Biologically unavoidable sequences

Samuel A. Alexander*†

Department of Mathematics, the Ohio State University

May 5, 2014

Abstract

A biologically unavoidable sequence is an infinite gender sequence which occurs in every gendered, infinite genealogical network satisfying certain tame conditions. We show that every eventually periodic sequence is biologically unavoidable (this generalizes König’s Lemma), and we exhibit some biologically avoidable sequences. Finally we give an application of unavoidable sequences to cellular automata.

1 Introduction

The following definition is motivated by biological considerations. The idea of modelling the biosphere with a directed graph goes at least back to Hennig [10], more recently (and more formally) to Dress, Moulton, Steel and Wu [6]. The idea of including vertices for future organisms was made explicit in Kornet, Metz, and Schellinx [12]. The (simplifying) assumption of infinitely many vertices was made explicit in Alexander [2]. We hope to submit the results of this paper as an answer to Sturmfels [16].

Definition 1. Let $n > 0$. An infinite $n$-gendered population is a directed graph $G = (V, E)$ together with a map $v \mapsto t(v) \in \mathbb{R}$ assigning birthdates to vertices and a map $e \mapsto g(e) \in \{1, \ldots, n\}$ assigning genders to edges, satisfying the following conditions (we call $v$ a parent of $u$ if $(v, u) \in E$, we define child similarly, and we define ancestor and descendant in the obvious (strict) way):

- (A1) There are only finitely many parentless vertices (call them roots).
- (A2) Each vertex has only finitely many children.
- (A3) For every $r \in \mathbb{R}$, $\{v \in V : t(v) \leq r\}$ is finite, and for each $(u, v) \in E$, $t(u) < t(v)$.
- (A4) $|V| = \infty$.
- $(n$-Gendered) For every non-root $u$ and every $1 \leq i \leq n$, $u$ has a parent $v$ with $g(v, u) = i$.

Note that A3 implies $G$ is a DAG and is reverse-well-founded, i.e., has no infinite reverse-directed path. Thus every non-root is a descendant of some root. Although Definition 1 stipulates that edges be gendered, we will often consider special cases where vertices are gendered, implicitly giving each edge the gender of its initial vertex.

Definition 2. Let $n > 0$. An infinite sequence $s = (s_1, s_2, \ldots) \in \{1, \ldots, n\}^\mathbb{N}$ is biologically unavoidable if every infinite $n$-gendered population realizes $s$ – by which we mean there is a vertex sequence $v_1, v_2, \ldots$ with each $v_i$ a parent of $v_{i+1}$ and each $g(v_i, v_{i+1}) = s_i$. If $s$ is not biologically unavoidable, it is biologically avoidable.

A priori, biological unavoidability appears ill-defined. The following lemma shows that it is well-defined.

Lemma 1. Suppose $s \in \{1, \ldots, n\}^\mathbb{N}$ and at the same time $s \in \{1, \ldots, m\}^\mathbb{N}$. Then $s$ is realized in every infinite $n$-gendered population if and only if $s$ is realized in every infinite $m$-gendered population.
Proof. Without loss of generality, \( n < m \). If there is an infinite \( m \)-gendered population which fails to realize \( s \), delete all edges with genders \( > n \) to obtain an infinite \( n \)-gendered population which fails to realize \( s \). Conversely, suppose there is an \( n \)-gendered population \( P \) failing to realize \( s \). Inductively it suffices to show there’s an \((n+1)\)-gendered population failing to realize \( s \). Let \( P' \) be a disjoint copy of \( P \). To the graph \( P' \cup P \), add two \((n+1)\)-gendered edges, \((v', u)\) and \((v, u')\), for every \( n \)-gendered edge \((v, u)\) in \( P \). It’s easy to see this makes \( P \cup P' \) an \((n+1)\)-gendered population which fails to realize \( s \).

In Section 2, we will show that every eventually periodic sequence is biologically unavoidable.

In Section 3, we will establish the existence of biologically avoidable sequences. This nontrivial fact involves a couple of unexpected enumerative combinatorial arguments.

In Section 4, we will show that there are biologically avoidable sequences from \( \{1, 2\}^N \) which contain no gender repeated thrice in a row.

In Section 5, we will give an unexpected application to cellular automata: an alternate proof of a result about spaceship speed limits in Conway’s Life-like Games, first proved by Nathaniel Johnston [11].

2 Eventually periodic implies biologically unavoidable

Before proving the unavoidability of eventually periodic sequences, a small amount of machinery must be developed. For the remainder of this section, let \( P = (V, E) \) be an infinite \( n \)-gendered population, \( n > 0 \).

Definition 3. For \( i \in \mathbb{N} \), we define the set \( V_i \subseteq V \) as follows: a vertex \( u \in V \) lies in \( V_i \) if and only if there is some root \( r \) and some directed path from \( r \) to \( u \) of length \( \leq i \).

