Pacific Journal of Mathematics

AVOIDABLE PATTERNS IN STRINGS OF SYMBOLS

Dwight Richard Bean, Andrzej Ehrenfeucht
and George Frank McNulty

Vol. 85, No. 2 October 1979

PACIFIC JOURNAL OF MATHEMATICS
Vol. 85, No. 2, 1979

AVOIDABLE PATTERNS IN STRINGS OF SYMBOLS

DWIGHT R. BEAN, ANDRZEJ EHRENFEUCHT AND
GEORGE F. MCNULTY

A *word* is just a finite string of letters. The word W *avoids* the word U provided no substitution instance of U is a subword of W. W is *avoidable* if on some finite alphabet there is an infinite collection of words each of which avoids W. W is kth *power-free* if W avoids x , where x is a letter. We develope the theory of those endomorphisms of free semigroups which preserve kth power-freeness and employ this theory to investigate kth power-free words. We go on to prove that every kth power-free word on n letters is a subword of a maximal word of the same kind. Next we examine avoidable words in general and prove that all words of length at least 2^n on an alphabet with n letters are simultaneously avoidable. We show that on any finite alphabet the collection of avoidable words is simultaneously avoidable. We provide an effective (recursive) characterization of avoidability. Finally we show how our work can be extended to infinite words, to n-dimensional arrays, and to circular words. We give an application to the Burnside problem for semigroups. The present work is chiefly concerned with certain combinatorial properties of strings of symbols. As such, it belongs to formal linguistics, to the theory of free semigroups, and to the theory of partitioned linear orders. While we have taken all of these points of view in the body of this work, it has proven most convenient to base our exposition on an attitude between linguistics and free semigroups.

By an *alphabet* we mean any set, the members of which are called *letters* and can be regarded in all subsequent discussions as indivisible. A *word on* the alphabet N is a finite string of letters belonging to N. For example, if $N = \{a, b, c, d\}$ then $abacd$ is a word on N. The *empty word* is the string with no letters and it is regarded as a word on every alphabet. Words can be concatenated: whenever U and V are words, the result of concatenating U and V is expressed by juxtaposition. If $U = abacd$ and $V = bdaca$, then $UV = abacdbdaca$. If $W = UV$ then U is an *initial segment* of W and V is a *final segment*. U is a *subword* of W provided $W = XUY$ for some words X and Y. For any word W and any natural number k, W^k is defined so that

$$W^0 \text{ is the empty word}$$
$$W^{k+1} = W^k W .$$

The *length* of the word W, denoted by $|W|$, is the number of occurrences of letters in W. Hence $|abacd| = 5$. The word U is a *substitution instance* of the word W provided $W = e_0e_1 \cdots e_{n-1}$, where W is a word on the alphabet $\{e_0, e_1, \cdots, e_{n-1}\}$, and there are words $E_0, E_1, \cdots, E_{n-1}$ such that $U = E_0E_1 \cdots E_{n-1}$ with $E_i = E_j$ if $e_i = e_j$.

$$abxcab$$

is a substitution of

$$xyx\ .$$

(Let $e_0 = e_2 = x$, $e_1 = y$, and $E_0 = ab = E_2$ with $xc = E_1$.)[1]

Concatenation is associative and it is well known that given any alphabet N, the collection of nonempty words on N under concatenation is one way to represent the semigroup freely generated by N. Homomorphisms between free semigroups correlate with substitutions in a natural way. The map h between $\{x, y\}$ and $\{a, b, c, x\}$ given by

$$h(x) = ab$$
$$h(y) = xc$$

can be extended (uniquely) to a homomorphism between the semigroups freely generated by $\{x, y\}$ and $\{a, b, c, x\}$. As in the last example of the previous paragraph, under this homomorphism xyx is carried to $abxcab$. The semigroup freely generated by N is denoted by $\mathscr{F}_N$. Frequently the only relevant fact about the alphabet is its cardinality. So if $n = |N|$ (the cardinality of N), then we sometimes write $\mathscr{F}_n$ for $\mathscr{F}_N$.

DEFINITION 0.0. The word W *avoids* the word U provided no subword of W is a substitution instance of U. When $\mathscr{F}$ and $\mathscr{G}$ are sets of words, we say $\mathscr{F}$ *avoids* $\mathscr{G}$ if every member of $\mathscr{F}$ avoids every member of $\mathscr{G}$. The word W is *avoidable on the n letter alphabet* provided there is an infinite collection $\mathscr{F}$ of words on the n letter alphabet which avoids $\{W\}$. The collection $\mathscr{G}$ is *avoidable on the n letter alphabet* if there is an infinite collection $\mathscr{F}$ of words on the n letter alphabets which avoids G. The collection $\mathscr{G}$ is *avoidable* if $\mathscr{G}$ is avoidable on some finite alphabet.

1. Our description of words, alphabets, and concatenation lacks some precision. Perhaps the most convenient way to remedy this is to axiomatize our intuitive notion of concatenation. This has been done by Alfred Tarski [**28**] and Hans Hermes [**14**]; see Corcoran, Frank, and Maloney [**6**]. Alternatively the ambiguities involved in our discussion of words and alphabets can be avoided at the expense of introducing some simple set theoretic "tricks" and of complicating our notation. Since none of our results depend on such details, we have not found it necessary to do this.

It is not difficult to see that if x and y are letters, then x and xyx are unavoidable and that x^2 is unavoidable on a two letter alphabet. While it is not immediately evident that there are any avoidable words at all, it turns out that for any finite alphabet all but finitely many words are avoidable (see § 3 below).

Among the simplest words are those of the form x^k where x is a letter. If W avoids x^k, we say that W is kth *power-free*; W is *square-free* if W avoids x^2 and it is *cube-free* if it avoids x^3. Let W be a word on the alphabet N. W is a *maximal* kth *power-free word on* N provided W is kth power-free and neither aW nor Wa is kth power-free for any $a \in N$. If M is also an alphabet we call a homomorphism $h: \mathscr{F}_N \to \mathscr{F}_M$ kth *power-free* provided $h(W)$ is kth power-free whenever W is kth power-free.

We conceive a word of type ω as a string of letters extending to the right and arranged like the natural numbers. For example the decimal representation of π can be regarded as a word of type ω on the alphabet $\{\cdot, 0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$. We note the following lemma.

LEMMA 0.1. *Let n be a natural number. The collection $\mathscr{S}$ is avoidable on n letters iff there is a word W of type ω on an n letter alphabet such that W avoids $\mathscr{S}$.*

This lemma is an immediate consequence of König's Infinity Lemma, since the relation "is an initial subword of" is a well-founded partial order of the words on n letters.

Words of the same type as the integers could also be considered here and a statement like Lemma 0.1 would still be true.

ω denotes both the set of all natural numbers and the first infinite cardinal (ordinal); 2^ω is the cardinality of the set of real numbers.

In 1906, Axel Thue in [29] established

A. x^2 is avoidable on a three letter alphabet, and

B. x^3 is avoidable on a two letter alphabet.

Both of these results were independently rediscovered by S. E. Arshon [2] in 1937 and by Marston Morse and Gustav Hedlund (see Morse [20] and Morse and Hedlund [21]) around 1940. It is interesting to note that while Thue saw analogies with the theory of Diophantine equations, the work of Morse and Hedlund was grounded in the investigation of flows on surfaces of negative curvature, and Arshon's work was done in order to answer a question which A.Y. Khinchin posed in January 1933. The papers of J. Leech [18], Hawkins and Mientka [12], Evdokimov [10], Pleasants [25], Justin [16], Entringer, Jackson, and Schatz [9], and Dekking [7] all provide either new proofs of Thue's theorems or extensions of these theorems, which, however, do not deal directly with avoidable words. On the other hand all these papers are united by a common use of combinatorial properties of various homomorphisms. Apparently Thue [30] was the first to make use of such properties in 1912. The present work is a contribution to this line of investigation. Brown [4] and Hedlund [13] collect together remarks concerning these developments.

Thue's theorems have found diverse applications. They played a role of fundamental importance in the solution of the Burnside conjecture. (See Novikov [23], Novikov and Adjan [24], Adjan [1], and Britton [3].) V.L. Murskii [22] employed them in the construction of a finitely based undecidable equational theory of semigroups. Burris ane Nelson [5] use A to show that the lattice of equational theories satisfying $x^2 = x^3$ has an interval isomorphic with the lattice of all equivalence relations on the natural numbers. In the course of extending the work of Burris and Nelson, J. Jezek [15] proved the following amazing theorem.

There is an infinite set $\mathscr{F}$ of square-free words on three letters such that $\mathscr{F} \sim \{W\}$ avoids W for all $W \in \mathscr{F}$.

Our principal concern in § 1 is kth power-free homomorphisms. We establish the existence of a homomorphism $h: \mathscr{F}_\omega \to \mathscr{F}_3$ which is kth power-free for all $k \geqq 2$ and a homomorphism $g: \mathscr{F}_\omega \to \mathscr{F}_2$ which is kth power-free for all $k \geqq 3$. We also prove that there are 2^ω square-free words of type ω on three letters, no two of which have any common final segments. A similar result holds for cube-free words on two letters.

Section 2 deals with maximal kth power-free words. It is shown that every kth power-free word on n letters is a subword of a maximal kth power-free word on n letters. We also prove that for all $n, k \geqq 2$ except $n = 2 = k$ there are infinitely many maximal kth power-free words on n letters.

Section 3 takes up avoidable sets of words in general. After establishing that certain sets of words on an infinite alphabet are avoidable (i.e., avoided on a finite alphabet), we prove that every word on n letters of length at least 2^n is avoidable. The collection of all avoidable words on an n letter alphabet also turns out to be an avoidable collection. Finally, an effective characterization of the notion of avoidable words is presented. The use of endomorphisms in this section was suggested by the work of Z. Harris in the early 1950's.

Some applications and extensions are collected in § 4. We extend the notion of square-freeness from finite linear orders to arbitrary ordinals and to the reals and the rationals. Every ordinal $\alpha \in (2^{\omega})^{+}$ can be "colored" in a square-free fashion with three colors, whereas two colors suffice for the reals and for the rationals. Next we discuss how the one dimensional notion of square-freeness might be extended to n-dimensional arrays. As a result, we show how the plane can be covered with square tiles of three colors so that no rectangular pattern is repeated adjacent to itself. After this we show that if Σ is a set of semigroup equations on k variables where each side of each equation has length at least 2^k then some finitely generated semigroup free with respect to Σ is infinite. Finally, we consider how to extend the notion of kth power-free words to periodic words of type ω. This is equivalent to extending the notion to necklaces of beads of different colors. As a result we find, for example, that there are arbitrarily large "square-free" necklaces using just beads of three colors.

Despite a period of investigation extending over seventy years highlighted by diverse motivations and surprising applications, no well understood coordinated theory has emerged and many problems remain open. In § 5, we collect some of these problems.

1. Homomorphisms between free semigroups which preserve square-freeness.

Axel Thue was among the first to investigate free semigroups. The theorem below plays a key role in the present work. It is a small improvement of Satz 17 in [30]. The proof we present is essentially Thue's; we include it here since [30] is not generally available.

