What can oracles teach us about the ultimate fate of life?

Ville Salo
University of Turku, Finland

Ilkka Törmä
University of Turku, Finland

Abstract

We settle two long-standing open problems about Conway's Life, a two-dimensional cellular automaton. We solve the Generalized grandfather problem: for all $n \geq 0$, there exists a configuration that has an nth predecessor but not an $(n+1)$st one. We also solve (one interpretation of) the Unique father problem: there exists a finite stable configuration that contains a finite subpattern that has no predecessor patterns except itself. In particular this gives the first example of an unsynthesizable still life. The new key concept is that of a spatiotemporally periodic configuration (agar) that has a unique chain of preimages; we show that this property is semidecidable, and find examples of such agars using a SAT solver.

Our results about the topological dynamics of Game of Life are as follows: it never reaches its limit set; its dynamics on its limit set is chain-wandering, in particular it is not topologically transitive and does not have dense periodic points; and the spatial dynamics of its limit set is non-sofic, and does not admit a sublinear gluing radius in the cardinal directions (in particular it is not block-gluing). Our computability results are that Game of Life's reachability problem, as well as the language of its limit set, are PSPACE-hard.

2012 ACM Subject Classification Theory of computation → Formal languages and automata theory

Keywords and phrases Game of Life, cellular automata, limit set, symbolic dynamics

Funding *Ville Salo*: Research supported by Academy of Finland grant 2608073211.

Acknowledgements We thank the anonymous referees for their useful suggestions, and the nice people of the ConwayLife forum for their interest.

1 Introduction

Conway's Game of Life is a famous two-dimensional cellular automaton defined by John Horton Conway in 1970 and popularized by Martin Gardner [14]. A cellular automaton can be thought of as zero-player game: the board is set up, and a simple rule determines the dynamics. In the case of Game of Life, the board is the two-dimensional infinite grid, where some grid cells are *live*, and some are *dead* (or *empty*); the evolution rule, executed simultaneously in all cells, is that a dead cell becomes live if it has exactly three live (cardinal or diagonal) neighbors, and a live cell stays live if and only if it has two or three live neighbors.

Iterating this rule gives rise to very complicated dynamics. Engineering patterns with interesting behaviors, and searching for such patterns by computer, has been an ongoing effort since the invention of the rule. For readers interested in delving into this world, we cite the very recent (and freely available) book [19] of Johnston and Greene. One result that exemplifies the complexity of Game of Life is that it is *intrinsically universal* [10], meaning that Game of Life can simulate any two-dimensional cellular automaton f (including proper self-simulation), so that the states of f correspond to large blocks with special content, and one step of f is simulated in multiple steps of Game of Life.

Game of Life can be thought of as a mathematical *complex system*, namely it is a system where complex global behavior arises from interacting (simple) local rules. Such systems can be notoriously difficult to study. We can often use computer simulations to make empirical observations about typical and eventual behavior, but it can be very difficult to actually prove that a particular behavior persists on larger scales (even if it seems like its failure would require a massive conspiracy). Usually one can only successfully analyze systems that are very simple [13], or their behavior simulates a phenomenon that is mathematically well-understood, say of an algebraic [8] or number theoretic [20] nature, or the systems are specifically constructed for some purpose [23]. Due to intrinsic universality, it seems unlikely that Game of Life fits in any of these classes.

Indeed, for Game of Life, despite decades of study by enthusiasts, almost no non-trivial mathematical results exist that state limitations on its eventual behavior. In other words, as a dynamical system, we know very little about it. From computer simulations, one can conclude that Game of Life is highly "chaotic", and one can make educated guesses about things like the typical population density after a large number of iterations; however, it is very hard to make such claims rigorous. Rigorous results about Game of Life do exist, but they concern mostly the behavior of Game of Life on nice configurations (the engineering feats discussed above are of this type), and no known pattern behaves predictably in a general context; alternatively, they deal with one-step or static behavior [12].

