

© World Scientific Publishing Company DOI: 10.1142/S0218196717500242

A family of formulas with reversal of high avoidability index

James Currie^{*}, Lucas Mol[†] and Narad Rampersad[‡]

Department of Mathematics and Statistics, University of Winnipeg 515 Portage Avenue, Winnipeg, Manitoba, Canada R3B 2E9 *j.currie@uwinnipeg.ca [†]l.mol@uwinnipeg.ca [‡]n.rampersad@uwinnipeg.ca

> Received 11 November 2016 Accepted 16 June 2017 Published

Communicated by M. Volkov

We present an infinite family of formulas with reversal whose avoidability index is bounded between 4 and 5, and we show that several members of the family have avoidability index 5. This family is particularly interesting due to its size and the simple structure of its members. For each $k \in \{4,5\}$, there are several previously known avoidable formulas (without reversal) of avoidability index k, but they are small in number and they all have rather complex structure.

 $Keywords\colon$ Pattern avoidance; pattern with reversal; formula with reversal; avoidability index.

Mathematics Subject Classification: 68R15

1. Preliminaries

Let Σ be a set of letters called *variables*. A *pattern* p over Σ is a finite word over alphabet Σ . A *formula* ϕ over Σ is a finite set of patterns over Σ . Instead of set notation, we usually use dot notation to denote formulas; that is, for $p_1, \ldots, p_n \in \Sigma^*$ we let

$$p_1 \cdot p_2 \cdot \ldots \cdot p_n = \{p_1, p_2, \ldots, p_n\}.$$

The elements of a formula ϕ are called the *fragments* of ϕ .

For an alphabet Σ , define the *reversed alphabet* $\Sigma^R = \{x^R : x \in \Sigma\}$, where x^R denotes the *reversal* or *mirror image* of variable x. A *pattern with reversal* is a pattern over $\Sigma \cup \Sigma^R$. A *formula with reversal* over Σ is a formula over $\Sigma \cup \Sigma^R$, i.e. a finite set of patterns with reversal over Σ .

For words over any alphabet A, we denote by $-^{R}$ the reversal antimorphism; if $a_{1}, a_{2}, \ldots, a_{n} \in A$, then

$$(a_1a_2\ldots a_n)^R = a_na_{n-1}\ldots a_1.$$

We say that a morphism $f : (\Sigma \cup \Sigma^R)^* \to A^*$ respects reversal if $f(x^R) = f(x)^R$ for all variables $x \in \Sigma$. Note that any morphism $f : \Sigma^* \to A^*$ extends uniquely to a morphism from $(\Sigma \cup \Sigma^R)^*$ that respects reversal.

Let p be a pattern (with reversal). An *instance* of p is the image of p under some non-erasing morphism (respecting reversal). A word w avoids p if no factor of w is an instance of p. Let ϕ be a binary formula (with reversal). We say that ϕ occurs in w if there is a non-erasing morphism h (which respects reversal) such that the h-image of every fragment of ϕ is a factor of w. In this case we say that ϕ occurs in w through h, or that w encounters ϕ through h. If ϕ does not occur in w then we say that w avoids ϕ . For example, the word singing avoids the pattern xyzyx, but encounters the formula $xy \cdot yx$ (through several morphisms including $x \mapsto in$ and $y \mapsto g$; and $x \mapsto ing$ and $y \mapsto ing$). The formula with reversal $xyx \cdot x^R \cdot y^R$ occurs in the word senses through the morphism defined by $x \mapsto se$ and $y \mapsto n$.

For a positive integer k, let A_k denote an alphabet on k letters. We say that formula ϕ is k-avoidable if there are infinitely many words of A_k^* which avoid ϕ ; or equivalently, if there is an ω -word \mathbf{w} over A_k such that every finite prefix of \mathbf{w} avoids ϕ . If ϕ is not k-avoidable we say that ϕ is k-unavoidable. We say that ϕ is avoidable if it is k-avoidable for some $k \in \mathbb{N}$; otherwise, we say that ϕ is unavoidable. Finally, the avoidability index of ϕ , denoted ind (ϕ) , is the least $k \in \mathbb{N}$ such that ϕ is k-avoidable if ϕ is avoidable, and is ∞ if ϕ is unavoidable.

An open question in pattern avoidance is whether patterns of arbitrarily high avoidability index exist. At the time of writing it is unknown whether any patterns of avoidability index strictly greater than 5 exist. Formulas have been important to the search for patterns with high avoidability index because of the following connection between patterns and formulas. For every formula, there is an associated pattern of the same avoidability index obtained by replacing every dot with a new distinct letter (this fact was first proven in [4], and a similar argument demonstrates the truth of this fact for formulas with reversal as well). Indeed, this was the primary reason that formulas were introduced, as they are somewhat easier to analyze than the associated patterns.

While the fact that the pattern xx has avoidability index 3 is a well-known classical result, the first known pattern of avoidability index 4, presented by Baker, McNulty, and Taylor in [2] arises from the much longer formula $ab \cdot ba \cdot ac \cdot ca \cdot bc$. Later, Clark [5] demonstrated that every avoidable formula on at most three variables is 4-avoidable, but found several formulas on four or more variables that have avoidability index 5. All of these formulas are rather long and complex.

Relatively little is known about the possible values of the avoidability index of patterns with reversal. Currie and Lafrance [6] found the avoidability index of every *binary* pattern with reversal; that is, every pattern on $\Sigma \cup \Sigma^R$ where $|\Sigma| = 2$.

Formulas with reversal of high avoidability index 3

In particular, they found that every avoidable binary pattern with reversal has avoidability index at most 3. Currie and Rampersad have shown that the growth of the number of binary words avoiding the formula xx^Rx is intermediate between polynomial and exponential [7], a surprising result which they have also shown to hold for the formula xxx^R [8]. These are the first known instances of such an intermediate growth rate in the context of pattern avoidance, and this suggests that patterns with reversal may be quite different from patterns in the usual sense (i.e. without reversal).

In this paper, we present an infinite family of formulas with reversal whose avoidability index is bounded between 4 and 5, and we show that several members of the family have avoidability index 5. This family is actually part of a larger family of formulas with reversal whose members are shown to have avoidability index between 4 and 7 in general (although we suspect that the true upper bound is 5). The simplicity of our examples makes their high avoidability index all the more surprising, as the previously known formulas (without reversal) of avoidability index 4 or 5 are quite complex.