THEOREM 1. *Let M and N be alphabets and let h be a homomorphism from $\mathscr{F}_M$ into $\mathscr{F}_N$. If*

( ) *$h(W)$ is square-free whenever W is a word on M which is square-free and of length no greater than three, and*

(1) *$a = b$ whenever $a, b \in M$ with $h(a)$ a subword of $h(b)$, then $h(U)$ is square-free whenever U is a square-free word on M.*

Proof. Suppose h is a homomorphism fulfilling conditions (0) and (1).

Claim. If $a, e_0, e_1, \cdots, e_n \in M$ and $h(e_0 e_1 \cdots e_n) = Xh(a)Y$ for some (possibly empty) words X and Y on N, then $a = e_j$, $h(e_0 \cdots e_{j-1}) = X$, and $h(e_{j+1} \cdots e_n) = Y$ for some $j = 0, 1, \cdots, n$.

Proof of claim. It does no harm to suppose $n \geqq 1$. From condition (1) we know that $h(a)$ is a subword of $h(e_j e_{j+1})$ for some

$j = 0, \cdots, n-1$. Suppose the claim fails. Then $h(e_j) = AB$, $h(a) = BC$, and $h(e_{j+1}) = CD$ for some words A, B, C and D on N where B and C are nonempty. This means that neither $h(ae_ja)$ nor $h(ae_{j+1}a)$ is square-free. According to condition (0)$a = e_j = e_{j+1}$ and hence $h(a) = AB = BC = CD$. Therefore A and C are words of the same length and moreover $B = D$. That is $h(a) = BC = CB$. Without loss of generality we can say that $C = BE$ for some possibly empty word E. But then $h(a) = BC = BBE$, in violation of condition (0). So the claim must hold.

Now suppose U is a word on M but $h(U) = XYYZ$ where X, Y, and Z are words on N with Y nonempty. We will prove that U is not square-free. Let $U = e_0e_1 \cdots e_n$, wiht $E_0 = h(e_0)$, $E_1 = h(e_1), \cdots, E = h(e_n)$. By shortening U if necessary we can let $E_0 = XE_0''$ and $E_n = E_n'$ where E_0'' and E_n' are nonempty and $YY = E_0''E_1 \cdots E_n'$. By condition (0) we know $n \geqq 3$. By condition (1) Y is not a subword of either E_0'' or E_n'. So there is j with $0 < j < n$ and $E_j = E_j'E_j''$ with E_j'' nonempty and

$$Y = E_0''E_j \cdots E_j'$$
$$Y = E_j''E_{j+1} \cdots E_n' .$$

Now by the claim $E_0'' = E_j''$, $E_j' = E_n'$, $n = 2j$ and $E_i = E_{j+i}$ for all i with $0 < i < j$. But then $h(e_0e_je_n) = E_0E_jE_n = XE_0''E_j'E_0''E_j'Z$ which is not square-free. By condition (0) either $e_0 = e_j$ or $e_j = e_n$. Say $e_0 = e_j$, the other case being similar. Therefore $E_i = E_{j+i}$ for all $i < j$. On the basis of condition (0) it is easy to establish that h is one-to-one on M. Hence $e_i = e_{j+i}$ for all $i < j$. This means that U is not square-free. (In fact $U = e_0e_1 \cdots e_{j-1}e_0e_1 \cdots e_{j-1}e_{2j}$.) So the proof of the theorem is complete.

Theorem 1.0 furnishes an easily applicable sufficient condition that will allow us to construct square-free homomorphisms. This is so because a homomorphism between free semigroups is uniquely determined by its behavior at the generators. Moreover any map from M into $\mathscr{F}_N$ can be (uniquely) extended to a homomorphism from $\mathscr{F}_M$ into $\mathscr{F}_N$. If M is a finite alphabet, then the hypothesis of the theorem requires checking finitely many cases. For example, if $M = \{a, b, c\}$ then to check condition (0) one need only verify that the images of the twelve words

aba	bab	cab
abc	bac	cac
aca	bca	cba
acb	bcb	cbc

are square-free. Establishing condition (1) is easier.

COROLLARY 1.1. (A. Thue [30]) *The maps* $h: \mathscr{F}_3 \to \mathscr{F}_3$ *and* $g: \mathscr{F}_3 \to \mathscr{F}_3$ *determined by*

$$h(a) = abcab$$
$$h(b) = acabcb$$
$$h(c) = acbcacb$$

and

$$g(a) = abacb$$
$$g(b) = abcbac$$
$$g(b) = abcacbc$$

are both square-free.

Notice that both h and g map the letters to words of length 5, 6, and 7. Aside from trivial maps obtained by renaming the letters, these two maps which Thue found in 1912 are the simplest square-free maps which do not take some letter to a word of length at least eight. In fact, every other nontrivial square-free map from $\mathscr{F}_3$ into $\mathscr{F}_3$ that does not require a word of length eight or more can be obtained from Thue's maps by renaming the letters and/or reversing the words. For example,

$$a \longrightarrow cbacb$$
$$b \longrightarrow cacbab$$
$$c \longrightarrow cabacab$$

induces a square-free map.

COROLLARY 1.2. (A. Thue [29]) *The word* xx *is avoidable on a three letter alphabet.*

Proof. The word a is square-free. $\{h^n(a): n \in \omega\}$ avoids xx by Corollary 1.1, where h is the map described there.

J. Leech in [18] and P. A. B. Pleasants [25] also established Corollary 1.2 by similar arguments. In each case, a particular homomorphism is shown to be square-free. In fact, Pleasants uses the same map h that Thue used, while Leech employs the more symmetrical

$$a \longrightarrow abcbacbcabcba$$
$$b \longrightarrow bcacbacabcacb$$
$$c \longrightarrow cabacbabcabac\ .$$

Both Leech and Pleasants were not aware of Thue's work.

Corollary 1.1 establishes the existence of a nontrivial square-free endomorphism of $\mathscr{F}_3$. Square-free endomorphisms on $\mathscr{F}_n$, where $n > 3$, are easier to construct. For example, the endomorphism of $\mathscr{F}_4$ induced by

$$\begin{aligned} a &\longrightarrow abdba \\ b &\longrightarrow bcacb \\ c &\longrightarrow cdbdc \\ d &\longrightarrow dacad \end{aligned}$$

is easily seen to be square-free and to enjoy symmetry properties similar to Leech's endomorphism of $\mathscr{F}_3$.

COROLLARY 1.3. *There is a square-free homomorphism from $\mathscr{F}_5$ into $\mathscr{F}_4$.*

Proof. By Theorem 1.0, the homomorphism induced by

$$\begin{aligned} a &\longrightarrow abcd \\ b &\longrightarrow abdc \\ c &\longrightarrow acbd \\ d &\longrightarrow abcd \\ e &\longrightarrow adbc \end{aligned}$$

is easily seen to be square-free.

COROLLARY 1.4. *There is a square-free homomorphism from $\mathscr{F}_\omega$ into $\mathscr{F}_5$.*

Proof. Let $a_0, a_1, a_2, \cdots$ be a nonrepetitive listing of the denumerable alphabet (so $\mathscr{F}_\omega$ is the semigroup freely generated by $\{a_i: i \in \omega\}$). By Corollary 1.1 there are infinitely many square-free words on $\{a, b, c\}$. Let $W_0, W_1, W_2, \cdots$ be a nonrepetitive list of them. The map induced by

$$a_i \longrightarrow dW_ieW_i \quad \text{for all} \quad i \in \omega$$

is square-free according to Theorem 1.0.

COROLLARY 1.5. *There is a square-free homomorphism from $\mathscr{F}_4$ into $\mathscr{F}_3$.*

Proof. The endomorphism of $\mathscr{F}_4$ induced by

$$\begin{aligned} a &\longrightarrow babcbd \\ b &\longrightarrow bacabd \\ c &\longrightarrow bcacbd \\ d &\longrightarrow bcbabd \end{aligned}$$

is square-free. Call it f. Let $\hat{h}$ denote the extension of $h \circ h$ (where h is defined in Corollary 1.1) by

$$d \longrightarrow abcabacabcbacabacbabcabacabcb\ .$$

Then $\hat{h} \circ f$ is a square-free homomorphism from $\mathscr{F}_4$ into $\mathscr{F}_3$.
The image of a is
abcabacbcacbabcabacabcbacbcacbacabcbabcabacabcbacbcacbabcabacabcbabc
abacbbacabcbacacbacacbcacbacabcbabcabacbcacbacabcbacbcacbabcabacbcac
bacabcbabcabacbcacbabcabacabcbacbcacbacabcbabcabacabcbacabacbabcaba
cabcb
The image of a has length 209.
The image of b is
abcabacbcacbabcabacabcbacbcacbacabcbabcabacabcbacbcacbabcabacabcbabc
abacbcacbacabcbacbcacbabcabacbcacbacabcbabcabacabcbacbcacbabcabacab
cbabcabacbcacbabcabacabcbacbcacbacabcbabcabacabcbacabacbabcabacabcb
The image of b has length 202.
The image of c is
abcabacbcacbabcabacabcbacbcacbacabcbabcabacbcacbacabcbacbcacbabcabac
bcacbacabcbabcabacabcbacbcacbabcabacabcbabcabacbcacbacabcbacbcacbabc
abacbcacbacabcbabcabacbcacbabcabacabcbacbcacbacabcbabcabacabcbacaba
cbabcabacabcb
The image of c has length 216.
The image of d is
abcabacbcacbabcabacabcbacbcacbacabcbabcabacbcacbacabcbacbcacbabcabac
bcacbacabcbabcabacbcacbabcabacabcbacbcacbacabcbabcabacabcbacbcacbabc
abac abcbabcabacbcacbabcabacabcbacbcacbacabcbabcabacabcbacabacbabca
bacabcb
The image of d has length 209.
That $\hat{h} \circ f$ fulfills condition (1) of Theorem 1.0 merely takes some patience to verify, but that it fulfills conditions (0) required roughly nine hours of run time on a PDP-11/45 computer.

EXAMPLE 1.6. There is an endomorphism of $\mathscr{F}_3$ which fulfills condition (0) of Theorem 1.0 and yet fails to be square-free.

Proof. We construct such an endomorphism of $\mathscr{F}_5$. Let k be the endomorphism induced by

$$\begin{aligned} a &\longrightarrow ad \\ b &\longrightarrow b \\ c &\longrightarrow cdbadce \\ d &\longrightarrow cdabdce \\ e &\longrightarrow cdadbce \end{aligned}$$

k fulfills condition (0) but $k(abac) = adbadcdbadce$, which is not square-free. In view of Corollaries 1.3 and 1.5, let q be a square-free homomorphism from $\mathscr{F}_5$ into $\mathscr{F}_3$. Finally, let p be the restriction of $q \circ k$ to $\mathscr{F}_3$. Then p fulfills condition (0) since k does and since q is square-free. However $p(abac)$ is not square-free.