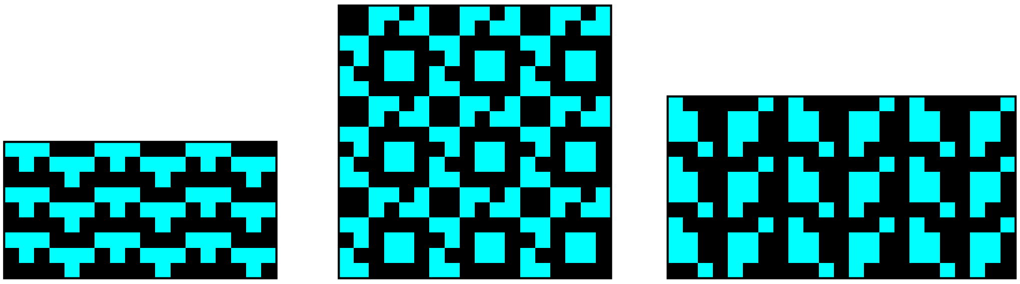
In this paper, we study Game of Life through its *agars*, which are the Game of Life community's term for spatiotemporally periodic points. More specifically, we observe that a simple algorithm (essentially Wang's partial algorithm from [32]) can be used to find all agars with small enough periodicity parameters that have a unique chain of predecessors. We then study finite patches of these agars, and find some with interesting backwards forcing properties. Namely, these patterns behave deterministically in the (a priori nondeterministic) backwards dynamics of Game of Life. This intuitively allows us to study the "last iterations" of Game of Life (after an unknown number of steps), and leads to a wealth of results about how Game of Life behaves "in the limit".

One practical difficulty is that the algorithm we use is not of the usual kind, but rather it is in the class $\mathrm{FP}^{\mathrm{NP}}$ of problems that are solvable in polynomial time with an NP oracle. Throughout this work, modern SAT solvers have constantly impressed and even humbled us by how freely they can be used as such oracles.[1] While their role is not very explicit in the final write-up of the paper, this work would not have been possible without them.

We do not expect the method of studying the eventual dynamics through self-enforcing patterns to be specific to Game of Life. Indeed, one can apply it directly to any cellular automaton rule (with any number of dimensions), and the idea can presumably be adapted to other systems as well. The reason we study a single example cellular automaton is that the results require us to find a "witness", usually a self-enforcing agar, and there is no *a priori* bound on how long this pattern-crunching will take – or whether it will succeed at all – for a particular rule. A single agar also tends to only work for a single rule or trivial modifications thereof. The choice of precisely Game of Life as the example rule to study is not mathematically motivated, it is simply a well-known simple rule that has already been studied extensively. Some of our programs are available on GitHub at [30], for readers interested in trying the methods out on other rules.

[1] For example, in our experience, general-purpose constraint-solvers and our own CA-specific solvers often fail even on the basic problem of finding a Game of Life preimage, while SAT solvers happily tell us, say, whether a preimage exists with particular constraints, and can find preimages for higher powers of Life.

1.1 The protagonists



(a) Köynnös. **(b)** Kynnös. **(c)** Marching band.

Figure 1 Patches of the agars. A 3-by-3 grid of the repeating patterns is shown for each. Cyan cells are live.

We begin with a brief discussion of the agars that we use to prove our results. These will be explained in more detail in separate sections.

Figure 1a shows the pattern we call köynnös[2]. The infinite agar obtained by repeating this pattern infinitely in each direction has no preimage other than itself; we say it is *self-enforcing*. Up to symmetries there are exactly 11 self-enforcing agars of size 6×3. Köynnös has the special property that it is impossible to stabilize a finite difference to this configuration: if one modifies the agar in finitely many cells, the difference spreads at the speed of light (one column of cells per time step, which is the maximal speed at which information can be transmitted by Game of Life). We say it *cannot be stabilized from the inside*.

Figure 1b shows the pattern we call kynnös[3]. The corresponding infinite agar is again self-enforcing. Up to symmetries there are at least 52 self-enforcing agars of size 6×6 (we were unable to finish the search, so it is possible that more exist). Kynnös has two special properties. First, it contains a finite patch such that if a configuration has this patch in its image, then the configuration already had that patch in place, i.e. one cannot synthesize it from any other patch. Second, unlike köynnös, it can be stabilized from the inside.

Figure 1c shows the pattern we call marching band[4]. This agar has temporal period two. Its most important property is that an infinite south half-plane of this pattern must shrink if there is a difference on its border, meaning that in the nondeterministic inverse dynamics of Game of Life, an infinite south half-plane of this pattern "marches" to the north.