For each $k \ge 1$, define

$$\psi_k = xy_1y_2\dots y_kx \cdot y_1^R \cdot y_2^R \cdot \dots \cdot y_k^R.$$

In Sec. 2, we show that $\operatorname{ind}(\psi_1) = 4$ and $\operatorname{ind}(\psi_2) = 5$. Then we move on to general bounds on the avoidability index of ψ_k for $k \ge 3$. In Sec. 3, we bound the avoidability index of ψ_k from above for all $k \ge 3$. We show that $\operatorname{ind}(\psi_{3k}) \le 5$ for $k \ge 1$, $\operatorname{ind}(\psi_{3k+1}) \le 6$ for all $k \ge 1$, and $\operatorname{ind}(\psi_{3k+2}) \le 6$ for all $k \ge 2$. The only remaining case is ψ_5 , for which we demonstrate $\operatorname{ind}(\psi_5) \le 7$. While we can use backtracking to show that $\operatorname{ind}(\psi_k) \ge 5$ for all $3 \le k \le 6$ (so in particular $\operatorname{ind}(\psi_3) = \operatorname{ind}(\psi_6) = 5$), this method becomes impractical for larger values of k. In Sec. 4, we present a general argument that shows $\operatorname{ind}(\psi_k) \ge 4$ for all $k \ge 7$.

2. The Avoidability Index of ψ_1 and ψ_2

This section is devoted to proving that $\operatorname{ind}(\psi_1) = 4$ and $\operatorname{ind}(\psi_2) = 5$. First of all, one can demonstrate that ψ_1 is 3-unavoidable and that ψ_2 is 4-unavoidable by using a standard backtracking algorithm. A longest word on three letters avoiding ψ_1 has length 14 while a longest word on four letters avoiding ψ_2 has length 45. It remains to show that ψ_1 is 4-avoidable and ψ_2 is 5-avoidable. In fact, we show that there are exponentially many words on four letters that avoid ψ_1 and exponentially many words on five letters that avoid ψ_2 . Below we define words that have a particular cyclic structure as they will be used in the constructions that follow.

Definition 2.1. Let $k \ge 1$ and let $w = w_1 w_2 \dots$ be a word in $\{1, \dots, k\}^*$. Define the (a_1, \dots, a_m) -cyclic w-word $c_m^{(w)}$ on distinct letters a_1, \dots, a_m by

$$c_m^{(w)} = a_1^{w_1} a_2^{w_2} \dots a_m^{w_m} a_1^{w_{m+1}} a_2^{w_{m+2}} \dots a_m^{w_{2m}} \dots$$

Any word isomorphic to $c_m^{(w)}$ is called an *m*-cyclic *w*-word.

We will find infinitely many words for which the corresponding 4-cyclic word avoids ψ_1 , and infinitely many words for which the corresponding 5-cyclic word avoids ψ_2 . Further, we will show that the growth of the number of these words is exponential in each case. We begin with a lemma that gives a condition on a word $w \in \{1, \ldots, k+1\}^*$ that is satisfied if and only if $c_m^{(w)}$ avoids ψ_k (with $m \ge k+2$). Hence this lemma allows us to determine whether a given m-cyclic w-word avoids ψ_k by considering w alone.

Lemma 2.2. Let $k \ge 1$ and $m \ge k+2$. If $w \in \{1, \ldots, k+1\}^*$ then the m-cyclic w-word $c_m^{(w)}$ avoids ψ_k if and only if for all $j \in \{1, \ldots, k\}$, w has no factor of the form $x'\alpha_1\alpha_2\ldots\alpha_j x''$, where

- $|\alpha_i| = 1$ for all $i \in \{1, \dots, j\}$, *i.e.* $\alpha_i \in \{1, \dots, k+1\}$;
- $\sum_{i=1}^{j} \alpha_i \ge k \text{ (in } \mathbb{Z});$
- |x'| = |x''| = n with $n \equiv m j \pmod{m}$; and if $x' = x'_1 \dots x'_n$ and $x'' = x''_1 \dots x''_n$ then $x'_1 \ge x''_1$, $x''_n \le x''_n$, and $x'_i = x''_i$ for all $i \in \{2, \ldots, n-1\}.$

Proof. First suppose that w has a factor u of the form $x'\alpha_1 \dots \alpha_i x''$ for some $j \in \{1, \ldots, k\}$ satisfying the conditions of the lemma statement. We have

$$c_m^{(u)} = a_1^{x_1'} a_2^{x_2'} \dots a_{m-j}^{x_n'} a_{m-j+1}^{\alpha_1} \dots a_m^{\alpha_j} a_1^{x_1''} a_2^{x_2''} \dots a_{m-j}^{x_{m-j}''},$$

and this word occurs as a factor of $c_m^{(w)}$ (up to a shifting of the letters a_1, \ldots, a_m). Note that n is not necessarily equal to m - j here, but that the factors of c_n^{\prime} corresponding to x' and x'' must end in a_{m-j} since $n \equiv m - j \pmod{m}$. For each $i \in \{2, \ldots, n-1\}$, we have $x'_i = x''_i$, so define $x_i = x'_i = x''_i$. Notice that the factor

$$a_1^{x_1''}a_2^{x_2}\dots a_{m-j-1}^{x_{n-1}}a_{m-j}^{x_n'}$$

appears on either side of $a_{m-j+1}^{\alpha_1} \dots a_m^{\alpha_j}$ in $c_m^{(u)}$, since $x_1'' \leq x_1'$ and $x_n' \leq x_n''$. Further, the facts that $j \leq k$ and $\sum_{i=1}^{j} \alpha_i \geq k$ (in \mathbb{Z}) allow us to factor $a_{m-j+1}^{\alpha_1} \dots a_m^{\alpha_j}$ into k nonempty words each of which is a power of a single letter. We conclude that the formula ψ_k occurs in $c_m^{(u)}$, and hence also occurs in $c_m^{(w)}$.

For the converse, suppose that ψ_k occurs in $c_m^{(w)}$ through the morphism h. First note that $h(xy_1...y_kx)$ cannot be a power of a single letter, as the powers in $c_m^{(w)}$ come from $w \in \{1, \ldots, k+1\}^*$. However, for each $i \in \{1, \ldots, k\}, h(y_i)$ must be a power of a single letter, as these are the only reversible factors of $c_m^{(w)}$ (since $m \ge 3$). Suppose without loss of generality that $h(y_1) = a_1^p$ and $h(y_k) = a_j^q$. We necessarily have $j \leq k$ due to the cyclic ordering of letters in $c_m^{(w)}$. We also see that we must have $h(x) = a_i^r v a_1^s$, for some $r, s \ge 0$, and

$$v = a_{j+1}^{t_1} a_{j+2}^{t_2} \dots a_m^{t_n}$$

with $n \equiv m - j \pmod{m}$, $t_i \in \{1, \dots, k+1\}$ for all $i \in \{1, \dots, n\}$. Define morphism h' by h'(x) = v, $h'(y_1) = a_1^s h(y_1)$, $h'(y_k) = h(y_k)a_i^r$, and $h'(y_i) = h(y_i)$ for all $i \in \{2, \ldots, k-1\}$. Then h' is also an occurrence of ψ_k in $c_m^{(w)}$.