Thus \( V_0 \) is the set of roots, \( V_1 \) contains the roots and their children, and so on.

Lemma 2. For every \( i \in \mathbb{N} \), \( V_i \) is finite.

Proof. Follows trivially from A1 and A2.

Lemma 3. For every \( 1 \leq i \leq n \), there is a map \( i^* : V \setminus V_0 \to V \) such that for every \( u \in V \setminus V_0 \) we have \((i^* u, u) \in E \) and \( g(i^* u, u) = i \).

Proof. By the axiom of choice and the \( n \)-Gendered assumption on \( P \).

In case \( n = 2 \), one might refer to \( 1^* \) and \( 2^* \) as motherhood and fatherhood maps, in some order. We will write \( i^* j^* \) for \( i^* \circ j^* \).

Definition 4. If \( s = (s_1, s_2, \ldots) \) is a periodic sequence on \( \{1, \ldots, n\} \), with period \( p \), and if \( u \in V \), then we define the \( s \)-path to \( u \), a finite directed path, written \( s^* u \), as follows.

1. If \( u \in V_{p-1} \) then \( s^* u = (u) \) (the length 0 path).

2. Otherwise,

\[
\begin{align*}
\begin{array}{c}
V \setminus V_0 \xrightarrow{(i^*, \text{Id})} E \\
\downarrow \\
\{i\} \xrightarrow{\cdot} \{1, \ldots, n\}
\end{array}
\end{align*}
\]

\[
s^* u = (s_1^* \cdots s_p^* u, s_2^* \cdots s_p^* u, \ldots, s_{pq-1}^* s_p^* u, s_p^* u, u),
\]

where \( q \) is maximal such that \( s_1^* \cdots s_p^* u \) is defined. (Some such \( q > 0 \) exists lest \( u \) lie in \( V_{p-1} \), and such \( q \) are bounded above due to Assumption A3.)

For example, suppose \( n = 2 \) and \( \{1, 2\} = \{M, F\} \). If \( s = (M, M, M, \ldots) \), then \( s^* u \) is obtained as follows: start with \( u \), go to \( u \)'s father, go to his father, and so on until a root is reached; then reverse the order of the resulting path. If \( s = (M, F, F, M, F, F, \ldots) \), then \( s^* u \) is obtained by starting at \( u \), taking his mother, taking her mother, taking her father, and repeating this three-step process until too close to a root to continue; and then reversing the resulting path.

Lemma 4. Let \( s, p \) be as in Definition 4. For any \( u \in V \), \( s^* u \) is a finite directed path starting at a vertex in \( V_{p-1} \) and ending at \( u \), and if \( s^* u = (v_1, \ldots, v_r) \) then for all \( 1 \leq i < r \), \( g(v_i, v_{i+1}) = s_i \).
Proof. To see that \( s^*u \) starts in \( V_{p-1} \), write \( s^*u = (v_1, \ldots, v_t) \) and assume \( v_1 \not\in V_{p-1} \). Then \( s_1^* \cdots s_p^*v_1 \) is defined. Let \( q \) be as in the definition of \( s^*u \). By periodicity of \( s \),
\[
    s_1^* \cdots s_p^*v_1 = s_1^* \cdots s_p^*(s_1^* \cdots s_p^*u) \\
    = s_1^* s_p^*s_1^* \cdots s_p^*u \\
    = s_1^* \cdots s_p^*(q+1)u,
\]
violating the maximality of \( q \). The rest of the lemma is clear. \( \square \)

**Proposition 5.** Every periodic sequence is biologically unavoidable. In fact, if \( s \) is a sequence with period \( p \), it is realized by a path in \( P \) which begins at a vertex in \( V_{p-1} \).

Proof. Consider the set (call it \( X \)) of finite directed paths \( \{s^*u \mid u \in V \} \) (note \( |X| = \infty \) by Assumption A4). Each begins (by Lemma 3) in \( V_{p-1} \), and \( V_{p-1} \) is finite by Lemma 2. Thus, there is some \( u_1 \in V_{p-1} \) such that infinitely many members (call them \( X_1 \)) of \( X \) begin at \( u_1 \).

Inductively, suppose I’ve defined a finite directed path \( u_1, \ldots, u_k \) and an infinite set \( X_k \subseteq X \) such that

1. For all \( \Pi \in X_k \), the \( i \)th vertex in \( \Pi \) is \( u_k \); and
2. \( g(u_i, u_{i+1}) = s_i \) for all \( 0 < i < k \).

It follows from Assumptions A1 and A2 that \( X_k \) contains only finitely many paths of length \( \leq k \), we may assume it contains no paths so short.

Since each path in \( X_k \) has \( u_k \) as \( k \)th vertex, each path in \( X_k \) has some child of \( u_k \) as \( (k+1) \)th vertex. By Assumption A2, \( u_k \) only has finitely many children. Thus there is a child \( u_{k+1} \) of \( u_k \) such that an infinite subset \( X_{k+1} \) of paths in \( X_k \) have \( u_{k+1} \) as \( (k+1) \)th vertex.