To find the marching band, we searched through all $w \times h$-patterns such that the corresponding agar with periods $(w, 0)$ and $(0, h)$ is temporally (exactly) t-periodic, for the parameter range $2 \leq w \leq 9$, $2 \leq h \leq 5$, $2 \leq t \leq 3$. There were no self-enforcing agars with temporal period 3 in this range, and there were exactly 14 self-enforcing agars with period 2. The marching band is the only one that has the marching property in any direction.

1.2 Results

Denote by $g : \{0,1\}^{\mathbb{Z}^2} \to \{0,1\}^{\mathbb{Z}^2}$ the Game of Life cellular automaton, where dead cells are represented by 0 and live cells by 1. In this section, we list all our new technical contributions

[2] Finnish for vine.
[3] Finnish for that which is tilled.
[4] English for marssiorkesteri.

about g. The reader should consult Section 2 for precise definitions of terms used in this section. First, we solve the Generalized grandfather problem: for all $n \geq 0$, there exists a configuration that has an nth predecessor but not an $(n+1)$st one.

▶ **Theorem 1** (Generalized grandfather problem). *For each $n \geq 0$, there exists $x \in \{0,1\}^{\mathbb{Z}^2}$ with $g^{-n}(x) \neq \emptyset$ and $g^{-(n+1)}(x) = \emptyset$.*

The case of $n = 0$ (that g is not surjective) was resolved by R. Banks in 1971, only a year after the introduction of Game of Life. Conway stated the Grandfather problem, namely the case $n = 1$ of the above, in 1972, and promised \$50 in the *Lifeline* newsletter [31] for its solution. This stayed open until 2016, when the cases $n \in \{1, 2, 3\}$ were proved by the user mtve of the ConwayLife forum. We note (see Lemma 16 for the proof) that while Theorem 1 refers to infinite configurations, the analogous statements for finite patterns or finite-population configurations are equivalent to it. Cellular automata satisfying the conclusion of Theorem 1 are sometimes called "unstable" [24], though we avoid this terminology here, as "stable" has another meaning in Game of Life jargon.

More specifically, we prove the following two results, which strengthen Theorem 1 in different directions. The first result is proved using köynnös and is based on the fact it cannot be stabilized from the inside. The notation $g^{-n}(p)$ for a finite pattern p of shape $D \subset \mathbb{Z}^2$ stands for the set of patterns of shape $D + [-n, n]^2$ that evolve into p in n steps.

▶ **Theorem 1.1.** *There exists a polynomial time algorithm that, given $n \geq 0$ in unary, produces a finite pattern p with $g^{-n}(p) \neq \emptyset$ and $g^{-(n+1)}(p) = \emptyset$.*

The algorithm is very simple: change the value of one cell in the agar, apply the Game of Life rule n times, and pick the central $[-30 - 6n, 30 + 6n] \times [-27 - 8n, 27 + 8n]$-patch of the resulting configuration as p. An example with $n = 28$ is shown in Figure 2 (with insufficient padding: the periodic background should extend 164 cells further to the left and right, and 220 cells up and down).

Figure 2 A "level-29 orphan" obtained by perturbing köynnös: these angry deities could be found 28 seconds after the Big Bang, then went extinct.

The second result is proved using kynnös, and is based on the facts that kynnös admits a self-enforcing patch and that it can be stabilized from the inside. The following was first pointed out by Adam Goucher [16].

▶ **Theorem 1.2.** *For any large enough n, there exists an $n \times n$ pattern which appears in the kth image of Game of Life, but does not appear in its $(k+1)$st image, where* $2^{n^2/368-O(n)} \leq k \leq 2^{n^2}$.

This is (up to a suitable equivalence relation) the optimally slow growth rate for higher level orphans. The same idea can be used to obtain the following results about the limit set of Game of Life. Also called the eventual image, it is the set of configurations with arbitrarily long chains of predecessors. The language of the limit set refers to the set of finite patterns occurring in it.

▶ **Theorem 2.** *The limit set of Game of Life has PSPACE-hard language.*

The language might well be much harder. Even for one-dimensional cellular automata it can be Π^0_1-complete [6, 18]; we do not know if Game of Life reaches this upper bound.