Formulas with reversal of high avoidability index 5

Now if we write $h'(y_1 \dots y_k) = a_1^{\alpha_1} a_2^{\alpha_2} \dots a_j^{\alpha_j}$, then each instance of the factor $h'(xy_1 \dots y_k x) = vh'(y_1 \dots y_k)v$ in $c_m^{(w)}$ corresponds to a factor

 $u = t_1^+ t_2 t_3 \dots t_n \alpha_1 \alpha_2 \dots \alpha_j t_1 t_2 \dots t_n^+$

of w, where $t_1^+ \ge t_1$ and $t_n^+ \ge t_n$ (since we do not know what happens to the left and right of an instance of $h'(xy_1 \dots y_k x)$ in $c_m^{(w)}$). The factor u clearly has the form $x'\alpha_1 \dots \alpha_j x''$ described in the lemma statement.

We are now ready to prove that ψ_1 is 4-avoidable. The proof relies on an 8uniform morphism f that takes any square-free ternary word v to a binary word f(v) for which the associated 4-cyclic word avoids ψ_1 . From here, the exponential growth of the number of words avoiding ψ_1 is straightforward to prove using the fact that v can be any square-free ternary word.

Theorem 2.3. The formula ψ_1 is 4-avoidable. Further, the growth of the number of words on four letters avoiding ψ_1 is exponential.

Proof. Define a morphism $f : \{0, 1, 2\}^* \rightarrow \{1, 2\}^*$ by

f(0) = 11112122,f(1) = 12112222, and f(2) = 21111222.

We claim that if v is a square-free ternary word, then f(v) has no factor of the form $x'\alpha x''$, where $|\alpha| = 1$, |x'| = |x''| = n with $n \equiv 3 \pmod{4}$, and x' = x'' except possibly at the first and last letters (and thus $c_4^{(f(v))}$ avoids ψ_1 by Lemma 2.2). We prove the contrapositive of this claim below.

Let $w = f(v) = w_0 w_1 w_2 \dots$ and suppose that for some $k \ge n$, the factor

 $w_{k-n}\ldots w_{k-1}w_kw_{k+1}\ldots w_{k+n}$

has the form $x'\alpha x''$ described above. In particular, $n \equiv 3 \pmod{4}$, $w_{k-n} \ge w_{k+1}$, $w_{k-1} \le w_{k+n}$, and $w_{k-n+i} = w_{k+1+i}$ for all $i \in \{1, \ldots, n-2\}$. Note that here k represents the position of α in the reading of $x'\alpha x''$ as a factor of w. We have verified that no factor of this form occurs in the image of any square-free ternary word of length 2, so we may assume that n > 3. In what follows, we refer to the image of a single letter of v as a *code word* in w.

We first demonstrate that if $w_{k-n} \dots w_{k+n}$ has the form $x' \alpha x''$ described above, then $n \equiv 7 \pmod{8}$. By assumption, $n \equiv 3 \pmod{4}$, so it suffices to show that $n \not\equiv 3 \pmod{8}$. Suppose toward a contradiction that $n \equiv 3 \pmod{8}$. Using the assumption that n > 3, in particular we have

$$w_{k-n+1}w_{k-n+2}w_{k-n+3} = w_{k+2}w_{k+3}w_{k+4}.$$
(2.1)

Note that the third and fourth letters of every code word are 1 while the seventh and eighth letters are 2. Some letter from $w_{k-n+1}w_{k-n+2}w_{k-n+3}$ must sit in either

the third, fourth, seventh, or eighth position of a code word. However, since $n+1 \equiv 4 \pmod{8}$ the letter in the corresponding position in $w_{k+2}w_{k+3}w_{k+4}$ has the opposite identity. This contradicts (2.1).

Now we may assume that $n \equiv 7 \pmod{8}$. A key observation used below is that if two code words match at their first and second letters, then they are equal. This follows directly from the fact that f(0), f(1), and f(2) all have distinct prefixes of length 2. Similarly, if two code words match at their fifth and sixth letters, then they are equal. Finally, since f is 8-uniform we can tell exactly where code words begin and end in $w_{k-n} \dots w_{k+n}$ from the value of $k \mod 8$.

Case I: If $k \equiv 0 \pmod{8}$ then the code words

 $w_{k-n-1} \dots w_{k-n+6}$ and $w_k \dots w_{k+7}$

are equal since they match at their fifth and sixth letters. If n = 7 we are done, as we have two identical code words in a row, which must have come from a square (of length 2) in w. Otherwise, the code words

 $w_{k-8} \dots w_{k-1}$ and $w_{k+n-7} \dots w_{k+n}$

are also equal since they match at their fifth and sixth letters. Altogether we have

$$w_{k-n-1}\ldots w_{k-1}=w_k\ldots w_{k+n},$$

and hence the preimage of $w_{k-n-1} \dots w_{k+n}$ in f is a square. The same argument works when $k \equiv 7 \pmod{8}$ with all indices shifted to the right by 1.

Case II: If $k \equiv 1 \pmod{8}$ then the code words

 $w_{k-n-2} \dots w_{k-n+5}$ and $w_{k-1} \dots w_{k+6}$

are equal since they match at their fifth and sixth letters. Thus we have

 $w_{k-n-2} \dots w_{k-2} = w_{k-1} \dots w_{k+n-1},$

and hence the preimage of $w_{k-n-2} \dots w_{k+n-1}$ in f is a square. The same argument works when $k \equiv 2 \pmod{8}$ with all indices shifted to the left by 1.

Case III: If $k \equiv 3 \pmod{8}$, then the code words

 $w_{k-3} \dots w_{k+4}$ and $w_{k+n-2} \dots w_{k+n+5}$

are equal since they match at their first and second letters. Thus we have

 $w_{k-n+4} \dots w_{k+4} = w_{k+5} \dots w_{k+n+5},$

and hence the preimage of $w_{k-n+4} \dots w_{k+n+5}$ in f is a square. The same argument works when $k \equiv 4, 5, \text{ and } 6 \pmod{8}$ with all indices shifted to the left by 1, 2, and 3, respectively.

Formulas with reversal of high avoidability index 7

Therefore, if $v \in \{0, 1, 2\}^*$ is square-free, then the binary word f(v) avoids factors of the form $x'\alpha x''$. By Lemma 2.2, the 4-cyclic f(v)-word $c_4^{(f(v))}$ avoids ψ_1 . Since there are infinitely many square-free ternary words, we conclude that ψ_1 is 4 avoidable. It remains to show that the growth of the number of words on four letters avoiding ψ_1 is exponential.