THEOREM 1.7. *There is a square-free homomorphism from* $\mathscr{F}_\omega$ *into* $\mathscr{F}_3$.

Proof. It is only necessary to compose the homomorphisms provided by Corollaries 1.4, 1.3, and 1.5.

THEOREM 1.8. *On there letters, three are* 2^ω *square-free words of type* ω; *no two of which have common final segments.*

Proof. With ω letters it is easy to arrange 2^ω square-free words of type ω no two of which have common final segments. A collection H of sets is *almost disjoint* provided $M \cap N$ is finite whenever $M, N \in H$ and $M \neq N$. Let H be an almost disjoint collection of subsets of the alphabet such that $|H| = 2^\omega$. Such a collection has been constructed by W. Sierpinski [27]. For each $M \in H$ let W_M be a word of tpye ω listing M without repetitions. $\{W_M : M \in H\}$ is the collection of 2^ω square-free words no two of which have common final segments. Now let f be a square-free endomorphism from $\mathscr{F}_\omega$ into $\mathscr{F}_3$. f can be extended to words of type ω. Then $\{f(W_M) : M \in H\}$ is the desired collection of words. f is one-to-one since it is square-free. So $\{f(W_M) : M \in H\}$ has cardinality 2^ω. Now suppose $M, N \in H$ with $M \neq N$ and yet $f(W_M)$ and $f(W_N)$ have a common final segment. Say $M \cap N = \{c_0, \cdots, c_{n-1}\}$ and $M = \{c_0, \cdots, c_{n-1}\} \cup \{a_0, a_1, \cdots\}$ and $N = \{c_0, \cdots, c_{n-1}\} \cup \{b_0, b_1, \cdots\}$. Let $A_i = f(a_i)$ and $B_i = f(b_i)$ for all $i \in \omega$. Since $f(W_M)$ ann $f(W_N)$ have a common final segment, they must have a common final segment in which the image of no c_i occurs. So for some $j, k \in \omega$ such that $A_j = XA_j''$, where X may be empty we have

$$F = A_j'' A_{j+1} A_{j+2} \cdots$$
$$F = B_k B_{k+1} B_{k+2} \cdots$$

where F is the common final segment. Evidently, either some initial segment of some B_ℓ is a final segment of some A_i or else some initial segment of some A_i is a final segment of some B_ℓ. So $f(a_i b_\ell a_i)$ is not square-free which is contrary to the square-freeness of f and $a_i b_\ell a_i$ for all $i, \ell \in \omega$. So $f(W_M)$ and $f(W_N)$ have no common final segments. The proof is complete.

Alfred Manaster pointed out to us that using the work of Kakutani, and Morse and Hedlund (see Gottschalk and Hedlund [11] p. 109) it was easy to prove that there are 2^ω square-free words of type ω on three letters.

Now we take up the investigation of kth power-free homomorphisms. The analog of Theorem 1.0 is the following theorem.

THEOREM 1.9. *Let M and N be alphabets and let h be a homomorphism from $\mathscr{F}_M$ into $\mathscr{F}_N$. Let $k > 2$. If*

(0) *(W) is kth power-free whenever W is a kth power-free word on M with length no greater than $k+1$.*

(1) *$a = b$ whenever $a, b \in M$ with $h(a)$ a subword of $h(b)$.*

(2) *If $a, b, c \in M$ and $X h(a) Y = h(b)h(c)$, where X and Y may be empty, then either X is empty and $a = b$ or else Y is empty and $a = c$.*

Then h is kth power-free.

The only essential difference between the proof of this theorem and the proof of Theorem 1.0 lies in the proof of the claim. Here condition (2) is used to establish the claim. The details of this proof are omitted.

There are square-free maps which are not cube-free. The endomorphism of $\mathscr{F}_4$ induced by

$$\begin{aligned} a &\longrightarrow abacbab \\ b &\longrightarrow cdabcabd \\ c &\longrightarrow cdacabcbd \\ d &\longrightarrow cdacbcacbd \end{aligned}$$

is square-free according to Theorem 1.0 but $h(a^2) = abacbababacbab = abac(ba)^3cbab$, and so h is not cube-free. On the other hand, some square-free homomorphisms are kth power-free for all $k > 2$.

THEOREM 1.10. *Let h be a homomorphism from $\mathscr{F}_M$ into $\mathscr{F}_N$, where M and N are alphabets. If*

(0) *h is square-free,*

(1) *$a = b$ whenever $a, b \in M$ with $h(a)$ a subword of $h(b)$, and*

(2) *No proper initial segment of $h(a)$ is a final segment of $h(a)$, for all $a \in M$,*

then h is kth power-free for all $k > 1$.

Proof. Let $e_0, e_1, \cdots, e_n \in M$ and $E_0 = h(e_0), \cdots,$ and $E_n = h(e_n)$. Suppose $k > 2$, and $h(e_0 e_1 \cdots e_n) = XY^kZ$ where X, Y, and Z are words on N with Y nonempty. We will show that $e_0 e_1 \cdots e_n$ is not kth power-free. Let $E_0 = XE_0''$ and $E_n' = E_nZ$ where E_0'' and E_n' are nonempty. (If it were not so we could simplify $e_0 e_1 \cdots e_n$.) So $Y^k = E_0''E_1 \cdots E_n'$. Now let ℓ be the length of Y. Then any subword of Y^k of length 2ℓ is a square; hence the length of E_i is less than 2ℓ for all i with $0 < i < n$. Moreover the length of E_i is no greater than ℓ, for all i with $0 < i < n$, since otherwise E_i would have a proper initial segment that would be a final segment. From the hypotheses and $k > 2$ it follows that Y is not a subword of either E_0'' or E_n'. This means that there are $j_0, j_1, \cdots, j_{k-2}$ all between 0 and n with $E_{j_i} = E_{j_i}'E_{j_i}''$ and E_{j_i}'' nonempty for all $i < k - 1$ and

$$Y = E_0''E_1 \cdots E_{j_0}'$$
$$Y = E_{j_0}''E_{j_0+1} \cdots E_{j_1}'$$
$$\vdots$$
$$Y = E_{j_{k-2}}''E_{j_{k-2}+1} \cdots E_n' .$$

Now h fulfills the conditions of Theorem 1.0 and so the claim proved there holds. Hence $E_0'' = E_{j_0}'' = E_{j_1}'' = \cdots = E_{j_{k-2}}''$ and $E_{j_0}' = E_{j_1}' = \cdots = E_n'$ and $E_i = E_{j_0+i}$ for all $i < j_0, \cdots, E_{j_{k-3}+i} = E_{j_{k-2}+i}$ for all $i < j_0$. Therefore $e_0e_1 \cdots e_n = e_0e_1 \cdots e_{j_0-1}(e_{j_0}e_1 \cdots e_{j_0-1})^{k-2}e_1 \cdots e_{j_0-1}e_n$. But $h(e_0e_{j_0}e_n)$ is not square-free, so $e_0 = e_{j_0}$ or $e_{j_0} = e_n$. In either case $e_0 \cdots e_n$ is not kth power-free and the theorem is established.

COROLLARY 1.11. *There is a homomorphism from $\mathscr{F}_\omega$ into $\mathscr{F}_3$ which is* kth *power-free for all* $k \geqq 2$.

Proof. Each of the homomorphisms involved in the proof Theorem 1.7 fulfills condition (2) of Theorem 1.10.

COROLLARY 1.12. *There is a cube-free homomorphism from $\mathscr{F}_3$ into $\mathscr{F}_2$.*

Proof. The homomorphism induced by

$$a \longrightarrow abaabbaababaabaabba$$
$$b \longrightarrow abaababaabaabbaababaabba$$
$$c \longrightarrow abaababaabbaababaabaababaabba$$

fulfills the conditions of Theorem 1.9 and hence in cube-free.

THEOREM 1.13. *There is a cube-free homomorphism from $\mathscr{F}_\omega$ into $\mathscr{F}_2$.*

Proof. The composition of the homomorphisms established by Corollaries 1.11 and 1.12 suffices.

COROLLARY 1.14. *There is a cube-free endomorphism of $\mathscr{F}_2$.*

Proof. The endomorphism induced by

$$a \longrightarrow abaabbaababaabaabba$$
$$b \longrightarrow abaababaabaabbaababaabba$$

fulfills all the conditions of Theorem 1.9.

COROLLARY 1.15. (A. Thue [29]) *xxx is avoidable on a two letter alphabet.*

Proof. This result can be regarded as a corollary of either Corollary 1.14 (similar to Corollary 1.2) or Theorem 1.13.

THEOREM 1.16. *On two letters, there are 2^ω cube-free words of type ω no two of which have common final segments.*

The proof of this theorem does not differ in any important way from the proof of Theorem 1.8 and we omit the proof. Again, the construction of Kakutani mentioned in Gottschalk add Hedlund [11] gives 2^ω cube-free words of type ω on two letters.

2. Maximal kth power-free words. Recall that a word W on an alphabet N is *maximal* kth *power-free on* N provided W is kth power-free and both aW and Wa fail to be kth power-free, for all $a \in N$. In [2] S. E. Arshon constructed maximal square-free words on every finite alphabet and S. R. Li in [19] has characterized maximal square-free words. This section is devoted to demonstrating the following theorem.

THEOREM 2.0. *For any natural numbers n and k, any alphabet N with $|N| = n$, and any* kth *power-free word W on N, W is a subword of some maximal* kth *power-free word on N.*

Proof. Call a kth power-free word U on N *right maximal* (kth *power-free on* N) if Ua fails to be kth power-free, for all $a \in N$. By considering symmetry, it is enough to show that every kth power-free word on N is an initial segment of some right maximal word on N. A kth power-free word U on N is *contrary* (*for* kth *power-freeness on* N) provided U is not an intital segment of any right-maximal word. The word W' is a conjugate of W if W and W' are substitution instances of one another. W is *vulnerable* (*for* kth *power-freeness on* N) if W is a word on N, and given any contrary word U there is a conjugate W' of W and a word X of positive length such that UXW' is kth power-free on N. The proof will be complete if a right maximal vulnerable word can be produced.

In order to avoid trivial cases, let $n, k \geqq 2$ with not both $n = 2$ and $k = 2$. There are two cases.

Case I. $k = 2$ and $n > 2$.

Vulnerable words are produced according to the following rules.

Rule 0. Every letter is vulnerable.

Rule 1. If W is vulnerable and x is a letter such that no proper initial segment of Wx is a final segment of Wx, then Wx is vulnerable.

Rule 1 is a special case of the next rule.

Rule 2. Suppose W is vulnerable and x is a letter. If $Wx = BXB$ with $|B| > 0$ implies XBX is not square-free and Wx is square-free then Wx is vulnerable.