We also obtain information about the symbolic dynamical nature of the limit set. A set of configurations is sofic if it can be defined by Wang tiles, or squares with colored edges: in a valid tiling of $\mathbb{Z}^2$, colors of adjacent edges are required to match, and the tiles can additionally be marked with 0 and 1 to project each valid tiling to a binary configuration. The set of those projections is called a sofic shift. Sofic systems form a large and varied class of subshifts, for example their one-dimensional projections can be essentially arbitrary (subject only to an obvious computability condition) [11, 1]. We show that the limit set of Game of Life cannot be defined by a tile set in this way.

▶ **Theorem 3.** *The limit set of Game of Life is not sofic.*

Besides illuminating the iterated images of Game of Life and its limit set, the self-enforcing kynnös patch itself solves a second open problem, namely the Unique father problem stated by John Conway in [31, 5]: is there a still life (a finite-population configuration that is a fixed point of g) whose only predecessor is itself, "with some fading junk some distance away not being counted"? We solve one interpretation of this problem.

▶ **Theorem 4** (Unique father problem)**.** *There exists a finite still life configuration x that contains a finite subpattern p such that every preimage of x also has subpattern p.*

One can also imagine stronger variants of the Unique father problem: for example, we could require p to contain all live cells of x, or all cells in their convex hull. These stay open.

Theorem 4 also tells us something about the dynamics of Game of Life *restricted to* its limit set, i.e. its limit dynamics. The chain-wandering property essentially means that there is a finite pattern that occurs in the limit set of Game of Life, but never returns to itself under the dynamics no matter how we fill the surrounding infinite plane. In fact, we are even allowed to completely rewrite the entire configuration on every step, apart from the domain of the pattern.

▶ **Theorem 5.** *Game of Life is chain-wandering on its limit set.*

Much is known about the kinds of things that can happen in Game of Life orbits, in particular it is well known that Game of Life is computationally universal and can simulate any cellular automaton. Nevertheless, to our knowledge all existing methods of simulating unbounded computation require the rest of the configuration to be empty (or at least stay out of the way). With our methods, we can enforce computations in a finite region (conditioned on its end state) even when it is completed into an infinite configuration by an adversary.

▶ **Theorem 6.** *The reachability problem of Game of Life is PSPACE-hard, i.e. given two finite patterns $p, q \in \{0,1\}^D$ whose domain D has polynomial extent, it is PSPACE-hard to tell whether there exists a configuration x with $x|D = p$ and $g^n(x)|D = q$ for some $n \geq 0$.*

Finally, the properties of the marching band's backwards dynamics imply that the limit set contains patterns that cannot be "glued" together too close: there are two $n \times n$ patterns such that no configuration of the limit set contains both of them separated by a distance less than $n/15$.

▶ **Theorem 7.** *For all large enough n there exist patterns $p, q \in \{0,1\}^{[0,n-1]^2}$ such that p and q appear in the limit set, but $p \sqcup \sigma_{(0,\lfloor n/15 \rfloor)}(q)$ does not.*

▶ **Corollary 8.** *The limit set of Game of Life is not block-gluing (thus has none of the gluing properties listed in [4]).*

1.3 Programs

Some of the programs we used can be found on GitHub at [30]. We have included Python scripts enumerating self-enforcing agars, and scripts checking the claimed properties of our three agars. In particular one can find implementations of Algorithms 1 and 2. The scripts use the PySAT [28] library to call the Minisat [26] SAT solver (the library supports many other solvers as well).

2 Definitions

Our intervals are discrete. To simplify formulas, we denote by

$$\begin{bmatrix} a & b & c & d \\ e & f & g & h \end{bmatrix} = ([-a,b] \times [-c,d]) \setminus ([-e,f] \times [-g,h])$$

a rectangular discrete annulus when the second rectangle fits fully inside the first, that is, $-a \leq -e \leq f \leq b$ and $-c \leq -g \leq h \leq d$.