It is well known that the number of square-free ternary words grows exponentially [3]. However, this does not immediately imply that the number of words on four letters avoiding ψ_1 grows exponentially, because there are square-free words uand v in $\{0,1,2\}^*$ of the same length for which $c_4^{(f(u))}$ and $c_4^{(f(v))}$ have different lengths. We would like f(u) and f(v) to have the same number of 1's and 2's so that we have $|c_4^{(f(u))}| = |c_4^{(f(v))}|$. One way to ensure this is to start with words u and v that have the letters 0, 1, and 2 in the same proportions. Fortunately, the number of square-free ternary words in which each alphabet letter occurs with proportion exactly 1/3 grows exponentially (while not explicitly stated there, this fact can easily be gleaned from [10, Sec. 4.1]; apply any of the Brinkhuis triples given there starting from initial word abc). From this fact we may conclude that the number of words on four letters avoiding ψ_1 grows exponentially.

Next we show that ψ_2 is 5-avoidable. The proof is much shorter than the preceding proof that ψ_1 is 4-avoidable because we make use of automatic theorem-proving software. We show that there is an infinite binary word **w** such that the 5-cyclic **w**-word $c_5^{(w)}$ avoids ψ_2 .

Theorem 2.4. The formula ψ_2 is 5-avoidable.

Proof. Define morphism
$$\rho : \{1,2\}^* \to \{1,2\}^*$$
 by $\rho(1) = 22$ and $\rho(2) = 21$. Let

 $w = \rho^{\infty}(2) = 21222121....$

We claim that $c_5^{(\mathbf{w})}$ avoids ψ_2 . We prove this claim using the automatic theoremproving software *Walnut* [9]. A description of this method is given in Appendix A.

The fact that the number of words avoiding ψ_2 grows exponentially follows fairly easily from Theorem 2.4. Essentially, we can replace a fixed number of 2's in $\rho^k(2)$ by 3's, and this change does not introduce any occurrence of ψ_2 to the corresponding 5-cyclic word. This gives us a set of words avoiding ψ_2 which we show grows exponentially in size.

Corollary 2.5. The growth of the number of words on five letters avoiding ψ_2 is exponential.

Proof. From the proof of Theorem 2.4, the infinite word

$$\rho^{\infty}(2) = 21222121...$$

has no factor of the form $x'\alpha_1 x''$ or $x'\alpha_1\alpha_2 x''$ satisfying the conditions listed in Lemma 2.2. Consider the finite prefix $\rho^k(2)$ for some $k \ge 2$ and let $n = 2^k = |\rho^k(2)|$.

Note that $|\rho^k(2)|_2 \ge \frac{n}{2}$. Note also that any word obtained from $\rho^k(2)$ by changing any number of 2's to 3's still has no factors of the form $x'\alpha_1x''$ or $x'\alpha_1\alpha_2x''$.

Let U_k be the set containing all words obtained from $\rho^k(2)$ by replacing exactly $\frac{n}{4}$ of the 2's by 3's. Clearly no word in U_k has a factor of the form $x'\alpha_1 x''$ or $x'\alpha_1\alpha_2 x''$, as this would imply that $\rho^k(2)$ has such a factor. Thus for any word $u \in U_k$, the 5-cyclic *w*-word $c_5^{(u)}$ avoids ψ_2 by Lemma 2.2. Further, any two words $u, v \in W_k$ have $|u|_i = |v|_i$ for all $i \in \{1, 2, 3\}$, so that every word in the set

$$\{c_5^{(u)}: u \in U_k\}$$

has the same length, which is at most 3n. Since there are at least $\binom{n/2}{n/4}$ words in U_k , we conclude that the growth of the number of words on five letters avoiding ψ_2 is exponential.

After discovering that $\operatorname{ind}(\psi_1) = 4$ and $\operatorname{ind}(\psi_2) = 5$, we were led to wonder whether $\operatorname{ind}(\psi_k)$ grows indefinitely along with k. We provide a negative answer to this question in the next section.

3. An Upper Bound on $ind(\psi_k)$ for $k \ge 3$

Our main result in this section is a general construction that shows the 5avoidability of ψ_{3k} for all $k \ge 1$. This leads us to believe that ψ_k is 5-avoidable for all $k \ge 3$. Although we are unable to verify this conjecture, we adapt our construction for ψ_{3k} to show that ψ_{3k+1} is 6-avoidable for $k \ge 1$ and ψ_{3k+2} is 6-avoidable for $k \ge 2$. Finally, we address the only remaining formula ψ_5 , showing that it is 7-avoidable.

We start with the main result concerning ψ_{3k} . The proof makes use of some new terminology for ease of reading. A *reversed variable* in a formula ϕ is a variable z such that z and z^R both appear in ϕ . In particular, in ψ_k the reversed variables are y_1, \ldots, y_k .

Theorem 3.1. For all $k \ge 1$, the formula ψ_{3k} is 5-avoidable. Further, the number of words on five letters avoiding ψ_{3k} grows exponentially.

Proof. Fix $k \ge 1$ and let $w = w_0 w_1 \ldots$ be a word in $\{0, 1, 2\}^*$. Define the morphism $d_k : \{0, 1, 2\}^* \to \{0, 1, 2\}^*$ by $i \mapsto i^{k+1}$ for all $i \in \{0, 1, 2\}$ and the morphism $g : \{0, 1, 2\}^* \to \{0, 1, 2, a, b\}^*$ by $i \mapsto iab$ for all $i \in \{0, 1, 2\}$. We claim that if ψ_{3k} occurs in $g(d_k(w))$, then w contains a square.

Suppose that ψ_{3k} occurs in $g(d_k(w))$ through morphism h. Several observations can be made from the fact that $g(d_k(w))$ alternates between a letter from $\{0, 1, 2\}$ and the factor ab:

- the *h*-image of each reversed variable in ψ_{3k} must have length 1, as these are the only reversible factors of $g(d_k(w))$;
- exactly k of the 3k reversed variables have h-image in $\{0, 1, 2\}$; and
- $|h(x)| \equiv 0 \pmod{3}$, hence h(x) contains at least one letter from $\{0, 1, 2\}$.

IJAC 1750024

Formulas with reversal of high avoidability index 9

Let u be the word obtained by dropping all a's and b's from h(x) and let v be the word obtained by dropping all a's and b's from $h(y_1 \ldots y_{3k})$. Clearly uvu is a factor of $d_k(w)$ as we are essentially taking a preimage in g, and by the remarks above we have $|u| \ge 1$ and |v| = k. Up to relabeling of letters, we have either $v = 0^k$ or $v = 0^i 1^j$ with i, j positive and i+j = k. We show below that in each of these cases the factor uvu extends to a factor in $d_k(w)$ that can only have come from a square in w.

Case I: $v = 0^k$.