Proof. Suppose Wx is not vulnerable. Let U be a contrary word such that $UXWx$ is not square-free for any word X. Pick X such that $|X| > 2(|U| + |W|)$ and UXW is square-free. There is a word Z such that ZZ is a final segment of $UXWx$. If Wx is a final segment of Z, then there is a word V with $|V| > 0$ and $UVWx$ square-free, since X is so long. This conclusion violates the choice of U, so Wx is a final segment of ZZ but Z is a proper final segment of Wx. That is, there are words A and B such that $AWx = ZZ$ and $Wx = BZ$. So $AB = Z$ and then $Wx = BAB$ and $Z = ABAB$. Hence ABA is a subword of UXW. Thus ABA must be both square-free and not square-free. Consequently Wx is vulnerable and rule 2 is verified.

Let $N = \{a_0, a_1, a_2, \cdots, a_{n-1}\}$. Set $T_3 = a_0a_1a_2a_0a_2a_1a_0a_1a_2a_0a_2$. For $n \geqq 3$, let $T_{n+1} = T_na_nT_n$. Observe that for all $n \geqq 3$, T_n is a right maximal square-free word on the n letter alphabet.

Claim. a_2T_n is vulnerable for all $n \geqq 3$.

Proof. $a_2a_0a_1$ is vulnerable according to rules 0, 1, 1 in that order. Suppose $a_2a_0a_1a_2$ is not vulnerable. Let U be a contrary word such that UXW fails to be square-free whenever W is conjugate to $a_2a_0a_1a_2$. Pick X with $|X| > 2(|U| + 4)$ and $UXa_2a_0a_1$ square-free. Now $UXa_2a_0a_1a_2$ is not square-free and since X is so long a_0a_1 must be a final segment of X (otherwise $a_2a_0a_1a_2$ is a subword of X). But then $UYa_0a_1a_2a_0a_1 = UXa_2a_0a_1$ for some Y and $UYa_0a_1a_2a_0$ is square-free. However $a_0a_1a_2a_0$ is a conjugate of $a_2a_0a_1a_2$ and so $a_2a_0a_1a_2$ is vulnerable; $a_2a_0a_1a_2a_0$ is vulnerable by a similar argument; $a_2a_0a_1a_2a_0a_2$ is vulnerable by rule 2. $a_2a_0a_1a_2a_0a_2a_1a_0a_1$ is vulnerable by rule 1 (three applications). $a_2a_0a_1a_2a_0a_2a_1a_0a_1a_2a_0a_2$ is vulnerable by three applications of rule 2. So a_2T_3 is vulnerable.

For the sake of induction, suppose that a_2T_n is vulnerable. $a_2T_na_n$ is vulnerable by rule 1. Suppose XBx is an initial segment of T_n and $a_2T_na_nXB$ is vulnerable. Assume $BxY = a_2T_n$ and $T_n = TBxW$. Since a_2T_n is square-free and T_n does not begin with a_2 it follows that there must be a word V of positive length such that $a_2T_n = BxVBxW$. Let $a_2C = B$. Then $T_n = CxVa_2CxW$ and $a_2T_na_nXBx = a_2CxVa_2CxWa_nVa_2Cx$. Finally $Va_2CxWa_nVa_2CxVa_2CxWa_nV$ fails to be square-free. Consequently rule 3 applies and $a_2T_na_nXBx$ is vulnerable. In this way $a_2T_na_nT_n(=a_2T_{n+1})$ can be shown to be vulnerable and the claim is established.

Since a_2T_n is both vulnerable and right-maximal on the n letter alphabet, where $n \geqq 3$, Case I is finished.

Case II. $k > 2$ and $n \geqq 2$.

Vulnerable words are produced according to the following rules.

Rule 0. Every letter is vulnerable.

Rule 1'. If W is vulnerable, x is a letter, Wx is kth power-free, and $W \neq Y'(xY)^{k-1}$ for any nonempty Y where Y' is a final segment of Y, then Wx is vulnerable.

Proof. Suppose not. Let U be a contrary word such that for all X and all conjugates V' of Wx, UXV' fails to be kth power-free. Pick X with $|X| \geqq k(|U| + |Wx|)$ such that UXW is kth power-free. (Exchanging U for one its conjugates if necessary). So there is a nonempty word Z such that z^k is a final segment of $UXWx$. Since Wx is kth power-free it must be a proper final segment of z^k. Since $W \neq Y'(xY)^k$ where Y is any word with final segment Y', it follows that Wx is a final segment of Z^{k-1}. Since X is so long there must be a word V of positive length such that $UVWx$ is an initial segment of UXW. Hence $UVWx$ is kth power-free violating the choice of U. Rule 1' is established.

Let $T_1 = a_0^{k-1}$ and $T_{n+1} = (a_nT_na_n)^{k-1}a_nT_n$ for all $n > 1$. Then T_n is a right maximal kth power-free word on the n letter alphabet $\{a_0, a_1, \cdots, a_{n-1}\}$. Theorem 2.0 is established by the following claim.

Claim. T_n is vulnerable, for $n \geqq 2$.

Proof. It suffices to show that $a_{n-2}T_n$ is vulnerable. After the initial application of rule 0, rule 1′ always applies-trivially if $n > 2$ since $a_{n-2}a_{n-1}$ is an initial segment of $a_{n-2}T_n$ which occurs nowhere else in this word. Suppose $n = 2$. $a_0T_2 = a_0(a_1a_0^{k-1}a_1)^{k-1}a_1a_0^{k-1}$. a_0a_1 occurs k times in a_0T_2 and $a_0a_1a_0$ is an initial segment of a_0T_2 which occurs nowhere else in this word. With these two observations in mind it is easy to see that one use of rule 0 followed by repeated use of rule 1′ will yield a_0T_2. Hence T_2 is vulnerable.

This completes the proof of the theorem.

COROLLARY 2.1. *For all $k, n \geqq 2$ with $k > 2$ or $n > 2$, there are infinitely many maximal* kth *power-free words on the n letter alphabet.*

Proof. The corollary follows from Theorem 2.0 in view of Thue's theorems, Corollaries 1.2 and 1.15 above.

3. Avoidable sets of words. In § 1, we saw that $\{x^2\}$ is avoidable on $\{a, b, c\}$, while $\{x^3\}$ is avoidable on $\{a, b\}$. A. Thue found these results in 1906. Evidently $\{x\}$ is not avoidable and some reflection reveals that $\{xyx\}$ is not avoidable. In this section we establish results about avoidable sets of words. Our principal tools are the canonical endomorphisms defined below. All these are endomorphisms on semigroups of the form $\mathscr{F}_{4n}$ for some $n > 0$. Let $n > 0$. We represent the alphabet with $4n$ letters as

$$\{a_i\colon 0 \leqq i < n\} \cup \{b_i\colon 0 \leqq i < n\} \cup \{c_i\colon 0 \leqq i < n\} \cup \{d_i\colon 0 \leqq i < n\}\ .$$

It is clear what is intended by the *index* of a letter.

DEFINITION 3.0. Let $n > 0$. The *canonical endomorphism* h on the alphabet with $4n$ letters is the endomorphism induced by

$$\begin{aligned} a_i &\longmapsto a_jb_kd_l \\ b_i &\longmapsto a_jc_kd_l \\ c_i &\longmapsto a_jd_kd_l \\ d_i &\longmapsto b_jb_kc_l \end{aligned}$$

where $j = 3i \bmod(n)$, $k = j + 1 \bmod(n)$, and $l = k + 1 \bmod(n)$. In most cases $3 \nmid n$ and then each letter is assigned a distinct word by h.

Consider $\{h^p(a_0)\colon p \in \omega\}$, where h is canonical. All words which belong to this set, with the exception of a_0, result from the concatenation of words of the form

$$a_j b_k d_l,\ a_j c_k d_l,\ a_j d_k d_l, \quad \text{and} \quad b_j b_k c_l$$

where j, k, and l are subject to the constraints in Definition 3.0. If $3 \nmid n$, and $n > 2$ then there are $4n$ such words. We call these *fundamental words.*

The following lemma is easy to verify.

LEMMA 3.1. *Consider the alphabet with $4n$ letters where $n > 0$. Let x and y be letters. Then xy can occur as a subword in at most one fundamental word; moreover, if A and B are fundamental and x is the right most letter in A while y is the left most letter of B, then xy is not a subword of any fundamental word.*

A word xy of length two which fulfills the hypothesis of the moreover clause in Lemma 3.1 is called a *border word.*

LEMMA 3.2. *If h is the canonical endomorphism for an alphabet with $4n$ letters where $3 \nmid n$ and $n > 3$, then h is square-free.*

Proof. Lemma 3.1 makes it easy to check the conditions of Theorem 1.0.

LEMMA 3.3. *If h is the canonical endomorphism for an alphabet with $4n$ letters where $3 \nmid n$ and $n > 3$, and $p > 0$ and xy is a two letter subword of $h^p(a_0)$, then modulo n the index of y is one greater than the index of x.*

The proof is by induction on p.

It follows from Lemma 3.3 that scanning the indices of $h^p(a_0)$ one would see $0, 1, 2, \cdots, n-1, 0, 1, 2, \cdots, n-1, 0, \cdots$ etc.

DEFINITION 3.4. Let h be the canonical endomorphism for an alphabet with $4n$ letters where $n \geqq 3$. Let W be a word with $W = UFV$ and $h(W) = XCY$. F is a *father* of C provided $|X| \leqq |h(U)| + 1$, $|Y| \leqq |h(V)| + 1$, and $|C| \leqq |h(F)|$.

In the definition above, F is a father of C if each letter of F contributes at least two letters to C under the map h. Notice that F is a father of C whenever there are words W, U, V, X, and Y which fulfill the definition.

LEMMA 3.5. *If h is the canonical endomorphism for an alphabet with $4n$ letters where $3 \nmid n$ and $n > 3$, then every word has at* *most one father, (i.e., fathers are unique.)*

This lemma is an immediate application of Lemma 3.1. A subword of $h^p(a_0)$ (where $3 \nmid n$ and $n > 3$) is fatherless if and only if it is of length 1 or else it is a border word.

DEFINITION 3.6. A word W (on any alphabet) is *scrambled* provided

(i) if a letter x occurs in W, then it occurs at least twice in W, and

(ii) if x and y are distinct letters occurring in W, then both xy and yx are subwords of W.

THEOREM 3.7. *The set of all scrambled words on a denumerable alphabet is avoidable on the twenty letter alphabet.*

Proof. We pick $n = 5$ and, letting h be the canonical endomorphism on the alphabet with $4 \cdot 5$ letters, we will show that $\{h^p(a_0): p \in \omega\}$ avoids each scrambled word.

Suppose, to the contrary, that W is scrambled and not avoided by $\{h^p(a_0): p \in \omega\}$. Let $W = e_0 \cdots e_m$ where $e_0, \cdots, e_m$ are letters. Then there is some substitution instance $W^* = e_0^* e_1^* \cdots e_m^*$ of W such that for some p, W^* is a subword of $h^p(a_0)$. Pick p as small as possible so that each of $e_0^*, e_1^*, \cdots, e_m^*$ has a father. Since fathers are unique (Lemma 3.5) let F_0 be the father of e_0^*, F_1 be the father of $e_1^*, \cdots,$ and F_m be the father of e_m^*. Then $F_0 F_1 \cdots F_m$ is a subword of $h^{p-1}(a_0)$ and it is also substitution instance of W. At least one of $F_0, F_1, \cdots, F_m$ is fatherless.