In particular, \( X_{k+1} \) has at least one path, \( s^*v \) for some \( v \in V \). By Lemma 4 \( g(u_k, u_{k+1}) = s_k \).

This inductively defines \( u_1, u_2, \ldots \) with all the desired properties. \( \square \)

It is also possible to prove Proposition 5 using the compactness theorem from first-order logic. Compare the proof (see Simpson’s book [15]) that weak König’s Lemma is equivalent (over RCA\(_0\)) to the compactness theorem.

**Theorem 6.** Every eventually periodic sequence is biologically unavoidable.

Proof. If \( s \) is eventually periodic, write it as \( s = t \cdot t' \) where \( t \) is finite of length \( k \) and \( t' \) is periodic. Let \( P' \) be the population obtained from \( P \) by discarding all vertices in \( V_{k-1} \); it’s easy to see \( P' \) remains an infinite gendered population. By Proposition 5 there is a directed path in \( P' \) realizing \( t' \). This defines a path \( u_1, u_2, \ldots \) in \( P \), realizing \( t' \), and avoiding \( V_{k-1} \). Back-extend this path to
\[
    (t_1^* \cdots t_k^*u_1, \ t_2^* \cdots t_k^*u_1, \ \ldots, \ t_k^*u_1, \ u_1, \ u_2, \ \ldots),
\]
which realizes \( s \) as desired; this is possible because if not, that would imply \( u_1 \in V_{k-1} \). \( \square \)

This generalizes König’s Lemma for trees, which can be seen as the 1-gender case of Theorem 6 with the additional constraint that vertices have only one parent. The following corollary paraphrases this idea even further.

**Corollary 7.** There is a subset \( V_0 \subseteq V \), ancestrally closed (whenever \( v \in V_0 \) and \( u \) is an ancestor of \( v \) then \( u \in V_0 \)), such that for every eventually periodic sequence \( s \in \{1, \ldots, n\}^\mathbb{N} \), \( G \) realizes \( s \) with a path \( p \) entirely in \( V_0 \), with the additional property that every vertex in \( V_0 \) has a descendant on \( p \).

Proof. By Theorem 6 above, combined with Theorem 3 and Proposition 5 from Alexander [2]. \( \square \)

Using Theorem 6 we can give a game-theoretical characterization of unavoidable sequences using the notion of guessability discovered by Wadge [17] (pp. 141–145) (and independently by Alexander [1]). Let \( s \) be a sequence on \( \{1, \ldots, n\} \). In the game \( G_s \), I starts by playing an infinite \( n \)-gendered population \( P \). Thereafter, \( II \) plays a path \( p \) in \( P \) (one vertex per turn) and I tries to guess (making one guess per turn) whether \( p \)'s genders have the form \( t \cdot s \) for some finite \( t \). I wins if I’s guesses converge to the correct answer, \( II \) wins otherwise. We leave it an exercise that I has a winning strategy iff \( s \) is biologically avoidable. (This holds whether or not \( II \) can see I’s guesses.)
3 Existence of Biologically Avoidable Sequences

One might hope to cleverly generalize the argument from the previous section to non-periodic gender sequences. In this section we’ll show that’s impossible. There are sequences which are biologically avoidable. Populations lacking certain gender sequences are analogous to Aronszajn trees (first introduced in [13]) in the sense that both provide counterexamples to plausible-seeming generalizations of König’s Lemma.

In this and the next section we restrict attention to populations with gendered vertices, implicitly gendering edges according to their initial vertices.

Definition 5. Suppose \( P = (V, E) \) is an infinite \( n \)-gendered population, \( s = (s_1, \ldots, s_n) \) is a finite sequence, \( k > 0 \), and \( V \) is partitioned into heights \( H_1, H_2, \ldots \). We say that \( s \) is impossible in \( P \) at height \( k \) if there is no finite directed path \( v_1, \ldots, v_k \), gendered by \( s \), with \( v_1 \in H_k \).

Lemma 8. Suppose \( s \) is an infinite \( \{1, \ldots, n\} \)-sequence. If there is an infinite \( n \)-gendered population \( P = (V, E) \) and a partition of \( V \) into heights \( H_1, H_2, \ldots \) such that for every \( k > 0 \), some finite restriction of \( s \) is impossible in \( P \) at height \( k \), then \( s \) is biologically avoidable.

Proof. If \( s \) were realized by \( P \), it would be realized by some path, starting with a vertex in some height \( H_k \), yet \( s \) would have some finite restriction impossible at height \( k \), impossible. So \( s \) is not realized by \( P \), so \( s \) is biologically avoidable.

We will now define a specific family of infinite \( 2 \)-gendered populations (generalizing an example suggested by Timothy J. Carlson) designed to take advantage of Lemma 8. Let \( \{M, F\} = \{1, 2\} \), we will refer to \( M \)-gendered vertices as males, \( F \)-gendered vertices as females, and adopt terminology such as son, daughter with the obvious meanings.