If the factor 00 occurs in w, then we are done. So we may assume that 0^{k+2} is not a factor of uvu. Hence u must either start with a single 0 and end with a different letter or end with a single 0 and start with a different letter. But then uvu appears internally as $0^k u 0^k u$ or $u 0^k u 0^k$, respectively. The preimage in d_k of each of these factors is a square in w.

Case II: $v = 0^{i}1^{j}$ for positive i, j with i + j = k.

In this case, the factor uvu in $d_k(w)$ always appears inside the factor

 $1^{j}u0^{i}1^{j}u0^{i}$,

whose preimage is a square in w.

We conclude that if w is square-free, then $g(d_k(w))$ avoids the formula ψ_{3k} . Since the growth of the number of square-free words on $\{0, 1, 2\}^*$ is exponential [3], we conclude that the number of words avoiding ψ_{3k} grows exponentially as well, as $|g(d_k(w))| = (3k+3)|w|$, i.e. $g(d_k(w))$ is only a constant factor longer than w.

Now we obtain an upper bound on $\operatorname{ind}(\psi_{3k+1})$ and $\operatorname{ind}(\psi_{3k+2})$ using a similar idea to the one used above to show $\operatorname{ind}(\psi_{3k}) \leq 5$. It is easily verified that any long enough word of the form $g(d_k(w))$ encounters both ψ_{3k+1} and ψ_{3k+2} whether w is square-free or not, so the exact construction used for ψ_{3k} will not work for ψ_{3k_1} or ψ_{3k+2} . However, it turns out that we only need to make a slight modification to $g(d_k(w))$ to make it avoid ψ_{3k+1} (or ψ_{3k+2}) whenever w is square-free. All we need to do is add a new letter c to $g(d_k(w))$ at certain carefully chosen locations. The details are given in the corollaries presented below.

Corollary 3.2. For all $k \ge 1$, the formula ψ_{3k+1} is 6-avoidable and the number of words on six letters avoiding ψ_{3k+1} grows exponentially.

Proof. Let w be a word in $\{0, 1, 2\}^*$ and let d_k and g be as in Theorem 3. From $g(d_k(w))$, create a word u_w on alphabet $\{0, 1, 2, a, b, c\}^*$ by inserting the letter c after every 3k letters of $g(d_k(w))$. Now if ψ_{3k+1} occurs in u_w through morphism h, it is easily verified that the following conditions hold:

- the *h*-image of each reversed variable has length 1;
- exactly k of the 3k + 1 reversed variables have h-image in $\{0, 1, 2\}$; and

• h(x) contains at least one letter from $\{0, 1, 2\}$.

Therefore, by arguments very similar to those used in Theorem 3, if w is square-free then u_w avoids ψ_{3k+1} , and it follows that there are exponentially many words on six letters avoiding ψ_{3k+1} .

Corollary 3.3. For all $k \ge 2$, the formula ψ_{3k+2} is 6-avoidable and the number of words on six letters avoiding ψ_{3k+2} grows exponentially.

Proof. Let w be a word in $\{0, 1, 2\}^*$ and let d_k and g be as in Theorem 3. From $g(d_k(w))$, create a word v_w on alphabet $\{0, 1, 2, a, b, c\}^*$ by inserting the letter c after an appearance of the letter b whenever the total number of b's that have occurred so far is equivalent to 0 or 1 modulo k. Then in any factor of length 3k + 2 of v_w we have exactly k appearances of letters from $\{0, 1, 2\}$, k appearances of a, k appearances of b, and two appearances of c. The remainder of the proof is analogous to that of Corollary 3.2, and is omitted.

The construction of Corollary 3.3 does not work for k = 1 (that is, for ψ_5). We show below that $\operatorname{ind}(\psi_5) \leq 7$ using a construction similar to the ones we have already seen. This does not seem optimal, but we have not found a construction using fewer letters.

Corollary 3.4. The formula ψ_5 is 7-avoidable and the number of words on seven letters avoiding ψ_5 grows exponentially.

Proof. Let w be a word in $\{0, 1, 2\}^*$ and let d_2 be as in Theorem 3. Define morphism $g' : \{0, 1, 2\}^* \to \{0, 1, 2, a, b, c, d\}^*$ by $i \mapsto iabcd$ for all $i \in \{0, 1, 2\}$. By arguments similar to those already seen in Theorem 3, if w is square-free then $g'(d_2(w))$ avoids ψ_5 , and it follows that there are exponentially many words on 7 letters avoiding ψ_5 .

Now that we have an upper bound on $\operatorname{ind}(\psi_k)$ for all $k \ge 3$, we prove a nontrivial lower bound on $\operatorname{ind}(\psi_k)$ in the next section.

4. A Lower Bound on $ind(\psi_k)$ for $k \ge 3$

It is trivially true that $\operatorname{ind}(\psi_k) \ge 2$ for all $k \ge 1$ since we have proven that ψ_k is avoidable and it is obvious that ψ_k is 1-unavoidable (this is of course true for every formula with reversal). Using a standard backtracking algorithm, we have found that ψ_k is 4-unavoidable for $2 \le k \le 6$. Table 1 shows the length of a longest word on four letters avoiding ψ_k for $2 \le k \le 6$.

The issue with our backtracking calculation is that it gets more computationally intensive as the value of k grows, becoming infeasible rather quickly. The main result of this section is that $\operatorname{ind}(\psi_k) \geq 3$ for the remaining cases $k \geq 7$. We begin with a lemma that will be required to prove this main result.



Formulas with reversal of high avoidability index 11

Table 1. Length of a longest word on four letters avoiding ψ_k for $2 \le k \le 6$.

k	Length of a longes
	word avoiding ψ_k
2	45
3	34
4	29
5	29
6	41

Lemma 4.1. For any $k, n \ge 1$, the 3-cyclic w-word $c_3^{(w)}$ encounters ψ_k for any word $w \in \{1, \ldots, n\}^{\omega}$.

Proof. Let $w = w_0 w_1 \dots$ and let $c_3^{(w)} = 0^{w_0} 1^{w_1} 2^{w_2} 0^{w_3} \dots$ in this proof for ease of reading. We have three cases, one for each possible value of $k \mod 3$.

Case I: $k \equiv 2 \pmod{3}$.

In this case the formula ψ_k occurs in $c_3^{(w)}$ as follows:

$$\underbrace{\underbrace{0}_{\overset{1}{\underline{1}}}\underbrace{1^{w_1}}_{\overset{1}{\underline{y}_1}}\ldots\underbrace{2^{w_k}}_{\overset{1}{\underline{y}_k}}\underbrace{0}_{\overset{1}{\underline{x}}}$$

Case II: $k \equiv 0 \pmod{3}$.