Claim 0. If F_i is fatherless, then both F_{i-1} and F_{i+1}, where they exist, have fathers.

Proof of claim 0. Without loss of generality we assume that F_i and F_{i+1} are both fatherless and seek a contradiction. Since W is scrambled it follows that $F_i F_{i+1}$ and $F_{i+1} F_i$ are both subwords of $h^{p-1}(a_0)$. Since both F_i and F_{i+1} are fatherless, they both have length no more than two. Examination of the indices involved reveals a violation of Lemma 3.3.

Claim 1. Only F_0 and F_m can be fatherless.

Proof of claim 1. Suppose F_i is fatherless with $0 < i < m$. By claim 0 F_{i-1} and F_{i+1} have fathers, so $F_{i-1} \neq F_i \neq F_{i+1}$. Let j be the index of the last letter in F_{i-1} and k be the index of the first symbol in F_{i+1}. Then $k \equiv j + 3 \bmod(5)$ or $k \equiv j + 2 \bmod(5)$, since $1 \leqq |F_i| \leqq 2$ $F_{i-1} \neq F_{i+1}$ according to Lemma 3.3. Since W is scrambled $F_{i-1}F_{i+1}$ is a subword of $h^{p-1}(a_0)$. Hence $k \equiv j + 1 \bmod(5)$. This is a contradiction.

Since W is scrambled $F_0 = F_m$, sinec e_0 occurs at least twice in W. $m \neq 1$ since h is square-free. If $0 < j < m$, then F_0F_j must occur in $F_0 \cdots F_m$ since W is scrambled. Therefore $F_1 = F_j$ for all $j = 1, \cdots, m - 1$. Hence $W = e_0e_1^{m-1}e_0$ and $m \geqq 3$. But this violates the square-freeness of $h^{p-1}(a_0)$. So $\{h^p(a_0): p \in \omega\}$ avoids W and the theorem is proven.

DEFINITION 3.8. Let $n \in \omega$ with $3 \nmid n$ and $n > 3$. The *ancestry relation* is the transitive closure of the fatherhood relation on the alphabet with $4n$ letters. That is, U is an *ancestor* of W provided for some $k \in \omega$ there are words $V_0, V_1, \cdots, V_{k-1}$ with $U = V_0$, $V_k = W$ and V_j is the father of V_{j+1} for all j with $0 \leqq j < k$.

The ancestry relation is a well-founded partial order.

DEFINITION 3.9. Let $n \in \omega$ with $3 \nmid n$ and $n > 3$. Let W be a word on the alphabet with $4n$ letters. The *lineage* of W is the number of ancestors of W.

DEFINITION 3.10. Let $k \in \omega$ and W be a word. W has *mesh* k provided whenever x is a letter and V is a word in which x does not occur with $|V| > k$, then xVx is not a subword of W.

DEFINITION 3.11. W is a *doubled* word provided every letter which occurs in W occurs at least twice in W.

THEOREM 1.12. *Let $k \in \omega$. The set of all doubled words of mesh k on a denumerable alphabet is avoidable on an alphabet with no more than $8k + 16$ letters.*

Proof. Let n be a number with $3 \nmid n$, $n > 3$, and $n > 2k$. Let h be the canonical endomorphism on the alphabet with $4n$ letters. Let W be a double word of mesh k. We argue that $\{h^p(a_0): p \in \omega\}$ avoids W.

Suppose not. As in the previous proof, let $W = e_0e_1 \cdots e_m$ and pick $p \in \omega$ and a substitution instance $W^* = e_0^*e_1^* \cdots e_m^*$ of W which is also a subword of $h^p(a_0)$.

Claim. For all $i = 0, \cdots, m$ there is some $j \leqq m$ such that e_j^* has greater lineage than e_i^*.

Proof of the claim. Since W is a doubled word of mesh k, there is $l = 0, \cdots, m$ with $1 < |i - l| \leqq k$ and $e_i = e_l$. For convenience suppose $i < l$. Suppose that for all q with $i < q < l$, the lineage of e_q^* is no greater than that of e_i^*. Let G be the fatherless ancestor of e_i^* and let r be the number of ancestors of e_i^*. Then GHG is a subword of $h^{p-r}(a_0)$ where $|H| \leqq 2(k-2) < 2k$ since in $h^{p-r}(a_0)$ each e_q^* has no ancestors (though possibly some "proposed fathers" of length one are available) or else a fatherless ancestor. Now let t be the index of the final symbol in G. In view of Lemma 3.3 either $t \equiv t + |H| + 1 \bmod n$ or else $t \equiv t + |H| + 2 \bmod n$, which violates $|H| + 2 \leqq 2k - 2 < 2k < n$. So there must be q with $i < q < l$ and the lineage of e_q^* is greater than the lineage of e_i^*. The claim is proven and with it, the theorem.

Theorems 3.7 and 3.12 reveal that certain sets of words involving infinitely many letters can be avoided on finite alphabets.

THEOREM 3.13. *Let N be an alphabet with n letters. $\{W\colon |W| \geqq 2^n$ and W is a word on $N\}$ is avoidable.*

Proof. The proof depends on the following lemma which is easily established by induction on n.

LEMMA 3.14. *If $|N| = n$ and W is a word on N with $|W| \geqq 2^n$, then W has a doubled subword U with $|U| \leqq 2^n$.*

Now for each W on N with $|W| \geqq 2^n$ pick a U_W according to Lemma 3.14. By Theorem 3.12 $\{U_W\colon W$ is a word on N with $|W| \geqq 2^n\}$ is avoidable, say by the infinite set $\mathscr{F}$ of words on some finite alphabet. So $\mathscr{F}$ avoids $\{W\colon W$ is a word on N with $|W| \geqq 2^n\}$ since U_W is a subword of W. The proof is complete.

LEMMA 3.15. *If $\mathscr{F}$ is an set of arbitrarily long words on the alphabet M and G is a set of arbitrarily long words on the alphabet N, then there is a set $\mathscr{H}$ of words on the alphabet $M \times N$ such that if $\mathscr{F}$ avoids W and $\mathscr{G}$ avoids U, then $\mathscr{H}$ avoids both W and U.*

Proof. Without loss of generality, we assume both $\mathscr{F}$ and $\mathscr{G}$ are closed under the formation of subwords. Let $A = a_0 \cdots a_{n-1}$, $B = b_0 \cdots b_{n-1}$ be words of length n respectively on M and N. Define $A \times B = (a_0, b_0) \cdots (a_{n-1}, b_{n-1})$. So $A \times B$ is a word on $M \times N$. Define $\mathscr{H} = \{A \times B\colon A \in \mathscr{F}, B \in \mathscr{G}$, and $|A| = |B|\}$. Notice that A and B are both substitution instances of $A \times B$. So $\mathscr{H}$ must avoid all words avoided by either $\mathscr{F}$ or $\mathscr{G}$.

COROLLARY 3.16. *Let N be a finite alphabet.* $\{W$: *W is a word on N and* $\{W\}$ *is avoidable*$\}$ *is avoidable.*

Proof. By Theorem 3.13, let $\mathscr{F}$ be a set which avoids $\{W$: W is a word on N with $|W| \geqq 2^n\}$. Only finitely many singletons remain. Let $\mathscr{G}_W$ be a set avoiding W provided $\{W\}$ is avoidable and $|W| < 2^n$. According to Lemma 3.15 (or more properly its obvious inductive extensions), $\{W$: W is a word on N with $\{W\}$ avoidable$\}$ is avoidable.

LEMMA 3.17. *Let N be a finite alphabet and $\mathscr{F}$ be an infinite set of words on N. There is $k \in \omega$ and an infinite set $\mathscr{G}$ of words of mesh k on N such that every word avoided by $\mathscr{F}$ is avoided by $\mathscr{G}$.*

Proof. Proceed by induction on $|N|$.

Initial step. The lemma is immediate if $|N| = 1$.

Inductive step. Suppose the lemma is true for all alphabets with fewer than $|N|$ letters and nevertheless $\mathscr{F}$ is an infinite set of words on N such that for all $k \in \omega$ and all sets $\mathscr{G}$ of words on N of mesh k there is a word W avoided by $\mathscr{F}$ but not by $\mathscr{G}$. So there must be a letter $x \in N$ such that for every $k \in \omega$, there is a distinct word $U_k \in \mathscr{F}$ such that for some V on $N \sim \{x\}$ with $|V| > k$ we can conclude that xV or Vx is a subword of U_k. Let $\mathscr{F}' = \{Y$: Y is a word of $N \sim \{x\}$ and Y is a subword of U_k for some $k \in \omega\}$. Evidently $\mathscr{F}'$ is a set of words on $N \sim \{x\}$ which is infinite and if W is avoided by $\mathscr{F}$, then W is avoided by $\mathscr{F}'$, since each $Y \in \mathscr{F}'$ is a subword of some $U \in \mathscr{F}$. By the inductive assumption there is an infinite set $\mathscr{G}'$ of words on $N \sim \{x\}$ and some $k \in \omega$ such that each word in $\mathscr{G}'$ is of mesh k, and moreover every word avoided by $\mathscr{F}'$ (and hence every word avoided by $\mathscr{F}$) is also avoided by $\mathscr{G}'$. This is contrary to the selection of $\mathscr{F}$ and hence the inductive step is completed, establishing the lemma.

Our next objective is to provide an effective characterization of the collection of avoidable words.

DEFINITION 3.18. Let W be a word. The letter x is *free for* W provided x occurrs in W and for no $n \in \omega$ is it possible to find letters $e_0, \cdots, e_n$ and $f_0, \cdots, f_n$ such that all of the following are subwords of W:

$$\begin{array}{l} xe_0 \\ f_0e_0 \\ f_0e_1 \\ f_1e_1 \\ \vdots \\ f_ne_n \\ f_nx\ . \end{array}$$

If x is free for W, then W^x is the word obtained from W be deleting all occurrences of x.

LEMMA 3.19. *Let W be a word on N such that the letter x is free for W. There are subsets A and S of N such that*
(i) $x \in A \sim S$.
(ii) *If $y \in A$, $z \in N$ and yz is a subword of W, then $z \in S$.*
(iii) *If $y \in S$, $z \in N$ and zy is a subword of W, then $z \in A$.*

Proof. Define A and S as follows.

$y \in A$ if and only if for some $n \in \omega$ there are letters $e_0, \cdots, e_n$ and $f_0, \cdots, f_{n-1}$ such that each of the following are subwords of W:

$$\begin{array}{l} xe_0 \\ f_0e_0 \\ f_0e_1 \\ \vdots \\ f_{n-1}e_n \\ ye_n\ . \end{array}$$

$y \in S$ if and only if for some $n \in \omega$, there are letters $e_0, \cdots, e_n$ and $f_0, \cdots, f_n$ such that each of the following are subwords of W:

$$\begin{array}{l} xe_0 \\ f_0e_0 \\ f_0e_1 \\ \vdots \\ f_ne_n \\ f_ny\ . \end{array}$$

$x \notin S$ since x is free for W. Properties (ii) and (iii) are immediate from the definitions of A and S.