We assume some familiarity with topological and symbolic dynamics and give only brief definitions, see e.g. [22] for a basic reference. We denote by S a finite *alphabet*. A *configuration* or *point* is an element of $S^{\mathbb{Z}^d}$. More generally, a *pattern* (or sometimes *patch* in more informal contexts) is a function $p : \mathrm{dom}(p) \to S$, where $\mathrm{dom}(p) \subset \mathbb{Z}^d$ is the *domain* of p. If $S \subset \mathbb{N}$, then by $\sum p$ we denote the sum $\sum_{\vec{v} \in \mathrm{dom}(p)} p(\vec{v})$. For $\vec{v} \in \mathbb{Z}^d$, a pattern p and $D \subset \mathbb{Z}^d$, we write $q = p|D$ for the restriction $\mathrm{dom}(q) = D \cap \mathrm{dom}(p), q(\vec{v}) = p(\vec{v})$. A pattern is *finite* if its domain is, and a configuration is *finite* if its sum as a pattern is finite. If p, q are patterns with disjoint domains, define $p \sqcup q = r$ by $\mathrm{dom}(r) = \mathrm{dom}(p) \cup \mathrm{dom}(q), r|\mathrm{dom}(p) = p, r|\mathrm{dom}(q) = q$. The *extent* of a pattern is the minimal hypercube containing the origin and its domain. For two patterns, write $\mathrm{eq}(q, q')$ for the set of vectors $\vec{v} \in \mathrm{dom}(q) \cap \mathrm{dom}(q')$ such that $q(\vec{v}) = q'(\vec{v})$, and $\mathrm{diff}(q, q')$ for those that satisfy $q(\vec{v}) \neq q'(\vec{v})$. For computer science purposes, we note that patterns with polynomial extent have an efficient encoding as binary strings.

The *full shift* is the set of all configurations $S^{\mathbb{Z}^d}$ with the product topology (where S has the discrete topology), under the action of $\mathbb{Z}^d$ by homeomorphisms $\sigma_{\vec{v}}(x)_{\vec{u}} = x_{\vec{v}+\vec{u}}$ called *shifts*. We use the same formula to define $\sigma_{\vec{v}}(p)$ for patterns p (of course shifting the domain correspondingly). A pattern p defines a *cylinder* $[p] = \{x \in S^{\mathbb{Z}^d} \mid x|\mathrm{dom}(p) = p\}$. Cylinders defined by finite patterns are a base of the topology, and their finite unions are exactly the clopen sets. The *symbol partition* is the clopen partition $\{[s] \mid s \in S\}$ where s is identified with the pattern $p : \{\vec{0}\} \to S$ with $p(\vec{0}) = s$. The space $S^{\mathbb{Z}^d}$ is homeomorphic to the Cantor space, and is metrizable. One possible metric is $\mathrm{dist} : (S^{\mathbb{Z}^2})^2 \to \mathbb{R}$, $\mathrm{dist}(x, y) = 2^{-\sup\{n \mid x|[-n,n]\times[-n,n]=y|[-n,n]\times[-n,n]\}}$ with $2^{-\infty} = 0$.

A *cellular automaton* (or *CA*) is a continuous self-map $f : S^{\mathbb{Z}^d} \to S^{\mathbb{Z}^d}$ that commutes with the shifts. The *neighborhood* is a set $N \subset \mathbb{Z}^d$ such that $f(x)_{\vec{0}}$ is determined by $x|N$; a finite neighborhood always exists by the Curtis-Hedlund-Lyndon theorem [17]. It is easy to show that there is always a unique *minimal neighborhood* under inclusion. A state $0 \in S$ is *quiescent* if $f(0^{\mathbb{Z}^d}) = 0^{\mathbb{Z}^d}$. A *subshift* is a closed subset X of $S^{\mathbb{Z}^d}$ invariant under shifts. Its *language* is the set of finite patterns p such that $[p] \cap X \neq \emptyset$, and we say these patterns *appear* or *occur* in the subshift. Patterns that do not appear in $f(S^{\mathbb{Z}^d})$ are usually called *orphans*. We say p is a *level-n orphan* if it appears in $f^{n-1}(S^{\mathbb{Z}^d})$ but not in $f^n(S^{\mathbb{Z}^d})$ (so the usual orphans are level-1). The *limit set* of a cellular automaton f is $\Omega(f) = \bigcap_n f^n(S^{\mathbb{Z}^d})$. It is a subshift invariant under f. A *subshift of finite type* is a subshift of the form $\bigcap \sigma_{\vec{v}}(C)$ where C is clopen. A *sofic shift* is a subshift which is the image of a subshift of finite type under a shift-commuting continuous function.