First of all, if $w = 1^{\omega}$ then ψ_k occurs in $c_3^{(w)}$ as follows:

$$\underbrace{\underbrace{012}}_{\frac{1}{x}}\underbrace{\underbrace{0}}_{\frac{1}{y_1}}\underbrace{1}_{\frac{1}{y_2}}\underbrace{\dots}\underbrace{2}_{\frac{1}{y_k}}\underbrace{\underbrace{012}}_{\frac{1}{x}}.$$

We may now assume that some letter of w is greater than 1. By shifting the index if necessary, we may assume that $w_1 > 1$. Then ψ_k occurs in $c_3^{(w)}$ as follows:

$$\underbrace{0}_{\frac{1}{k}}\underbrace{1}_{\frac{y_1}{j_1}}\underbrace{1}_{\frac{y_2}{j_2}}\underbrace{2}_{\frac{y_3}{j_3}}\underbrace{0}_{\frac{y_4}{j_4}}\cdots\underbrace{2}_{\frac{y_{k-1}}{j_k}}\underbrace{0}_{\frac{y_{k-1}}{j_k}}\cdots\underbrace{0}_{\frac{y_{k-1}}{j_k}}$$

Case III: $k \equiv 1 \pmod{3}$.

We have already demonstrated by backtracking that ψ_1 is 3-unavoidable, so certainly it occurs in every infinite word of the form $c_3^{(w)}$. We now handle $k \ge 4$. First of all, if $w = 1^{\omega}$ then ψ_k occurs in $c_3^{(w)}$ as follows:

$$\underbrace{\underbrace{01}}_{\stackrel{1}{x}}\underbrace{\underbrace{2}}_{\stackrel{1}{y_1}}\underbrace{0}_{\frac{1}{y_2}}\cdots\underbrace{2}_{\stackrel{1}{y_k}}\underbrace{01}_{\stackrel{1}{x}}$$

Not it suffices to show that ψ_k occurs in $c_3^{(w)}$ whenever some letter of w is at least 2. First suppose that some letter of w is strictly greater

than 2. By shifting the index if necessary, we can assume $w_1 > 2$. Then ψ_k occurs in $c_3^{(w)}$ as follows:

 $\underbrace{0}_{\frac{1}{k}} \underbrace{1}_{\frac{1}{k}} \underbrace{1}_{\frac{1}{2}} \underbrace{1}_{\frac{1}{2}} \underbrace{2}_{\frac{1}{2}} \underbrace{2}_{\frac{1}{2}} \underbrace{0}_{\frac{1}{2}} \underbrace{0}_{\frac{1}{2}} \cdots \underbrace{2}_{\frac{1}{2}} \underbrace{0}_{\frac{1}{2}} \cdots \underbrace{1}_{\frac{1}{2}} \cdots \underbrace{1}_{\frac{1}{2}} \underbrace{0}_{\frac{1}{2}} \cdots \underbrace{1}_{\frac{1}{2}} \cdots$

Suppose now that w contains the factor 22. By shifting the index if necessary we may assume $w_1 = w_2 = 2$. Then ψ_k occurs in $c_3^{(w)}$ as follows:

$$\underbrace{0}_{\frac{1}{x}}\underbrace{1}_{y_{1}^{1}}\underbrace{1}_{y_{2}^{1}}\underbrace{2}_{y_{3}}\underbrace{2}_{y_{4}^{2}}\underbrace{0}_{y_{5}^{1}}\underbrace{1}_{y_{6}^{1}}\underbrace{1}_{y_{6}^{2}}\underbrace{1}_{y_{k}^{1}}\underbrace{2}_{\frac{1}{x}}\underbrace{0}_{\frac{1}{x}}$$

Now suppose that w contains the letter 2; by shifting the index if necessary we may assume $w_0 = 2$. Further, we can take $w_1 = 1$ since we have shown that any w with the factor 22 encounters ψ_k . Finally, since we have shown that any w with a letter strictly greater than 2 encounters ψ_k , we can take $w_{k+2} \leq 2 = w_0$. Thus, we see that ψ_k occurs in $c_m^{(w)}$ as follows:

$$\underbrace{\underbrace{0^{w_{k+2}}}_{\frac{1}{x}}\underbrace{2^{w_2}}_{j_1^{1}}\underbrace{0^{w_3}}_{j_2^{1}}\cdots\underbrace{2^{w_{k+1}}}_{j_k^{1}}\underbrace{0^{w_{k+2}}1}_{\frac{1}{x}}$$

We conclude that $c_3^{(w)}$ encounters ψ_k for every word $w \in \{1, \ldots, n\}^{\omega}$.

In other words, Lemma 4.1 demonstrates that any ω -word on three letters avoiding ψ_k has at least one reversible factor containing two distinct letters. This fact will be important to the proof of the following theorem, the main result of this section.

Theorem 4.2. For all $k \ge 1$, the formula ψ_k is 3-unavoidable.

Proof. We have shown by backtracking that ψ_k is 3-unavoidable for $k \leq 6$, so it remains to verify the theorem statement for $k \geq 7$. We proceed by induction on k. Suppose for some $k \geq 6$ that ψ_k is 3-unavoidable. Suppose toward a contradiction that $w \in \{0, 1, 2\}^{\omega}$ is a recurrent word that avoids ψ_{k+1} .

By the induction hypothesis, we know that ψ_k occurs in w, say through morphism h. We claim that h must be 1-uniform. First suppose that |h(x)| > 1. Let h(x) = va, where $v \neq \varepsilon$ and |a| = 1. Then ψ_{k+1} occurs in w through g defined by

$$\begin{array}{ll} x\mapsto v, & \text{and} \\ \\ y_i\mapsto \begin{cases} a & \text{if } i=1 \\ h(y_{i-1}) & \text{if } i>1 \end{cases} \end{array}$$

Now suppose that $|h(y_j)| > 1$ for some $j \in \{1, ..., k\}$. Let $h(y_j) = va$, where $v \neq \varepsilon$ and |a| = 1. Since $h(y_j)$ is reversible in w by the fact that h is an occurrence of

Formulas with reversal of high avoidability index 13

 ψ_k , v must also be reversible in w, and we see that ψ_{k+1} occurs in w through g defined by

$$x \mapsto h(x), \quad \text{and}$$

$$y_i \mapsto \begin{cases} h(y_i) & \text{if } i < j, \\ v & \text{if } i = j, \\ a & \text{if } i = j+1, \\ h(y_{i-1}) & \text{if } i > j+1. \end{cases}$$

Thus h must be 1-uniform as claimed.

