[S]-sentences. While H supplies information about the infinite occurrence of certain segment types, FO[S]-sentences can only express such occurrences in numbers up to a certain finite bound. Indeed, it is well-known that the FO-theory of $M = (\mathbb{N}, S, P)$ is decidable iff for each isomorphism type τ of finite segments and each m, one can decide whether τ occurs $\geq m$ times in M (see, e.g., [16,19]).

Theorem 9. *There is a recursive predicate P such that the FO-theory of* $(\mathbb{N}, S, P)$ *is decidable but there is no recursive uniformly homogeneous set for* $(\mathbb{N}, S, P)$*.*

Proof. We use a predicate P, presented in [16], for which the FO-theory of $(\mathbb{N}, S, P)$ is decidable whereas the FO-theory of $(\mathbb{N}, <, P)$ is undecidable.

Suppose $\mathcal{P}$ is a procedure which runs through all pairs (i, j) of natural numbers in some order; we write (i_n, j_n) for the n-th pair in this order. $\mathcal{P}$ generates a bit sequence as follows: When treating (i_n, j_n), it checks whether the i_n-th Turing machine runs for at least j_n steps when started on the empty tape. $\mathcal{P}$ outputs $10^n10^{i_n}$ in this case, and otherwise generates just 10^n. The resulting bit sequence u is obtained as the concatenation of the $\mathcal{P}$-outputs. Clearly u is recursive, and it has the property that for each given $w \in \{0,1\}^*$ and threshold number m one can decide whether w occurs m times in u. (To verify this, note that by construction of u, the only segment types that occur infinitely often are in the languages 0^*, 0^*10^*, and, for certain values of i, $0^*10^i10^*$. To test whether a segment of the latter type occurs m times, check the output of procedure $\mathcal{P}$ up to the m-th treatment of Turing machine M_i.) From this fact one infers that the first-order theory of $(\mathbb{N}, S, P)$ is decidable (cf. [16]).

By construction of u, the i-th Turing machine does not halt on the empty tape iff the segment 10^i1 occurs infinitely often in u. We show that the latter can be decided if there is a recursive uniformly homogeneous set H for $(\mathbb{N}, S, P)$.

Let $H = \{h_0 < h_1 < \ldots\}$ be a recursive uniformly homogeneous set for $(\mathbb{N}, S, P)$. Given i choose k large enough such that from a k-type of a 1-labelled chain one can infer whether the following holds:

> $(*)$ there is a sequence of $i+2$ successive elements such that its first and last element are in P but the others are not.

Consider the segment $M[h_k, h_{k+1})$, which can be obtained effectively by recursiveness of H. $M[h_k, h_{k+1})$ satisfies $(*)$ iff 10^i1 occurs infinitely often in u. □

4.3 Algebraic and Automata Theoretic Types

In this section we discuss alternative ways of introducing "k-types", using semigroups or automata rather than formulas to describe properties of words. When referring to a logic L, we assume that it is compositional and expressive for the existence of homogeneous sets.

Recall that for such a logic L, for each k the set T_k^L of k-types of L with the $+$ operation is a finite semigroup. Let $\mathcal{S}_L$ be the family of finite semigroups defined as follows:

> $S \in \mathcal{S}_L$ iff there is $k \in \mathbb{N}$ and a semigroup homomorphism from T_L^k onto S

Note that $\mathcal{S}_{WMSO} = \mathcal{S}_{MSO}$ is the family of all finite semigroups and that $\mathcal{S}_{FO}$ is the family of finite aperiodic semigroups.

Let $\mathcal{S}$ be a family of finite semigroups. Define an equivalence relation $\sim_k^{\mathcal{S}}$ on Σ^+ as follows:

> $w_1 \sim_k^{\mathcal{S}} w_2$ iff $h(w_1) = h(w_2)$ for every $S \in \mathcal{S}$ of size at most k and for every morphism $h : \Sigma^+ \to S$.

The following lemma is technical but straightfoward.

Lemma 10. *1. For every $k \in \mathbb{N}$ there is $m \in \mathbb{N}$ computable from k such that if $w_1 \sim_m^{\mathcal{S}_L} w_2$ then $w_1 \equiv_k^L w_2$.*

2. For every $m \in \mathbb{N}$ there is $k \in \mathbb{N}$ computable from m such that if $w_1 \equiv_k^L w_2$ then $w_1 \sim_m^{\mathcal{S}_L} w_2$.

As representations of such semigroups one may take the transformation semigroups of finite automata extended by information about visited states. Then the parameter k can be set to be the cardinality of automata rather than of semigroups. Formally, one refers to a class $\mathcal{A}$ of finite automata and uses the congruences $\sim_k^{\mathcal{A}}$ over Σ^+ defined as follows: For $w_1, w_2 \in \Sigma^+$, define $w_1 \sim_k^{\mathcal{A}} w_2$ iff for any automaton $A \in \mathcal{A}$ with k states, any states p, q of A and any set $\mathcal{P}$ of states of A, there is a run of A on w_1 from p to q with states forming the set $\mathcal{P}$ iff this holds for w_2.