We are mainly interested in $d = 2$, $S = \{0, 1\}$, and the *Game of Life* cellular automaton $g : \{0,1\}^{\mathbb{Z}^2} \to \{0,1\}^{\mathbb{Z}^2}$ defined by

$$\begin{aligned} g(x)_{\vec{v}} = 1 \iff & (x_{\vec{v}} = 0 \wedge \sum(x|\vec{v} + K) = 3) \\ & \vee (x_{\vec{v}} = 1 \wedge \sum(x|\vec{v} + K) \in \{2, 3\}), \end{aligned}$$

where $K = [-1, 1]^2 \setminus \{(0, 0)\}$.

A *fixed point* (of a CA f) is $x \in S^{\mathbb{Z}^d}$ such that $f(x) = x$. In the context of Game of Life these are also called *stable configurations* or *still lifes*. *Spatial* and *temporal* generally refer respectively to the $\mathbb{Z}^d$-action of shifts and the action of a CA. In particular a *spatially periodic point* is a configuration $x \in S^{\mathbb{Z}^d}$ which has a finite orbit under the shift dynamics, and *temporal periodicity* means $f^n(x) = x$ for some $n \geq 1$. Spatiotemporal periodicity means that both hold; in the Game of Life context spatiotemporal points are also called *agars*.

If $f : X \to X$ is a continuous function, an *ϵ-chain from x to y* is $x = x_0, x_1, \ldots, x_k = y$ with $k \geq 1$ such that $\mathrm{dist}(f(x_i), x_{i+1}) < \epsilon$ for $0 \leq i < k$. We say f is *chain-nonwandering* if for all $\epsilon > 0$ and $x \in X$ there is an ϵ-chain from x to itself; otherwise f is *chain-wandering*. (In the literature, chain-nonwandering is more commonly known as chain-recurrence, but both terms are logical.) We say f is *topologically transitive* if for all nonempty open sets U, V we have $f^n(U) \cap V \neq \emptyset$ for some n. It is *sensitive* (to initial conditions) if there exists $\epsilon > 0$ such that for all $x \in X$ and $\delta > 0$ there exists $y \in X$ with $\mathrm{dist}(x, y) < \delta$ and $n \in \mathbb{N}$ such that $\mathrm{dist}(f^n(x), f^n(y)) \geq \epsilon$. We say f has *dense periodic points* if its set of temporally periodic points is dense.

3 Proofs

We begin by introducing a formalism for forced cells in the preimages of a given pattern or configuration. The general topological idea is the following: if we have a zero-dimensional space X and a family of closed sets $\mathcal{I}$ which is closed under arbitrary intersections and contains the empty set, then to any continuous $f : X \to X$ we can associate a map $\hat{f} : \mathcal{I} \to \mathcal{I}$ by

$$\hat{f}(A) = \bigcap \{B \in \mathcal{I} \mid f^{-1}(A) \subset B\}. \tag{1}$$

We call this the *dual map* of f with respect to $\mathcal{I}$.

In our situation, $X = S^{\mathbb{Z}^d}$ and $f : S^{\mathbb{Z}^d} \to S^{\mathbb{Z}^d}$ is a cellular automaton. We define three families of subsets of $S^{\mathbb{Z}^d}$:

- $\mathcal{I}$ consists of the cylinders $[p] \subset S^{\mathbb{Z}^d}$ defined by all patterns p, plus the empty set $\emptyset$, which we denote by $\top$. The entire space $S^{\mathbb{Z}^d}$, which is the cylinder of the empty pattern, is denoted by $\bot$.
- $\mathcal{F} \subset \mathcal{I}$ consists of all cylinders $[p]$ defined by finite patterns p.
- $\mathcal{C} \subset \mathcal{I}$ consists of the singletons $[x] = \{x\}$ for full configurations $x \in S^{\mathbb{Z}^d}$.

Note that $\mathcal{F} \cap \mathcal{C} = \emptyset$. The family $\mathcal{F}$ is naturally stratified into finite subsets $\mathcal{F}_M = \{[p] \mid p \in S^M\}$, where $M \subset \mathbb{Z}^d$ ranges over finite sets. For a cylinder $[p] \in \mathcal{I}$, equation (1) defines $\hat{f}([p]) \in \mathcal{I}$ as the cylinder $[q]$, where q contains exactly those cells whose values are the same in all f-preimages of p, or $\top$ if p has no f-preimages (i.e. is an orphan). Also, $\hat{f}(\top) = \top$.