Now take an instance of the fragment $xy_1 \dots y_k x$ under some occurrence of ψ_k in w through a 1-uniform morphism. Without loss of generality we may assume that $x \mapsto 0$, so an instance of $xy_1 \dots y_k x$ has the form

$$0a_1\ldots a_k0,$$

where $a_i \in \{0, 1, 2\}$. First note that this factor cannot be preceded or followed by a 0 in w as this gives an obvious occurrence of ψ_{k+1} . Thus this factor appears internally in w as either $10a_1 \dots a_k 01$ or $10a_1 \dots a_k 02$ (up to relabeling of letters).

Consider the former possibility $10a_1 \dots a_k 01$. If $a_1 = 0$ or $a_1 = 1$, then the factor $0a_1$ is reversible, so ψ_{k+1} occurs in w as follows:

$$\underbrace{1}_{\overset{1}{x}}\underbrace{0a_1}_{\overset{1}{y_1}}\underbrace{a_2}_{\overset{1}{y_2}}\ldots\underbrace{a_k}_{\overset{1}{y_k}}\underbrace{0}_{\overset{1}{y_{k+1}}}\underbrace{1}_{\overset{1}{x}}$$

Thus we may assume that $a_1 = 2$. By a symmetric argument, we may assume that $a_k = 2$. However, then ψ_{k+1} occurs in w as follows:

$$\underbrace{1}_{\overset{1}{x}}\underbrace{02}_{\overset{1}{y_1}}\underbrace{a_2}_{\overset{1}{y_2}}\ldots\underbrace{a_{k-1}}_{\overset{1}{y_{k-1}^{\dagger}}}\underbrace{2}_{\overset{1}{y_k}}\underbrace{0}_{\overset{1}{y_{k+1}^{\dagger}}}\underbrace{1}_{\overset{1}{x}}$$

We may now assume that $0a_1 \ldots a_k 0$ appears internally as $10a_1 \ldots a_k 02$. In fact, we can demonstrate that it must appear internally as $10a_1 \ldots a_k 02^t 1$ for some $t \ge 1$. Otherwise, ψ_{k+1} occurs in w as follows:

$$\underbrace{0}_{\frac{1}{x}}\underbrace{a_1}_{y_1}\underbrace{a_2}_{y_2}\cdots\underbrace{a_k}_{y_k}\underbrace{02^t}_{y_{k+1}^{\dagger}}\underbrace{0}_{\frac{1}{x}}.$$

Now to reach a contradiction it suffices to show that $0a_1a_2...a_k0$ can be factored as $b_1b_2...b_k$, where $b_i \neq \varepsilon$ and is reversible for each $i \in \{1,...,k\}$. If this is the case then ψ_k occurs in $10a_1...a_k02^t1$ as follows:

$$\underbrace{1}_{\overset{1}{x}}\underbrace{b_1}_{y_1}\underbrace{b_2}_{y_2}\cdots\underbrace{b_k}_{y_k}\underbrace{2^t}_{y_{k+1}}\underbrace{1}_{\overset{1}{x}}$$

Clearly if $0a_1 \ldots a_k 0$ contains a length 3 reversible factor then it can be factored into k reversible factors. Hence we may assume that $0a_1 \ldots a_k 0$ contains no factors of the form c^3 , cdc, or cddc for $c, d \in \{0, 1, 2\}$ (the first two factors are length 3 reversible factors while in the third factor cdd is reversible).

Case I: $0a_1 \dots a_k 0$ contains at least two squares of length 2. Clearly we can factor $0a_1 \dots a_k 0$ into k reversible factors by choosing any two length 2 squares.

Case II: $0a_1 \dots a_k$ contains no squares of length 2.

By the assumption that $0a_1 \dots a_k 0$ has no factors of the form cdc with $c, d \in \{0, 1, 2\}$, we necessarily have $k \equiv 2 \pmod{3}$ and $0a_1 \dots a_k 0$ must have the cyclic form $021021 \dots 0$. In particular, we must have $k \equiv 2 \pmod{3}$. By Lemma 4.1, we know that the entire ω -word w cannot have cyclic form, so that at least one of the factors 02, 21, and 10 is reversible in w. Each factor 02, 21, and 10 appears at least twice in $0a_1 \dots a_k 0$ (since $k \ge 6$), meaning that there will be at least two nonoverlapping reversible factors of length 2 in $0a_1 \dots a_k 0$, allowing us to factor it into k reversible factors.

Case III: $0a_1 \ldots a_k 0$ contains exactly one square of length 2.

By the assumption that $0a_1 \ldots a_k 0$ has no factors of the form cdc or cddc with $c, d \in \{0, 1, 2\}$, we see that $0a_1 \ldots a_k 0$ must have the cyclic form $0^{p_1} 2^{p_2} 1^{p_3} 0^{p_4} 2^{p_5} 1^{p_6} \ldots 0^{p_{k+1}}$ where $p_i = 2$ for some $i \in \{1, \ldots, k+1\}$ and $p_j = 1$ for all $j \neq i$. In particular, we must have $k \equiv 0 \pmod{3}$. Again, by Lemma 4.1 we know that at least one of the factors 02, 21, and 10 must be reversible in w. It is easy to verify that each of these factors appears at least once in $0^{p_1} 2^{p_2} 1^{p_3} 0^{p_4} \ldots 0^{p_{k+1}}$ without overlapping the square of length 2 (again, recall $k \geq 6$). Thus we can factor $0a_1 \ldots a_k 0$ into k reversible factors.

We have shown that ψ_{k+1} occurs in w, a contradiction. By induction, we conclude that ψ_k is 3-unavoidable for all $k \ge 1$.

While we conjecture that ψ_k is actually 4-unavoidable for all $k \ge 2$, the proof technique used for Theorem 4.2 does not seem tractable on a 4-letter alphabet. It appears that a different technique will be necessary in order to show that this conjecture holds.

5. Conclusion

The family $\{\psi_k: k \ge 1\}$ is the first infinite family of avoidable formulas (with reversal) that we know of whose members are all 3-unavoidable. While we have shown that $\operatorname{ind}(\psi_1) = 4$ and $\operatorname{ind}(\psi_k) = 5$ for $k \in \{2, 3, 6\}$, we have only demonstrated bounds on the avoidability index of the remaining formulas. It would be nice to know the exact avoidability index of ψ_k for all $k \ge 2$; we suspect that it is 5.

After discovering such simple formulas with reversal of avoidability index 5, it seems plausible to us that there are avoidable formulas with reversal of avoidability index 6 which can be found. This would be especially interesting as there are currently no known formulas (with or without reversal) of avoidability index 6.

Formulas with reversal of high avoidability index 15

Acknowledgments

The authors wish to thank the anonymous referee for his or her valuable comments and suggestions. This work was supported by the Natural Sciences and Engineering Research Council of Canada (NSERC), grant numbers 418646-2012 and 42410-2010.