Definition 6. Suppose \( h : \mathbb{N}^+ \rightarrow \mathbb{N}^+ \). The infinite \( 2 \)-gendered population \( T_h \) is defined as follows (Figure 1 shows \( T_{n \rightarrow n} \)). The vertices of \( T_h \) are partitioned into successive generations \( G_1, G_2, \ldots \), the \( n \)th generation consisting of \( h(n) \) males \( m_1^n, \ldots, m_{h(n)}^n \) and \( h(n) \) females \( f_1^n, \ldots, f_{h(n)}^n \). These vertices are given birthdates, \( v \mapsto t(v) \), so that vertices in \( G_i \) are born before those in \( G_j \) whenever \( i < j \), and within each generation \( G_n \), \( \max\{t(m_i^n), t(f_i^n)\} < \min\{t(m_j^n), t(f_j^n)\} \) whenever \( i < j \). Edges are defined as follows.

- \( m_1^1 \) and \( f_1^1 \) have no parents.
- \( \forall n > 0, m_{i+1}^n \) has parents \( m_i^n \) and \( f_i^n \).
- \( \forall n > 0, f_{i+1}^n \) has parents \( f_i^n \) and \( m_i^n \).
- \( \forall n > 0, \forall 0 < i < h(n), m_{i+1}^n \) has parents \( m_i^n \) and \( f_i^n \).
- \( \forall n > 0, \forall 0 < i < h(n), f_{i+1}^n \) has parents \( f_i^n \) and \( m_i^n \).

Lemma 9. Let \( h : \mathbb{N}^+ \rightarrow \mathbb{N}^+ \).

1. \( T_h \) is an infinite 2-gendered population with roots \( m_1^1 \) and \( f_1^1 \).
2. If \( m_i^n \) has a daughter, or \( f_i^n \) has a son, then \( i = 1 \).
3. No edge in \( T_h \) skips an entire generation: if an edge has initial vertex in \( G_i \), then it has terminal vertex in either \( G_i \) or \( G_{i+1} \).
4. If \( B = (v_1, \ldots, v_n) \) is a finite directed path of males from \( T_h \), such that \( v_1 \in G_i \) and \( v_n \in G_j \), then for every \( i < k < j \), \( B \) contains all the males in \( G_k \).

Proof. Left to the reader.
Definition 7. When a function \( h : \mathbb{N}^+ \rightarrow \mathbb{N}^+ \) is clear from context, we let \( \hat{m}_1, \hat{m}_2, \ldots \) denote the males in \( T_h \) (over all the generations), ordered ascending by birthdate. Similarly for \( \hat{f}_1, \hat{f}_2, \ldots \). We partition \( T_h \) into heights \( H_1, H_2, \ldots \) by letting each \( H_i = \{ \hat{m}_i, \hat{f}_i \} \).

The following technical lemma should be compared and contrasted with Cauchy’s polygonal number theorem (see [14]) which states that every positive integer can be written as a sum of \( n \) \( n \)-gonal numbers, for any \( n \geq 3 \). For example, every natural number has the form \( \sum_{p=1}^{b_1} p + \sum_{p=1}^{b_2} p + \sum_{p=1}^{b_3} p \). Also worth comparing is the work [1] of D. Cantor and B. Gordin, and more recently [8] of S. Gupta.

Lemma 10. Suppose \( h : \mathbb{N}^+ \rightarrow \mathbb{N}^+ \) and \( \lim_{n \rightarrow \infty} h(n) = \infty \). For every \( u \in \mathbb{N} \), there is some positive integer \( e \) which is not of the form

\[
a + \sum_{p=1}^{b} h(c + p)
\]

for any \( a, b, c \in \mathbb{N} \) with \( c \leq u \) and \( a \leq \max \{ h(1), \ldots , h(u) \} \).

Proof. Let \( A = \max \{ h(1), \ldots , h(u) \} \). Since \( \lim_{n \rightarrow \infty} h(n) = \infty \), there is some \( M_0 \) so big that \( h(c + M) > (u + 1)(A + 1) \) whenever \( M \geq M_0 \) and \( 0 \leq c \leq u \). Let

\[
M = \max \left\{ a + \sum_{p=1}^{b} h(c + p) : a \leq A, c \leq u, b \leq M_0 \right\}
\]

and let \( X = \{ M + 1, M + 2, \ldots , M + (u + 1)(A + 1) \} \).