LEMMA 3.20. *If x is free for W and W^x is unavoidable, then W is unavoidable.*

Proof. Let $\mathscr{F}$ be any infinite set of words of mesh k (where $k \in \omega$) on any finite alphabet N. According to Lemma 3.18 it suffices to show that $\mathscr{F}$ does not avoid W.

Let $a \in N$ and $m = |W|$. Since a^m is a substitution instance of W it is safe to assume that a^m is not a subword of any word in $\mathscr{F}$. We may also assume without loss of generality that a is the first letter of every word in $\mathscr{F}$ and that if $U \in \mathscr{F}$ then there is $V \in \mathscr{F}$ such that $V = XUY$ where a is the first letter of Y.

Let $E_0, \cdots, E_t$ be a listing without repetitions of all the words on N of the form $a^j X$ where $0 < j < m$ and $|X| \leqq k$ and a does not occur in X. Let $M = \{e_0, \cdots, e_t\}$ be an alphabet with $t + 1$ letters and let $g: \mathscr{F}_M \to \mathscr{F}_N$ be the homomorphism induced by $g(e_i) = E_i$ for each $i \leqq t$. Now every $U \in \mathscr{F}$ can be represented in a unique way as a concatenation of members of $\{E_0, \cdots, E_t\}$ such that only the last word in the concatenation is permitted to be of the form a^j. Let U^* be the word on M corresponding to the word $U \in \mathscr{F}$ with respect to this representation under g. Let $\mathscr{F}^* = \{U^*: U \in \mathscr{F}\}$. Since W^x is unavoidable, W^x has a substitution instance V which is a subword of U^* for some $U \in \mathscr{F}$. Let

$$W = w_0 w_1 \cdots w_m ,$$
$$W^x = z_0 z_1 \cdots z_r, \quad \text{and}$$
$$V = Z_0 Z_1 \cdots Z_r .$$

Let A and S be sets satisfying conditions (i), (ii), and (iii) of Lemma 3.19. For each $i = 0, 1, \cdots, m$, define

$$W_i = \begin{cases} a & \text{if } \ w_i = x \\ T & \text{if } \ w_i \in A \cap S \text{ and } \ g(Z_i)a = aT \\ T & \text{if } \ w_i \in S \sim A \text{ and } \ g(Z_i) = aT \\ g(Z_i)a & \text{if } \ w_i \in A \sim S \text{ and } \ w_i \neq x \\ g(Z_i) & \text{otherwise} . \end{cases}$$

Claim. $W_0 W_1 \cdots W_m$ is a subword of $g(V)a$.

The claim may be established by induction on m. By noting that $W_0 W_1 \cdots W_m$ is a substitution instance of W and that $g(V)a$ is a subword of Ua, which is itself a subword of a member of $\mathscr{F}$, we arrive at the conclusion that W is not avoided by $\mathscr{F}$, as desired.

LEMMA 3.21. *If x and y are letters occurring in W, U is the word obtained from W by substituting x for y, and U is unavoid-*

able, then W is unavoidable.

This lemma follows immediately since U is a substitution instance of W. Say that *U is obtained from W by identification of letters* whenever the first two hypotheses of Lemma 3.21 hold. *W reduces to U* provided there are words $V_0, V_1, \cdots, V_{n-1}$ with $W = V_0$, $V = V_{n-1}$ and $V_{i+1} = V_i^x$ for some letter x free in V_i or V_{i+1} is obtained from V_i by identification of letters, for all i with $0 \leqq in - 1$.

THEOREM 3.22. *The word W is unavoidable if and only if W reduces to a word of length one.*

Proof. Suppose W reduces to a word of length one. Since words of length one are unavoidable, Lemmas 3.20 and 3.21 yield that W is unavoidable.

Now suppose W is unavoidable. Pick n with $3 \nmid n$ and $n > k^{k/2}$ where $k = |W|$. Since W is unavoidable there is some $p \in \omega$ and some substitution instance W^* of W such that W^* is a subword of $h^p(a_0)$ where h is the canonical endomorphism on $4n$ letters. Moreover, we choose p to be the smallest number such that each subword of W^* which represents a letter of W is fatherless or legitimate. Let x be a letter occurring in W and X be the subword of W^* representing x. A letter y occurring in $h^q(a_0)$, with $q < p$, is a *proto-ancestor* of x provided there is some occurrence of y in $h^q(a_0)$ not in any ancestor of any subword of W^* representing a letter of W such that the $p - q$-fold image of that occurrence contributes at least one letter to some occurrence of X representing x in W^*. Proto-ancestors, unlike fathers, "grandfathers", etc. are not unique and may depend on the particular position of x in W.

Let $x \sim y$ if and only if x and y occur in W and there is some $q \leqq p$ such that proto-ancestors or fatherless ancestors of both x and y occur in $h^q(a_0)$. $\approx$is used to denote the transitive closure of$\sim$. $\approx$is an equivalence relation. Let x be a letter occurring in W which is represented by a fatherless subword of W^* and let W^+ be the word obtained from W by deleting all the letters $\approx$-equivalent to x. Evidently, there is $r < p$ such that some substitution instance of W^+ is a subword of $h^r(a_0)$. Therefore all that remains to prove is that W^+ can be obtained fɹom W by identification of letters and the elimination of free letters.

Since W^* is a subword of $h^p(a_0)$, it is clear that some substitution instance W^0 of W is a subword of T, where T is a sufficiently long word of the kind

$$012 \cdots 4n - 1\ 012 \cdots 4n - 1\ 012 \cdots$$

(using the first $4n$ natural numbers as letters). Moreover, n has been chosen so large that the subwords of T representing letters equivalent to x have a total length less that n. Hence, some of the letters of T do not occur in any of the subwords representing letters of W equivalent to x. Let y be a letter of W such that $y \not\approx x$ and Y represents y in T. Replace each such Y of W^0 by Y $i+1 i+2 \cdots 4n-1 012 \cdots i$ where i is the left most symbol of Y. Let W' be the word obtained from W^0 in this way. Then W' is a substitution instance of W which is a subword of T. Let $j_0 < j_1 < \cdots < j_p$ be the letters which occur in the subwords of W^0 (and hence of W') which represent letters of W equivalent to x.

Let $\bar{W}$ result from W' by the elimination of all occurrences of j_0. If the subword j_0 of W' represents the letter y of W, then y is free in W for otherwise there are letters $u_0 v_0 \cdots$ such that

$$\begin{array}{c} yu_0 \\ v_0u_0 \\ v_0u_1 \\ v_1u_1 \\ \vdots \\ v_\pi y \end{array}$$

are all subwords of W. In this event there are subwords $U_0 V_0 \cdots$ of W' such that

$$\begin{array}{c} j_0U_0 \\ V_0U_0 \\ \vdots \\ V_\pi j_0 \end{array}$$

are all subwords of W' and hence of T. But since j_0 is the last letter in each V_i then $j_0 j_0$ would be a subword of T, which is impossible. On the other hand y and y' may be letters of W equivalent to x and represented by Y and Y' in W' such that Y and Y' become identical when j_0 is deleted. At any rate $\bar{W}$ is a substitution instance of some word obtained from W by deletion of a free letter and/or the identification of letters (or else it is still a substitution instance of W). Moreover $\bar{W}$ retains all of the properties of W' above, but on a smaller alphabet. After $p+1$ such steps only a substitution instance of W^+ remains; consequently W^+ can be obtained from W by the deletion of free letters and/or the identification of letters. Therefore W can be reduced to a single letter, as desired.

The word *abaca'badaba'ca'ba'* reduces to a single letter (the first step must be the identification of a and a'), yet it cannot be reduced to a single letter by a series of deletions of free letters.

4. Extensions and applications.

A. *Partitions of linear orders.*

Thue's theorem (Corollary 1.2) concerning square-free words can be construed as

I. ω can be partitioned into three sets such that no two adjacent intervals are partitioned in the same way. Another way to say this is

II. ω can be partitioned into three sets such that for all $i, j \in \omega$ with $j > 0$ there is $k \in \omega$ such that $i \leqq k < i + j$ and k and $k + j$ lie in different blocks of the partition.

With this in mind we make the following definition.

DEFINITION 4.0. Let α be an ordinal and let $\mathscr{P}$ be a partition of α. $\mathscr{P}$ is *square-free* provided for all $\beta, \gamma \in \alpha$ with $0 < \gamma$ and $\beta + \gamma + \gamma \leqq \alpha$ there exists $\delta \in \alpha$ such that $\beta \leqq \delta < \beta + \gamma$ and δ and $\delta + \gamma$ lie in different blocks of $\mathscr{P}$.

In a similar fashion we could define the notion of a cube-free partition of an ordinal and, at some cost in complexity, even the concept of a partition which avoids some set of words. In the previous section we investigated the latter notion for ordinals no larger than ω; little is so far known about it at ordinals beyond ω.

THEOREM 4.1[2]. *Let α be an ordinal with $\alpha < (2^\omega)^+$.*

(a) *There is a square-free partition of α into three pieces.*

(b) *There is a cube-free partition of α into two pieces.*

Proof. According to Theorem 1.8 there is a collection $\mathscr{S}$ of square-free words of type ω, on the letters a, b, and c such that $|\mathscr{S}| = 2^\omega$ and no two distinct members of $\mathscr{S}$ have common final segments. Let $L = \{\lambda: \lambda < \alpha$ and λ is a limit ordinal$\}$. Let f be a function mapping L one-to-one into $\mathscr{S}$. So $f(\lambda)$ is a square-free word of type ω. Now if $\beta \in \alpha$, there is a unique limit ordinal λ and a unique natural number n such that $\beta = \lambda + n$. Define $g(\beta)$ as the nth symbol of $f(\lambda)$. So $g: \alpha \to \{a, b, c\}$ and it induces a partition $\mathscr{P}$ of α into three pieces. To see that $\mathscr{P}$ is square-free let $\beta, \gamma \in \alpha$ with $\beta + \gamma + \gamma \leqq \alpha$ and $\beta = \lambda + n$ where λ is a limit

2. After this proof was found Jean Larson, Richard Laver and George McNulty proved that the theorem holds for all ordinals, not just those less than $(2^\omega)^+$. A note containing their proof has been submitted for publication.

ordinal and $n \in \omega$. In the case that $\gamma \in \omega$ the existence of the required δ follows from the square-freeness of $f(\lambda)$. In the case $\omega \leqq \gamma$, there is a limit ordinal $\mu \neq \lambda$ and a natural number m such that $\beta + \gamma = \mu + m$. The existence of the requisite δ now follows from the fact that $f(\mu)$ and $f(\lambda)$ have no common final segments. So (a) is proven and (b) can be established by a similar argument employing Theorem 1.16 in place of Theorem 1.8.

In the notion of a square-free partition of an ordinal, translations played a prominent role. So we are led to the next definition.