▶ **Example 9.** Consider $S = \{0, 1, 2\}$ and the cellular automaton $f : S^{\mathbb{Z}} \to S^{\mathbb{Z}}$ defined by

$$f(x)_0 = \begin{cases} 2, & \text{if } x_0 = 2, \\ \min(x_0, x_1), & \text{otherwise.} \end{cases}$$

The minimal neighborhood of f is $N = \{0, 1\}$. The pattern $p = 002$ of domain $\{0, 1, 2\}$ has preimages $f^{-1}(p) = \{0020, 0021, 0022, 1020, 1021, 1022\}$ of domain $\{0, 1, 2, 3\}$. Thus $\hat{f}([p]) = [q]$, where $q = 02$ has domain $\{1, 2\}$, since the values of these cells are the same in all preimages. The pattern $p' = 102$ has no preimages, so $\hat{f}([p']) = \top$.

We define a partial order on $\mathcal{I}$ by $[p] \leq [q]$ whenever $[q] \subset [p]$, and $\alpha \leq \top$ for all $\alpha \in \mathcal{I}$. The intuition is that $[p] \leq [q]$ corresponds to the pattern q specifying more cells than p, and thus containing more information. As the empty set $\top$ in a sense specifies the maximal amount of information – a contradiction – it is the largest element. Note that $\mathcal{C}$ consists of the maximal elements of $\mathcal{I} \setminus \{\top\}$.

We give $\mathcal{I}$ the topology with basis sets $U_p = \{\alpha \in \mathcal{I} \mid [p] \leq \alpha\}$ for $[p] \in \mathcal{F}$ as well as $\{\top\}$, making $\top$ an isolated point. This space is not Hausdorff (T_2), indeed it only satisfies the Kolmogorov (T_0) separation axiom. The induced topology on $\mathcal{C}$ is the standard compact Cantor topology, and $\mathcal{F}$ is a dense subset of $\mathcal{I} \setminus \{\top\}$. Every nonempty open set contains $\top$: "the contradiction is dense".

▶ **Lemma 10.** *The dual map $\hat{f} : \mathcal{I} \to \mathcal{I}$ is continuous.*

We are simply saying that if a (possibly infinite) pattern forces some particular value in some cell in the preimage, then actually some finite patch already forces it. The proof is a straightforward compactness argument.

Proof. Continuity at $\top$ is obvious. We show continuity at a cylinder $[p] \in \mathcal{I}$. Suppose first that $\hat{f}([p]) = [q]$, and let $[r] \in \mathcal{F}$ be such that $[q] \in U_r$. This means that p forces the pattern q in its f-preimages, and r is a finite subpattern of q. There exists a finite subpattern s of p that forces r, for otherwise we could take larger and larger subpatterns of p along with two preimages that disagree on $\mathrm{dom}(r)$, and in the limit obtain two preimages of p that disagree on $\mathrm{dom}(r)$. Hence $\hat{f}(U_s) \subset U_r$.

Suppose then that $\hat{f}([p]) = \top$, meaning that p is an orphan. It is well known that p contains a finite subpattern r that is also an orphan. Then $\hat{f}(U_r) = \{\top\}$. ◀

We list some other easy properties of $\hat{f}$. For $\alpha, \beta \in \mathcal{I}$ write $\alpha \parallel \beta$ for $\alpha \cap \beta \neq \top$. In the case of cylinders, this means that the corresponding patterns agree on the intersection of their domains. For a pattern p, write $f(p)$ for the pattern obtained by applying the local rule of f (with the minimal neighborhood) in every position whose neighborhood is contained in the domain of p (and only those positions are included in the domain of $f(p)$). Note that with this definition $f([p]) \subset [f(p)]$, and the inclusion may be strict.