I claim that for every \( c \leq u \), \( X \) contains at most \( A + 1 \) different numbers of the form \( a + \sum_{p=1}^{b} h(c + p) \) with \( a \leq A \). If not, by the pigeonhole principle there is some particular \( a \leq A \) and some \( b_1 < b_2 \) such that \( a + \sum_{p=1}^{b_1} h(c + p) \) and \( a + \sum_{p=1}^{b_2} h(c + p) \) are both in \( X \), let \( d \) be their difference. Then \( d = h(c + b_1 + 1) + \cdots + h(c + b_2) \geq h(c + b_2) \). Since \( a + \sum_{p=1}^{b_2} h(c + p) > M \) (by virtue of being in \( X \)), by definition of \( M \) this implies \( b_2 \geq M_0 \), whereby \( d \geq h(c + b_2) > (u + 1)(A + 1) \). This is absurd: \( X \) is made up of \( (u + 1)(A + 1) + 1 \) consecutive points, no two of them can have a difference \( > (u + 1)(A + 1) \). The claim is proved.

Given the above claim, the number of numbers in \( X \) with the form \( a + \sum_{p=1}^{b} h(c + p) \), \( a \leq A, c \leq u \), is at most \( (u + 1)(A + 1) \): \( u + 1 \) choices for \( c \), times \( A + 1 \) numbers of the given form for each \( c \). Since \( |X| > (u + 1)(A + 1) \), \( X \) contains an \( e \) as desired. \( \square \)

Proposition 11. If \( h : \mathbb{N}^+ \rightarrow \mathbb{N}^+ \), \( \lim_{n \rightarrow \infty} h(n) = \infty \), \( k > 0 \), and \( s \) is any finite \( \{ M, F \} \)-sequence, there is some \( e > 0 \) such that \( s \sim M^e F \) is impossible in \( T_h \) at height \( k \).

Proof. We may assume \( s \) nonempty. We may also assume the first gender in \( s \) is \( M \), the other case being similar. Let \( \ell = \text{length}(s) \). By Lemma [11] there is some \( e > 1 \) such that \( e - 1 \) is not of the form \( a + \sum_{p=1}^{b} h(c + p) \) for any \( a, b, c \in \mathbb{N} \) with \( c \leq \ell + k \) and \( a \leq \max \{ h(1), \ldots , h(\ell + k) \} \). Thus \( e \) itself is not of the form \( a + 1 + \sum_{p=1}^{b} h(c + p) \) for any such \( a, b, c \). We will show \( s \sim M^e F \) is impossible in \( T_h \) at height \( k \). Suppose not: suppose there is a finite directed path \( v_0, \ldots , v_{\ell}, v_{\ell+1}, \ldots , v_{\ell+e} \) in \( T_h \), gendered by \( s \sim M^e F \), with \( v_0 = \hat{m}_k \).

We would like to estimate in which generation does \( v_\ell \), the first vertex corresponding to the \( M^e F \) block, lie. We will be content with an overestimate. By Lemma [10] part 3, every edge either ends in the same generation where it began, or at most one generation further. Thus \( v_0 = \hat{m}_k \) is in at most the \( k \)th generation, and \( v_\ell \) is in at most the \( (k + \ell) \)th generation.

I claim \( e \) has the form \( a + 1 + \sum_{p=1}^{b} h(c + p) \) for some \( a, b, c \in \mathbb{N} \) with \( c \leq \ell + k \), \( a \leq \max \{ h(1), \ldots , h(\ell + k) \} \), a contradiction. Let \( i, j \) be such that \( v_i \) is in \( G_i \) (so \( i \leq k + \ell \)) and \( v_{\ell+k+1} \) is in \( G_j \). Since \( v_{\ell+k+1} \) is female, and \( v_{\ell+e-1} \) is male, by Lemma [9] part 2, \( v_{\ell+e-1} \) must be \( m^e_1 \). Thus \( B = \{ v_0, \ldots , v_{\ell+e-1} \} \) is a finite directed path of male vertices beginning with \( m^e_1 \) for some \( x \) and ending with \( m^e_1 \). All of these males lie in generations between \( i \) and \( j \) inclusive, and for every \( i < p < j \), Lemma [9] part 4 says that \( B \) contains all the males in \( G_p \). Let us count the vertices in \( B \):

1. The number of males from \( G_i \) included in \( B \) is at most all of them, so \( B \) has \( \leq h(i) \) vertices from \( G_i \).
2. For \( i < p < j \), \( B \) contains all \( h(p) \) males of \( G_p \). This is a total of
\[
h(i + 1) + \cdots + h(i + (j - i - 1)) = \sum_{p=1}^{j-i-1} h(i + p)
\]
vertices.

3. \( B \) has exactly one vertex from \( G_j \), namely \( m^j_i \).