DEFINITION 4.2. A partition $\mathscr{P}$ of the set R of real numbers (Q of rational numbers) is *square-free* provided for all $r, s \in R (r, s \in Q)$ with $s > 0$ there is a $t \in R (t \in Q)$ such that $r \leqq t < r + s$ and t and $t + s$ lie in different blocks of $\mathscr{P}$.

THEOREM 4.3. *There is a square-free partition of $R(Q)$ into two pieces.*

Proof. The constructions for R and Q are similar, so we provide only the construction for R. Let $R^+ = \{r : r \in R \text{ and } r > 0\}$.

Give both R and $R \times R^+$ well orderings of type 2^ω. We define A_α and B_α by the following recursion for ordinals $\alpha < 2^\omega$. Let (r, s) be the αth pair in $R \times R^+$. Let t be the least (in the well ordering of R) member of $[r, r + s)$ such that $\{t, t + s\} \notin \bigcup_{\beta<\alpha} (A_\beta \cap B_\beta)$. If $t \notin \bigcup_{\beta<\alpha} A_\beta$, let

$$A_\alpha = \{t\} \cup \bigcup_{\beta<\alpha} A_\beta \quad \text{and} \quad B_\alpha = \{t + s\} \cup \bigcup_{\beta<\alpha} B_\beta \ .$$

Otherwise, let

$$A_\alpha = \{ts + \} \cup \bigcup_{\beta<\alpha} A_\beta \quad \text{and} \quad B_\alpha = \{s\} \cup \bigcup_{\beta<\alpha} B_\beta \ .$$

Observe that $|A_\alpha \cup B_\alpha| < 2^\omega$ for all $\alpha < 2^\omega$.

Finally let $A = \bigcup_{\alpha<2^\omega} A_\alpha$ and $B = \bigcup_{\alpha<2^\omega} B_\alpha$. Evidently $A \cap B$ is empty. For any fixed $r \in R$, $|\{(r, s) : s > 0\}| = 2^\omega$ and so $\{(r, s) : s > 0\}$ is unbounded in the well ordering of $R \times R^+$. Thus if r is the βth member of R we can pick $\alpha > \beta$ so that (r, s) is the αth member of $R \times R^+$. Consequently $r \in A_\alpha \cup B_\alpha \subseteq A \cup B$. Therefore $\{A, B\}$ is a partition of R and it is square-free by construction.

B. *Multidimensional versions of square-freeness.* Is it possible to color the lattice points of the plane (or in general n-dimensional space) with three colors such that no rectangle occurs adjacent to a copy of itself? Is it possible to color the points of the Euclidean plane (Euclidean n-dimensional space) with two colors such that no two adjacent rectangles are colored in the same way? The answers to both these questions is yes. The proof of Theorem 4.3 can be extended to handle the last question.

To manage the first question, we observe that the notion of a square-free map can be suitably extended to the n-dimensional case, where, however, individual letters are mapped to square (cubic, etc.) arrays of letters. Once more, the orbit of an individual letter can be used to obtain a cover (of the plane instead of the line) which is square-free provided the map used is not trivial. The following map based on the work of J. Leech [18] (see our remark after Corollary 1.2) yields a square-free coloring of the lattice points in the plane with three colors "a", "b", and "c".

$$a \longmapsto \begin{array}{ccccccccccccc}
b & c & a & c & b & a & c & a & b & c & a & c & b \\
c & a & b & a & c & b & a & b & c & a & b & a & c \\
a & b & c & b & a & c & b & c & a & b & c & b & a \\
c & a & b & a & c & b & a & b & c & a & b & a & c \\
b & c & a & c & b & a & c & a & b & c & a & c & b \\
a & b & c & b & a & c & b & c & a & b & c & b & a \\
c & a & b & a & c & b & a & b & c & a & b & a & c \\
a & b & c & b & a & c & b & c & a & b & c & b & a \\
b & c & a & c & b & a & c & a & b & c & a & b & b \\
c & a & b & a & c & b & a & b & c & a & b & a & c \\
a & b & c & b & a & c & b & c & a & b & c & b & a \\
c & a & b & a & c & b & a & b & c & a & b & a & c \\
b & c & a & c & b & a & c & a & b & c & a & c & b
\end{array}$$

$$b \longmapsto \begin{array}{ccccccccccccc}
c & a & b & a & c & b & a & b & c & a & b & a & c \\
a & b & c & b & a & c & b & c & a & b & c & b & a \\
b & c & a & c & b & a & c & a & b & c & a & c & b \\
a & b & c & b & a & c & b & c & a & b & c & b & a \\
c & a & b & a & c & b & a & b & c & a & b & a & c \\
b & c & a & c & b & a & c & a & b & c & a & c & b \\
a & b & c & b & a & c & b & c & a & b & c & b & a \\
b & c & a & c & b & a & c & a & b & c & a & c & b \\
c & a & b & a & c & b & a & b & c & a & b & a & c \\
a & b & c & b & a & c & b & c & a & b & c & b & a \\
b & c & a & c & b & a & c & a & b & c & a & c & b \\
a & b & c & b & a & c & b & c & a & b & c & b & a \\
c & a & b & a & c & b & a & b & c & a & b & a & c
\end{array}$$

$$\begin{array}{rccccccccccccc}
 & a & b & c & b & a & c & b & c & a & b & c & b & a \\
 & b & c & a & c & b & a & c & a & b & c & a & c & b \\
 & c & a & b & a & c & b & a & b & c & a & b & a & c \\
 & b & c & a & c & b & a & c & a & b & c & a & c & b \\
 & a & b & c & b & a & c & b & c & a & b & c & b & a \\
 & c & a & b & a & c & b & a & b & c & a & b & a & c \\
c \longmapsto & b & c & a & c & b & a & c & a & b & c & a & c & b \\
 & c & a & b & a & c & b & a & b & c & a & b & a & c \\
 & a & b & c & b & a & c & b & c & a & b & c & b & a \\
 & b & c & a & c & b & a & c & a & b & c & a & c & b \\
 & c & a & b & a & c & b & a & b & c & a & b & a & c \\
 & b & c & a & c & b & a & c & a & b & c & a & c & b \\
 & a & b & c & b & a & c & b & c & a & b & c & b & a
\end{array}$$

Each row (and each column) of any one of these three 13×13 arrays is one Leech's thirteen letter words. This observation makes it easy to check that each array in the orbit of a under this map is indeed square-free.

C. *An application to the Burnside problem for semigroups.*

THEOREM 4.4. *If Σ is a set of semigroup equations in no more than k variables such that $\sigma = \tau \in \Sigma$ implies $|\sigma|, |\tau| \geqq 2^k$, then the semigroup freely generated by $8 \cdot 2^k + 16$ generators with respect to* $\mathrm{Mod}(\Sigma \cup \{x(yz) = (xy)z\})$ *is infinite.*

REMARK. $\mathrm{Mod}(\Sigma \cup \{z(yz) = (xy)z\})$ is the class of all semigroups in which each equation in Σ is universally true.

Proof. Let $m = 8 \cdot 2^k + 16$. According to Theorem 3.13 there is an infinite set $\mathscr{F}$ of words on m letters which avoids $T = \{\sigma\colon$ there is τ with either $\sigma = \tau \in \Sigma$ or $\tau = \sigma \in \Sigma\}$. Let A be the set of all words on the m letter alphabet. For $U, W \in A$ define

$U \sim W$ if and only if $U = W$ or neither U nor W avoids every word in T. $\sim$ is a congruence relation on the free semigroup with m generators. $A/\sim$ is infinite, has m generators, and each equation in Σ is true in this quotient semigroup. So the theorem is established.

This proof is reminescent of an idea credited to R. P. Dilworth in Morse-Hedlund [21] (see also § 6 of J. Rhodes [26]).

D. *Periodic words.*

DEFINITION 4.5. Let $W = e_0e_1 \cdots e_{n-1}$ be a word on the alphabet $\{e_0, \cdots, e_{n-1}\}$. W has *period* p provided $e_i = e_j$ whenever $|i - j| = p$. W is *periodic* if W has period p for some p with $0 < p < |W|+2/1$.

A periodic word is never square-free. An infinite periodic word is not kth power-free for any $k \in \omega$. Suppose W is a periodic word of type ω with p the smallest positive period of W. Then any subword of W of length a multiple of kp is a kth power of some word. This leads us to the next definition.

DEFINITION 4.6. Let W be a periodic word of type ω with p the smallest positive period of W and let $k > 1$. W is *almost* kth *power-free* provided $p \mid |U|$ whenever U^k is a subword of W.

We remark that $(abca)^\omega = abcaabcaabca\cdots$ is a word with minimal period 4 that fails to be almost square-free. Some interesting facts concerning periodic words can be found in Ehrenfeucht and Silberger [8]. Here we observe that every periodic word is an initial segment of a periodic word of type ω that has exactly the same minimal positive period. Note also that a final segment of a periodic word of type ω is itself a periodic word with the same minimal positive period.

NOTATION. $\lfloor r \rfloor$ denotes the largest integer no greater than r, whenever r is a real number.

LEMMA 4.7. *Let* $|W| = n$ *and* g.c.d. $(m, p) = d$. *If* W *has periods* m *and* p *and* $n \geqq m + p$, *then* W *has period* d.

Proof. To avoid trivial and symmetric cases we assume $p < m = qp + r$ with $0 < r < p$. Since g.c.d.$(p, r) = d$, it will suffice to show that W has period r. Let $W = e_0e_1 \cdots e_{n-1}$ where $e_0, e_1, \cdots, e_{n-1}$ are letters. Pick i with $0 \leqq i < i + r < n$.

If $pq \leqq i$, then $e_i = e_{i-pq} = e_{i-pq+m} = e_{i+r}$. So suppose $i < pq$. Let $x = \lfloor i/p \rfloor$ and note that $q - x > 0$. If $i - px + m < n$, then $e_i = e_{i-px} = e_{i-px+m} = e_{i-px+m-(q-x)p} = e_{i+r}$. So we are done unless $n \leqq i-px+m$. Since $m + p \leqq n$ we obtain

$$m + p \leqq i - px + m\ .$$

So $p \leqq i - px = i - p\lfloor i/p \rfloor < p$, a contradiction.

THEOREM 4.8. *Let W be a word of type ω with minimal positive period p. W is almost square-free if and only if every subword of W with length p is square-free.*

Proof. We prove the harder implication. Suppose UU is a subword of W with $|U| = m$ as small as possible such that $p \nmid m$. UU has periods m and p. If $m \geqq p$, thed $2m = |UU| \geqq m + p$ and Lemma 4.7 then asserts that UU has period $d = \text{g.c.d.}\ (m, p) < p, m$. In this event UU contains a square VV with $|V| = d < m$ and $p \nmid d$, violating the minimality of $|U|$. So $m < p$.

Now suppose $p < 2m$. Then there a words U' and U'' such that $U = U'U''$, $|U'| = p - m > 0$, and $|U''| = 2m - p > 0$. Since p is a period of W, we have U'' is both an initial segment and a final segment of U. Consequently $U''U''$ is a subword of UU. Since $2m - p < m$ we again contradict the minimality of $|U|$ (observe that $p \nmid (2m - p)$ since $m < p < 2m$). So $2m \leqq p$. This means that some subword of W with length p must fail to be square-free. The proof is finished.