Definition 11. *An ω-word u is* effectively homogeneous *for a family $\mathcal{S}$ of finite semigroups if there is a recursive ω-sequence $w_1, w_2, \ldots$ of finite words such that $u = w_1 w_2 \ldots$ and for every $k \in \mathbb{N}$ and semigroup $S \in \mathcal{S}$ of size at most k and morphism $h : \Sigma^+ \to S$ there is $s \in S$ such that $h(w_i) = s$ for all $i > k$.*

The following theorem is an immediate corollary of Theorem 6 and Lemma 10.

Theorem 12. *Let L be a logic which is both compositional and expressive for the existence of homogeneous sets. The L-theory of an ω-word u is decidable iff u is effectively homogeneous for $\mathcal{S}_L$.*

Hence we have

Corollary 13. *The FO-theory of an ω-word u is decidable iff u is effectively homogeneous for the family of finite aperiodic semigroups. The WMSO-theory and the MSO-theory of an ω-word u is decidable iff u is effectively homogeneous for the family of finite semigroups.*

5 Conclusion

We analyzed, for some natural logics L including first-order and monadic second-order logic, the decision problem for the L-theories of structures $M = (\mathbb{N}, <, \overline{P})$ where $\overline{P}$ is a tuple of unary predicates. Our main result gave two characterizations of the decidable theories of this form, using recursiveness conditions on two different versions of "homogeneous sets".

As already mentioned, it seems hard to apply the main theorem of this paper as a tool to find new predicates P where, say, the monadic-second theory of $(\mathbb{N}, <, P)$ is decidable, or to establish even an interesting predicate where this theory is undecidable.

Another kind of application, left open in this paper, is the generation of concrete classes of predicates P (by certain closure properties) such that say the MSO theory of $(\mathbb{N}, <, P)$ is decidable. This kind of application would yield decidability results via the transformation of uniformly homogeneous sets.

References

1. J. R. Büchi. On a decision method in restricted second order arithmetic In *Proc. International Congress on Logic, Methodology and Philosophy of Science*, E. Nagel at al. eds, Stanford University Press, pp 1-11, 1960.
2. J.R. Büchi, L.H. Landweber, Definability in the monadic second-order theory of successor. *J. Symb. Logic* 34, 166-170, 1969.
3. O. Carton and W.Thomas. The Monadic Theory of Morphic Infinite Words and Generalizations. In *Proc. MFCS 2000*, Springer LNCS 1893 (2000), 275-284.

4. O. Carton and W.Thomas. The Monadic Theory of Morphic Infinite Words and Generalizations. *Inf. Comput. 176*(1), pp. 51-65, 2002.
5. C. Elgot and M. O. Rabin. Decidability and Undecidability of Extensions of Second (First) Order Theory of (Generalized) Successor. *J. Symb. Logic*, 31(2), pp. 169-181, 1966.
6. Y. Gurevich. Monadic second order theories. In J. Barwise and S. Feferman eds.,em Model Theoretic Logics pp. 479-506, Springer Verlag, 1986.
7. S. Lifsches and S. Shelah. Uniformization and Skolem Functions in the Class of Trees. *J. Symb. Logic*, 63, 103–127, 1998.
8. A. Rabinovich. On decidability of monadic logic of order over the naturals extended by monadic predicates, manuscript, 2005, submitted.
9. R. M. Robinson. Restricted Set-Theoretical Definitions in Arithmetic. *Proceedings of the American Mathematical Society*, Vol. 9, No. 2. pp. 238-242, 1958.
10. H.R. Rogers, *Theory of Recursive Functions and Effective Computability*, McGraw-Hill, New York 1967.
11. A. Semenov. Logical theories of one-place functions on the set of natural numbers. *Mathematics of the USSR - Izvestia*, vol. 22, pp 587-618, 1984.
12. S. Shelah. The monadic theory of order. *Ann. of Math.* 102, 349-419, 1975.
13. H. Straubing. *Finite Automata, Formal Logic, and Circuit Complexity*, Birkhäuser, Boston, 1994.
14. D. Siefkes. Decidable extensions of monadic second-order successor arithmetic. In *Automatentheorie und Formale Sprachen*, J. Dörr and G. Hotz, Eds., pp 441-472, BI-Wissenschaftsverlag, Mannheim 1970.
15. W. Thomas. *Das Entscheidungsproblem für einige Erweiterungen der Nachfolger-Arithmetik*. Ph. D. Thesis Albert-Ludwigs Universität, Freiburg 1975.
16. W. Thomas. The theory of successor with an extra predicate. *Math. Ann.* 237, 121-132, 1978.
17. W. Thomas. Automata on infinite objects. In: *Handbook of Theoretical Computer Science* (J. v. Leeuwen, ed.), Vol. B, Elsevier, Amsterdam 1990, pp. 135-191.
18. W. Thomas. Ehrenfeucht Games, the composition method, and the monadic theory of ordinal words. In *Structures in Logic and Computer Science: A Selection of Essays in Honor of A. Ehrenfeucht*, *Lecture Notes in Computer Science* 1261:118-143, 1997, Springer-Verlag.
19. W. Thomas. Languages, automata, and logic. In: Handbook of Formal Languages (G. Rozenberg, A. Salomaa, eds.), Vol 3. Springer 1997, pp. 389-455.
20. B. A. Trakhtenbrot. Finite automata and the logic of one-place predicates. (Russian version 1961). In AMS Transl. 59, 1966, pp. 23-55.