Thus, the number \( e \) of vertices in \( B \) has the form \( a + 1 + \sum_{p=1}^{b} h(c + p) \) for some natural \( c \leq i \leq \ell + k \), some \( a \leq \max\{h(1), \ldots, h(\ell + k)\} \), and some natural \( b \), the desired contradiction. \( \square \)

**Corollary 12.** There is a biologically avoidable sequence.

**Proof.** Let \( h \) be as in Definition 8. Using Proposition 11 repeatedly, define finite sequences \( \{s^k\}_{k>0} \) such that each \( s^{k+1} \) strictly extends \( s^k \) and each \( s^k \) is impossible in \( T_h \) at height \( k \). Then \( s = \bigcup s^k \) is biologically avoidable by Lemma 8. \( \square \)

**Example 13.** \( M^a F M^{2b+1} F \cdots M^{2u} F \cdots \) is biologically avoidable, where each \( \epsilon_n > 0 \) is chosen minimal so as to avoid the form \( a + 1 + \sum_{p=1}^{b} (c + p) \) \((a, c \leq u \text{ where } u = \ell + n, \text{ where } \ell \text{ is the length of } M^a F \cdots M^{2u-1} F)\).

The above example is suboptimal, because our proof of Proposition 11 used such staggering overestimates.

**Example 14.** An alternate way to obtain avoidable sequences is to follow the proof of Corollary 12, but to obtain \( s_{k+1} \), rather than follow the instructions in Proposition 11, we can simply do a brute-force search to find the minimal \( e > 0 \) such that \( s_k \) is impossible at height \( k + 1 \) (Proposition 11 says we won’t get stuck). If we do this for the function \( h(n) = n \), we obtain the avoidable sequence \( M^2 F M^5 F M^8 F M^{11} F \cdots \) where each block of \( M \)'s is 3 longer than the last, with one exception at the beginning (the proof is tedious so we omit it). That one exception is annoying, so here’s how to further optimize the sequence: choose each \( s_k \) to be impossible at height \( k + 1 \). This leaves open the possibility, a priori, the sequence could occur in \( T_h \) at height 1. However, by fortune, it ends up being impossible at height 1 anyway. This yields a very nice well-behaved avoidable sequence,
\[
M^2 F M^5 F \cdots M^{3n-1} F \cdots
\]
(again we omit the formal proof).

All the avoidable sequences we obtain in this manner have the property that they contain arbitrarily long blocks of one gender.

4 **There is a biologically avoidable sequence in which no block of males or females has length more than 2**

Julian Ziegler Hunts discovered an interesting family of populations which witness the avoidability of sequences with very short blocks of males and females.

**Definition 8.** Suppose \( h : \mathbb{N}^+ \rightarrow \mathbb{N}^+ \). The infinite 2-gendered population \( H_h \) is defined in the same way as \( T_h \) was defined (Definition 6) except for its edges, which are instead defined as follows.

- \( m^1_i \) and \( f^1_i \) have no parents.
- \( \forall n > 0, \ m^{n+1}_i \) has parents \( m^n_i \) and \( f^n_{h(n)} \).
- \( \forall n > 0, \ f^{n+1}_i \) has parents \( m^n_{h(n)} \) and \( f^n_{h(n)} \).
- \( \forall n > 0, \forall 0 < i < h(n), \ m^{i+1}_{i+1} \) has parents \( m^i_i \) and \( f^i_i \).
- \( \forall n > 0, \forall 0 < i < h(n), \ f^{i+1}_{i+1} \) has parents \( f^i_i \) and \( m^i_i \).

![Figure 2: The infinite 2-gendered population \( H_{m,n} \).](image)
Lemma 15. Let $h : \mathbb{N}^+ \to \mathbb{N}^+$.

1. $H_h$ is an infinite 2-gendered population with roots $m_1^i$ and $f_1^i$.

2. If $m_i^p$ has a son, then $i = 1$.

3. No edge skips an entire generation: if an edge of $H_h$ has initial vertex in $G_i$ then its terminal vertex is in $G_i$ or $G_{i+1}$.

4. If $B = (v_1, \ldots, v_n)$ is a finite directed-path in $H_h$, whose genders are alternating, and if $v_1$ lies in $G_i$ and $V_n$ lies in $G_j$, then for every $i < p < j$ and every $0 < k \leq h(p)$, precisely one of $\{m_i^p, f_i^p\}$ appears in $B$.

Proof. Left to the reader.

We define the heights of $H_h$, and the corresponding $\widehat{m}_i$ and $\widehat{f}_i$, in the same way as we did for $T_h$.

Proposition 16. Let $h : \mathbb{N}^+ \to \mathbb{N}^+$ be such that $\lim_{n \to \infty} h(n) = \infty$ and $h(n)$ is even for every $n$. If $k > 0$ and $s$ is any finite \{M,F\}-sequence, then there is some $e > 0$ such that $s \sim (FM)^eM$ is impossible in $H_h$ at height $k$.

Proof. We may assume $s$ nonempty and that the first gender in $s$ is $M$. Let $\ell = \text{length}(s)$. Since $h(n)$ is even for every $n$, by Lemma 15 there is an $e > 1$ such that $e - 1$ is not of the form $a + \sum_{p=1}^b \frac{1}{b} h(c + p)$ for any $a, b, c \in \mathbb{N}$ with $c \leq \ell + k$ and $a \leq \max\{h(1)/2, \ldots, h(\ell + k)/2\}$. Thus $e$ itself is not of the form $a + 1 + \sum_{p=1}^b \frac{1}{b} h(c + p)$ for any such $a, b, c$. We will show $s \sim (FM)^eM$ is impossible in $H_h$ at height $k$. If not, there is a finite directed path $v_0, \ldots, v_e, v_{e+1}, \ldots, v_{e+2e}$ in $H_h$, gendered by $s \sim (FM)^eM$, with $v_0 = \widehat{m}_k$.