THEOREM 4.9. *Let $k > 2$ and let W be a word of type ω with smallest positive period p. W is almost kth power-free if and only if every subword of W of length $\lfloor kp - 1/k - 1 \rfloor$ is kth power-free.*

Proof. We are only concerned with the harder implication. So pick U with $|U| = m$ as small as possible subject to $p \nmid m$ and U^k is a subword of W. U^k has periods m and p. If $km \geqq m + p$, then by Lemma 4.7 U^k has period $d = \text{g.c.d.}(m, p)$. Since $p \nmid m$ we know that $p \nmid d$ but that V^k is an initial segment of U^k for some word V with $|V| = d$. Since $d < m$, this violates the minimality of $|U|$. Consequently $(k - 1)m < p$. So $m \leqq p - 1/k - 1$. Finally $m \leqq \lfloor p - 1/k - 1 \rfloor$ and so $km \leqq \lfloor kp - 1/k - 1 \rfloor$ and $|U^k|$ is bounded as desired.

EXAMPLE. $W = (((ab)^{k-2}a)^{k-1}ab)^{\omega}$ is a periodic word that is not almost kth power-free but all subwords of it shorter than $k\lfloor p - 1/k - 1 \rfloor$ are kth power-free, where p is the smallest positive period of W and $k > 2$.

COROLLARY 4.10. *Let h be a kth power-free homomorphism from $\mathscr{F}_N$ into $\mathscr{F}_M$, where N and M are alphabets. If W is a word on N of type ω with least positive period greater than 1 and if W is almost kth power-free, then $h(W)$ is almost kth power-free.*

Proof. Let p be the least positive period of W and let V be the word such that $|V| = p$ and $W = V^{\omega}$. Let $p' = |h(V)|$. Let q be the least positive period of $h(W)$. Clearly $q | p'$. Suppose $h(W)$ is not almost kth power-free. Then there is a word U with $U| = m$ and $p' \nmid m$ and U^k a subword of $h(W)$. According to Lemma 4.7 we may suppose that $m|p'$. So there are words X and Y such that $|X| + |Y| < p$ and U^k is a subword of $h(XV^{k-1}Y)$. Hence $XV^{k-1}Y$ is not kth power-free. But since V^ω is almost kth power-free, this is a contradiction and the corollary is established.

By means of this corollary and the results of section one it is now simple to construct periodic almost kth power-free words with arbitrarily large poriods on small alphabets.

5. Problems. We gather here problems which have arisen during our investigations and which we have not yet been able to resolve.

1. Theorems 1.0 and 1.9 provide useful sufficient conditions for a homomorphism to be kth power-free. Characterize kth power-free homomorphisms in a similar manner.

2. For every avoidable word W is there an endomorphism (of a large enough alphabet) which is "W-free"? [That is, is there an endomorphism h such that $h(U)$ avoids W whenever U avoids W?].

3. For an arbitrary avoidable word W determine the smallest alphabet on which W is avoidable.

4[3]. For an arbitrary ordinal α, determine the smallest cardinal κ such that α has a square-free (cube-free) partition $\mathscr{P}$ with $|\mathscr{P}| = \kappa$.

5. Is it effectively decidable whether an arbitrary word is avoidable on an n letter alphabet? (Here n is regarded as fixed.)

ACKNOWLEDGMENTS. We would like to thank Bernhard Ganter, Jerry Malitz, Alfred Manaster, Jan Mycielski, Jeff Remmel, and Heinrich Werner for encouragement and for interesting conversations about these topics. Don Young translated Arshon [2] for us. The University of San Diego and the University of South Carolina each provided generous amounts of computer time. In addition, the University of South Carolina was very helpful with travel funds during the preparation of this paper. Besides our home institutions, parts of the research for this paper were accomplished at Dartmouth College, die Technische Hochschule Darmstadt, and the University of California at San Diego.

Note added in proof. Jean Bersted has recently characterized square-free homomorphisms in a manner close to Theorem 1.0. This settles part of Problem 1 above. See Berstel's paper "Sur les mots sans carre definis par un morphisme" in 6th ICALP, Mauer, ed., Lecture Notes in Computer Science, vol. 71, 1979, Springer-Verlag, Berlin.

3. See footnote 2.

REFERENCES

1. S. I. Adjan, *Burnside groups of odd exponent and irreducible systems of group identities*, in "Word Problems", ed. W. W. Boone et al., North-Holland Publishing Company, Amsterdam, 1973.

2. S. E. Arshon, *Dokazatel'stov suščestvovanija n-značnyh beskonečnyh asimmetricnyh posledovatel'nostei*, Mat. Sb. (N. S.) **2** (44) 1937), 769–779.

3. J. L. Britton, *The existence of infinite Burnside groups*, in "Word Problems", ed. W. W. Boone et al., North-Holland Publishing Company, Amsterdam, 1973.

4. T. C. Brown, *Is there a sequence on four symbols in which no two adjacent segments are permutations of one another?*, Amer. Math. Monthly, **78** (1971), 886–888.

5. S. Burris and E. Nelson, *Embedding the dual of* Π_∞ *in the lattice of equational classes of semigroups*, Algebra Universalis, **1** (1971), 248–253.

6. S. Corcoran, W. Frank, and M. Maloney, *String theory*, J. Symbolic Logic, **39** (1974), 625–637.

7. F. Dekking, *On repetitions of blocks in binary sequences*, J. Combinatorial Theory, Ser. A., **20** (1976), 292–299.

8. A. Ehrenfeunct and D. Silberger, *Periodicity and unbordered segments of words*, prepublication copy.

9. R. Entringer, D. Jackson and J. Schatz, *On nonrepetitive sequences*, J. Combinatorial Theory, Ser. A, **16** (1974), 159–164.

10. A. A. Evdokimov, *Strongly asymmetric sequences generated by a finite number of symbols*, Soviet Math. Dokl., **9** (1968), 536–539.

11. W. Gottschalk and G. Hedlund, *Topological Dynamics*, Amer. Math. Soc. Colloquium Publications, vol. **36**, Providence, 1955.

12. D. Hawkins and W. Mientka, *On sequences which contain no repetition*, Math. Student, **24** (1956), 185–187.

13. G. Hedlund, *Remarks on the work of Axel Thue on sequences*, Nordisk Mat. Tidskr, **15** (1967), 147–150.

14. H. Hermes, *Semiotik, eine Theorie der Zerchengestalten als Grundlage fur Untersuchungen von formalizieten Sprachen*, Forschungen zur Logik und zur Grundlage der exackten Wissurschaften, n.s., no. 5, Leipzig, 1938.

15. J. Jezek, *Intervals in the lattice of varieties*, Algebra Universalis.

16. J. Justin, *Characterization of the repetitive commutative semigroups*, J. Algebra, **21** (1972), 87–90.

17. ———, *Généralization du théorème de van der Waerden sur les semigroupes répétitifs*, J. Combinatorial Theory, Ser. A, **12** (1972), 357–367.

18. J. Leech, *A problem on strings of beads*, Math. Gazette, **41** (1957), 277–278.

19. S. R. Li, *Annihilators in nonrepetitive semigroups*, prepublication copy.

20. M. Morse, *A solution of the problem of infinite play in chess*, Bull. Amer. Math. Soc., **44** (1938), 632.

21. M. Morse and G. Hedlund, *Unending chess, symbolic dynamics, and a problem in semigroups*, Duke Math. J., **11** (1944), 1–7.

22. V. L. Murskii, *Examples of varieties of semigroups*, (Russian), Mat. Zametki, **3** (1968), 663–670.

23. P. S. Novikov, *On periodic groups*, Dokl. Akad. Nauk SSSR, **127** (1959), 749–752.

24. P. S. Novikov and S. I. Adjan, *Infinite periodic groups*, Math. USSR IZV., **2** (1968), 209–236, 241–479, 665–685.

25. P. A. B. Pleasants, *Nonrepetitive sequences*, Proc. Comb. Philos. Soc., **68** (1970), 267–274.

26. J. Rhodes, *Algebraic Theory of Finite Semigroups*, in "Semigroups" ed. K. W. Folley, Academic Press, New York, 1969.

27. W. Sierpinski, *Sur une décomposition d'ensembles*, Monatsch. Math., **35** (1928), 239–248.

28. A. Tarski, *The Concept of Truth in Formalized Languages*, in "Logic, Semantics, and Metamathematics," Clarendon Press, Oxford, 1956.
29. A. Thue, *Über unendliche Zeichenreihen*, Norske Vid. Selsk. Skr., I Mat. Nat. Kl., Christiania, **7** (1906), 1–22.
30. A. Thue, *Über die gegenseitigen Lage gleicher Teile gewisser Zeichenreilren*, Norske Vid. Selsk. Skr., I Mat. Nat. Kl., Christiania, **1** (1912), 1-67.

Received February 7, 1978 and in revised form June 14, 1979.

UNIVERSITY OF SAN DIEGO
SAN DIEGO. CA 92110

UNIVERSITY OF COLORADO
BOULDER, CO 80302
AND
UNIVERSITY OF SOUTH CAROLINA
COLUMBIA, SC 29208

PACIFIC JOURNAL OF MATHEMATICS

Mathematical papers intended for publication in the *Pacific Journal of Mathematics* should be in typed form or offset-reproduced, (not dittoed), double spaced with large margins. Please do not use built up fractions in the text of the manuscript. However, you may use them in the displayed equations. Underline Greek letters in red, German in green, and script in blue. The first paragraph or two must be capable of being used separately as a synopsis of the entire paper. Please propose a heading for the odd numbered pages of less than 35 characters. Manuscripts, in triplicate, may be sent to any one of the editors. Please classify according to the scheme of Math. Reviews, Index to Vol. **39**. Supply name and address of author to whom proofs should be sent. All other communications should be addressed to the managing editor, or Elaine Barth, University of California, Los Angeles, California, 90024.

50 reprints to each author are provided free for each article, only if page charges have been substantially paid. Additional copies may be obtained at cost in multiples of 50.

The *Pacific Journal of Mathematics* is issued monthly as of January 1966. Regular subscription rate: $84.00 a year (6 Vols., 12 issues). Special rate: $42.00 a year to individual members of supporting institutions.

Subscriptions, orders for numbers issued in the last three calendar years, and changes of address should be sent to Pacific Journal of Mathematics, P.O. Box 969, Carmel Valley, CA 93924, U.S.A. Older back numbers obtainable from Kraus Periodicals Co., Route 100, Millwood, NY 10546.

PUBLISHED BY PACIFIC JOURNAL OF MATHEMATICS, A NON-PROFIT CORPORATION
Printed at Kokusai Bunken Insatsusha (International Academic Printing Co., Ltd.).
8-8, 3-chome, Takadanobaba, Shinjuku-ku, Tokyo 160, Japan.

Pacific Journal of Mathematics

Vol. 85, No. 2 October, 1979