Appendix A. Walnut Predicates for Theorem 2.4

Let $\rho: \{1,2\}^* \to \{1,2\}^*$ be given by $\rho(1) = 22$ and $\rho(2) = 21$, and let $\mathbf{w} = \rho^{\infty}(2)$. We show that $c_5^{(\mathbf{w})}$ avoids ψ_2 . By Lemma 2.2, it suffices to show that \mathbf{w} avoids factors of the following two forms:

- $x'\alpha_1 x''$, where
 - $\begin{array}{l} \alpha_1 = 2, \\ |x'| = |x''| = n \text{ with } n \equiv 4 \pmod{5}, \text{ and} \\ \text{ if } x' = x'_1 \dots x'_n \text{ and } x'' = x''_1 \dots x''_n, \text{ then } x'_1 \ge x''_1, x'_n \le x''_n, \text{ and } x'_i = x''_i \text{ for all} \\ i \in \{2, \dots, n-1\}. \end{array}$
- $x'\alpha_1\alpha_2 x''$, where
 - $\begin{array}{l} \alpha_1, \alpha_2 \in \{1, 2\}, \\ |x'| = |x''| = n \text{ with } n \equiv 3 \pmod{5}, \text{ and} \\ \text{ if } x' = x'_1 \dots x'_n \text{ and } x'' = x''_1 \dots x''_n, \text{ then } x'_1 \ge x''_1, x'_n \le x''_n, \text{ and } x'_i = x''_i \text{ for all} \\ i \in \{2, \dots, n-1\}. \end{array}$

Throughout, our automata take a binary representation of a number as input, and read the binary digits from most significant to least significant. Since **w** is a fixed point of a 2-uniform morphism, **w** is 2-automatic, and is generated by the deterministic finite automaton with output shown in Fig. 1 (see the proof of [1, Theorem 6.3.2] for details). This automaton generates $\mathbf{w} = w_0 w_1 \dots$ in the following manner: given a binary representation of a number *i*, if we begin at the initial state 2 and follow the transitions indicated by reading the binary digits of *i* from most significant to least significant, halting when we reach the end of *i*, then we will have halted at state w_i . In other words, given input *i*, the automaton outputs w_i .



Fig. 1. The deterministic finite automaton with output that generates w.





Fig. 2. The deterministic finite automaton is4mod5base2.

We first create a text file which defines this automaton for use in *Walnut*. The text is as follows:

```
msd_2
0 2
0 -> 0
1 -> 1
1 1
0 -> 0
1 -> 0
```

We save this file as "w.txt" in the directory "/Walnut/Word Automata Library/", which allows us to refer to this automaton in later predicates. Explicitly, "w[i]" in a *Walnut* predicate returns w_i .

Next, we construct two deterministic finite automata for use in later predicates: one that accepts binary numbers equivalent to 4 modulo 5, and another that accepts binary numbers equivalent to 3 modulo 5. These automata are constructed by the following predicates, respectively:

```
def is4mod5base2 "?msd_2 Ei n=5*i+4";
def is3mod5base2 "?msd_2 Ei n=5*i+3";
```

To be completely clear, given input the binary number n, the automaton is4mod5base2 reads n from left to right and accepts n if and only if $n \equiv 4 \pmod{5}$. The automaton is4mod5base2 is pictured in Fig. 2. The automaton is3mod5base2 behaves analogously, accepting n if and only if $n \equiv 3 \pmod{5}$.

Now we construct an automaton Case1 with the following predicate:

```
eval Case1 "?msd_2 $is4mod5base2(n)
    & (Ei w[i]>=w[i+n+1] & w[i+n-1]<=w[i+2*n]
    & Aj (j>0 & j<n-1) => w[i+j]=w[i+n+j+1]
    & w[i+n]=w[0])";
```

Formulas with reversal of high avoidability index 17

Given the binary number n, the automaton Case1 reads n from left to right and accepts n if and only if

- $n \equiv 4 \pmod{5}$; and
- there exists an integer i such that
 - $w_i \ge w_{i+n+1},$
 - $w_{i+n-1} \le w_{i+2n},$
 - $-w_{i+j} = w_{i+n+j+1}$ for all $j \in \{1, \ldots, n-2\}$, and
 - $-w_{i+n} = w_0 = 2.$

In other words, it accepts n if and only if there is some factor $w_i \dots w_{i+2n}$ of \mathbf{w} of the form $x'\alpha_1 x''$ described above; explicitly, $x' = w_i \dots w_{i+n-1}$, $\alpha_1 = w_{i+n}$, and $x'' = w_{i+n+1} \dots w_{i+2n}$. The automaton with a single, nonaccepting state is returned, meaning that \mathbf{w} has no factors of this form.

Finally, we construct an automaton Case2 with the following predicate:

```
eval Case2 "?msd_2 $is3mod5base2(n)
    & (Ei w[i]>=w[i+n+2] & w[i+n-1]<=w[i+2*n+1]
    & Aj (j>0 & j<n-1) => w[i+j]=w[i+n+j+2])";
```

The automaton Case2 accepts all natural numbers n for which **w** has a factor of the form $x'\alpha_1\alpha_2x''$ described above. The automaton with a single, nonaccepting state is returned, meaning that **w** has no factors of this form.

References

- J.-P. Allouche and J. Shallit, Automatic Sequences: Theory, Application, Generalizations (Cambridge University Press, 2003).
- [2] K. A. Baker, G. F. McNulty and W. Taylor, Growth problems for avoidable words, *Theoret. Comput. Sci.* 69 (1989) 319–345.
- [3] F.-J. Brandenburg, Uniformly growing kth power-free homomorphisms, *Theoret. Comput. Sci.* 23 (1983) 69–82.
- [4] J. Cassaigne, Motifs évitables et régularité dans les mots, Ph.D. thesis, Université Paris VI, Paris, France (1994), 181 pages.
- [5] R. J. Clark, Avoidable formulas in combinatorics on words, Ph.D. thesis, University of California, Los Angeles, Los Angeles, California (2001), 78 pages.
- [6] J. D. Currie and P. Lafrance, Avoidability index for binary patterns with reversal, *Electron. J. Combin.* 23(1) (2016), #P1.36, 14 pages (electronic).
- [7] J. D. Currie and N. Rampersad, Binary words avoiding xx^Rx and strongly unimodal sequences, J. Integer Seq. 18 (2015) 1–7.
- [8] J. D. Currie and N. Rampersad, Growth rate of binary words avoiding xxx^r, Theoret. Comput. Sci. 609 (2016) 456–468.
- [9] H. Mousavi, Automatic theorem proving in Walnut, preprint (2016), arXiv:1603.06017.
- [10] C. Richard and U. Grimm, On the entropy and letter frequencies of ternary squarefree words, *Electron. J. Combin.* 11 (2004), #R14, 19 pages (electronic).