By similar reasoning to the proof of Proposition 11, $v_e$ is in at most the $(k + \ell)$th generation. I claim $e$ has the form $a + 1 + \sum_{p=1}^b \frac{1}{b} h(c + p)$, some $a, b, c$ as above, a contradiction. Let $i, j$ be such that $v_\ell \in G_i$ (so $i \leq k + \ell$) and $i < p < j$, $v_{e+2e-1} \in G_j$. Since $v_{e+2e-1} \in G_j$ and has a son $v_{e+2e}$, Lemma 15 part 2 ensures $v_{e+2e-1} = m_j^i$. Thus $B = (v_\ell, \ldots, v_{e+2e-1})$ is a finite directed path of alternating gender beginning with $f_\ell^i$ for some $x$ and ending with $m_j^i$. All these vertices lie in generations between $i$ and $j$ inclusive, and for each $i < p < j$, Lemma 15 part 4 implies that the number of vertices in $B \cap G_p$ is exactly $h(p)$. Count the male vertices in $B$:

1. The number of males from $G_i$ is at most half of them (since $B$ alternates genders), that is at most $h(i)/2$.

2. For any $i < p < j$, the number of males from $G_p$ is exactly half of them, $h(p)/2$, by Lemma 15 part 4 since $B$ alternates genders.

3. There is exactly one male from $G_j$, namely $m_j^i$.

Thus the number of males in $B$ is of the form $a + 1 + \sum_{p=1}^b \frac{1}{b} h(c + p)$ ($a, b, c$ as above). But the number of males in $B$ is $e$—absurd.

Corollary 17. There is a biologically avoidable sequence in which no gender occurs thrice in a row.

Proof. Similar to the proof of Corollary 12.

5 Application to cellular automata

In his paper [11], Nathaniel Johnston proved that in certain (a large family) of Conway’s Life-like games, spaceships have an orthogonal speed limit of $1/2$ cells per generation, and a diagonal speed limit of $1/3$ cells per generation. We will give an alternate proof using a technique which we believe, might be generalizable to obtain results of a wide variety [1]. We assume a novice-level familiarity with Life-like games (see [7]), and brush formal details under the rug.

---

1If nothing else, our proof would generalize with minimal changes to Life-like games in higher dimensions (see Bays [3]).
**Definition 9.** Suppose a Life-like game is played, with some initial configuration in generation 1, which generates a configuration in generation 2, and so on. A *lifeline* for this gameplay is a sequence \( c_1, c_2, \ldots \) of cells such that:

1. Each \( c_i \) is alive in generation \( i \).
2. For each \( i \), either \( c_i = c_{i+1} \) or \( c_i \) is adjacent to \( c_{i+1} \).

Thus, a lifeline is a (not necessarily simple) stroll through the cells which, at each \( i \)th step, visits a cell alive in the \( i \)th generation.

**Lemma 18.** (Two Forbidden Directions) Suppose \( x, y \) are any two of the following *directions*:

\[
N, S, E, W, NE, NW, SE, \text{ and } SW.
\]

Consider a Life-like gameplay with the following properties:

1. The initial configuration is finite.
2. Each generation contains at least one live cell.
3. Birth requires \( \geq 3 \) neighbors and survival requires \( \geq 1 \) neighbor.

For any such gameplay, there is a lifeline which never steps in direction \( x \) or \( y \) (that is, \( x_{i+1} \) is never located in the \( x \) or \( y \) direction from \( x_i \)).

To be clear, the third condition is to be understood as liberally as possible, making the lemma apply not only to the ruleset \( B345678/S12345678 \), but to any sub-ruleset thereof.

We postpone the proof (using biologically unavoidable sequences) of Lemma 18 so we can first see how it applies to spaceship speed limits.

**Theorem 19.** (Johnston 2010) For any Life-like ruleset where birth requires \( \geq 3 \) neighbors and survival requires \( \geq 1 \) neighbor, spaceships can move at most \( \frac{1}{2} \) cells per generation orthogonally and at most \( \frac{1}{3} \) cells per generation diagonally.

**Proof sketch.** Since a spaceship is initially finite and does not go extinct, the hypotheses of Lemma 18 are met.

(Orthogonal) By symmetry, it’s enough to show the spaceship cannot exceed \( \frac{1}{2} \) cells per generation northward. By Lemma 18, there is a lifeline which never steps in direction N or NE. The cells in this lifeline are living cells, hence cells in the spaceship, and it follows that the spaceship cannot travel faster than the lifeline. The only movement the lifeline can make with northward component is NW (N and NE being forbidden). Any such step also moves the lifeline westward, and so to maintain an overall northward direction, any such step must be compensated for by a step in one of the directions E or SE (NE is forbidden). So at least two total steps are required per unit of overall northward movement. Thus the speed limit, \( \frac{1}{2} \).

(Diagonal) By symmetry, it’s enough to show the spaceship cannot exceed \( \frac{1}{3} \) cells per generation north-eastward. By Lemma 18, there is a lifeline which never steps N or NE. The only way the lifeline may move northward is by moving NW, and two eastward steps must be added to produce overall NE movement. Thus the speed limit, \( \frac{1}{3} \).

**Proof sketch of Lemma 18.** Let \( V \) be the set of pairs \((c, i)\) such that \( c \) is a cell alive in generation \( i \). Direct an edge from \((c, i)\) to \((d, j)\) if \( j = i + 1 \) and either \( c = d \) or \( c \) is a neighbor of \( d \). By (3) of Lemma 18, if \( d \) is a cell born into generation \( i + 1 \), there must have been at least three distinct neighbors \( c_1, c_2, c_3 \) of \( d \) alive in generation \( i \). There are only two forbidden directions, so (possibly relabelling) we may assume \( d \) does not lie in a forbidden direction from \( c_1 \). Gender the edge \(((c_1, i), (d, i + 1))\) male, and gender all other edges terminating in \((d, i + 1)\) female. If \( d \) survives into generation \( i + 1 \), then gender the edge \(((d, i), (d, i + 1))\) male and gender all other edges terminating in \((d, i + 1)\) female. Let each vertex \((c, i)\) have birthdate \( i \). The reader may check (using the hypotheses of Lemma 18) this makes \( V \) an infinite 2-gendered population. By Theorem 6, there is an infinite directed path through this population with all edges male. By construction, male edges never step in a forbidden direction.
6 Further Questions

If $s$ is an avoidable gender sequence, to what extent can we find populations avoiding $s$ which are universal among all such, in a way analogous to that described by Cherlin and Shelah [5]?

Given a possibly avoidable sequence, can the infinite gendered populations which realize that sequence be characterized in some way? Of particular interest would be a characterization in terms of ordinal numbers, similar to R. Schmidt’s characterization of rayless graphs (see Halin [9]).

But perhaps the most important question remaining is, what are the biologically unavoidable sequences? Are there any which are not eventually periodic?

References

[1] Alexander S (2011) On Guessing Whether a Sequence has a Certain Property. J. Int. Seq. 14. 12 pages.
[2] Alexander S (preprint) Infinite graphs in systematic biology, with an application to the species problem. To appear in Acta Biotheoretica. http://arxiv.org/abs/1201.2869
[3] Bays C (1987) Candidates for the Game of Life in Three Dimensions. Comp. Syst. 1: 373–400.
[4] Cantor D, Gordin B (1973) Sequences of Integers with Missing Differences. J. Comb. Theor. Series A 14: 281–287.
[5] Cherlin G, Shelah S (2007) Universal graphs with a forbidden subtree. J. Comb. Theor. Series B 97: 293–333.
[6] Dress A, Moulton V, Steel M, Wu T (2010) Species, clusters and the ‘Tree of life’: A graph-theoretic perspective. J. Theor. Biol. 265: 535–542.
[7] Eppstein D (2002) Searching for Spaceships. MSRI Publications 42, 433–453. http://arxiv.org/abs/cs/0004003
[8] Gupta S (2000) Sets of Integers with Missing Differences. J. Comb. Theor. Series A 89: 55–69.
[9] Halin R (1998) The Structure of Rayless Graphs. Abh. Math. Sem. Univ. Hamburg 68: 225–253.
[10] Hennig W (1966) Phylogenetic Systematics. Urbana, IL: University of Illinois Press.
[11] Johnston N (2010) The B36/S125 “2 × 2” Life-Like Cellular Automaton. In: Adamatzky A, editor. Game of Life Cellular Automata. Springer-Verlag.
[12] Kornet D, Metz J, Schellinx H (1995) Internodons as equivalence classes in genealogical networks: building-blocks for a rigorous species concept. J. Math. Bio. 34: 110–122.
[13] Kurepa D (1935) Ensembles ordonnés et ramifiés. Doctoral dissertation. Publ. Math. Univ. Belgrade 4: 1–138.
[14] Nathanson M (1987) A Short Proof of Cauchy’s Polygonal Number Theorem. Proc. Amer. Math. Soc. 99: 22–24.
[15] Simpson S (1982) Subsystems of Second Order Arithmetic. 2nd Edition.
[16] Sturmfels B (2005) Can biology lead to new theorems? Annual report of the Clay Mathematics Institute.
[17] Wadge W (1983) Reducibility and Determinateness on the Baire Space. Doctoral dissertation, UC Berkeley.