▶ **Lemma 11.** ▬ *$\mathcal{F} \cup \{\top\}$ is preserved under $\hat{f}$. Indeed, we have $\hat{f}(\mathcal{F}_M) \subset \mathcal{F}_{M+N} \cup \{\top\}$ where N is the minimal neighborhood of f.*

- *$\hat{f}$ is monotone, i.e. $\alpha \le \beta$ implies $\hat{f}(\alpha) \le \hat{f}(\beta)$.*
- *$\hat{f} \circ \hat{g} \le \widehat{f \circ g}$ pointwise.*
- *for all $\alpha \in \mathcal{I}$, either $\hat{f}(\alpha) = \top$ or $\hat{f}(\alpha) = [p]$ with $[f(p)] \le \alpha$; in particular $[f(p)] \parallel \alpha$ in the latter case.*
- *for all $\alpha \in \mathcal{I}$, we have $\hat{f}(\sigma_{\vec{v}}(\alpha)) = \sigma_{\vec{v}}(\hat{f}(\alpha))$.*

The next few results refer to $\mathrm{FP}^{\mathrm{NP}}$, the class of function problems solvable in deterministic polynomial time with the help of an oracle that can solve an NP decision problem in one step. Of course, the oracle can be invoked repeatedly to construct NP certificates in a polynomial number of steps. This class naturally captures the method of using SAT solvers as black boxes to compute preimages of finite patterns.

▶ **Lemma 12.** *For a fixed CA f, given $p \in \mathcal{F}$, the image $\hat{f}([p])$ can be computed in $\mathrm{FP}^{\mathrm{NP}}$. It remains computable if f is also given as input.*

Proof. Since $\hat{f}(\mathcal{F}_M) \subset \mathcal{F}_{M+N}$, we only need to determine which coordinates in $M + N$ are forced in preimages. This requires at most $1 + |M + N|$ calls to an NP oracle: one to request a preimage, and for each $\vec{v} \in M + N$, one to request a pair of preimages which differ at $\vec{v}$. ◀

The proof above is the easiest way to get the theoretical result, but for practical purposes we give Algorithm 1, which tends to find the $\hat{f}$-image much quicker (and is just as quick to implement). It is written for an "incremental oracle", meaning we can only *add* constraints to it (represented by the set F) when we make a new query. In this case, we compute a single f-preimage q of the input pattern p, and then compute additional preimages that differ from q on progressively smaller sets of cells. Modern SAT solvers tend to support such incremental access – of course, on the side of theory it is easy to see that the class $\mathrm{FP}^{\mathrm{NP}}$ is the same whether or not queries are restricted to be incremental.

Algorithm 1 Finding $\hat{f}([p])$ for a finite pattern $p \in S^M$.

```
function HatCA(f, p)
    Let O ← NP oracle.
    if O finds a pattern q ∈ f^{-1}(p) then
        Let D ← M + N.
        Let F ← {(q, D)}.
        loop
            if O finds a pattern q' ∈ f^{-1}(p) with q'|E ≠ r|E for all (r, E) ∈ F then
                Let D ← D ∩ eq(q, q').
                Let F ← F ∪ {(q', D)}
            else
                return q|D
    else
        return ⊤
```

Our results rely on the existence of patterns p that force large patterns into their preimages, meaning that $\hat{f}([p])$ is large in the sense of $\le$. We say a pattern p is *self-enforcing* under f if $[p] \le \hat{f}([p])$. In a slight abuse of terminology, we also say that a temporally t-periodic configuration $x \in S^{\mathbb{Z}^d}$ is *self-enforcing* if $\widehat{(f^t)}([x]) = [x]$. A self-enforcing agar is then a spatially and temporally periodic configuration that has a unique chain of preimages.

▶ **Lemma 13.** *The set of all pairs (f, x) such that f is a CA on $S^{\mathbb{Z}^d}$ and $x \in S^{\mathbb{Z}^d}$ is a self-enforcing agar is recursively enumerable.*

Proof. Let $x \in S^{\mathbb{Z}^d}$ be a self-enforcing agar with spatial periods $n_1\vec{e}_1, \ldots, n_d\vec{e}_d$ and temporal period t. Denote the iterated CA by $h = f^t$, and let $B = [0, n_1 - 1] \times \cdots \times [0, n_d - 1]$. We need to find a certificate for $\hat{h}([x]) = [x]$. For this, observe that by continuity of $\hat{h}$ there is a finite subpattern p of x such that $\hat{h}([y]) \geq [x|B]$ for every configuration $y \in [p]$. This implies $\hat{h}([p]) \geq [x|B]$. By Lemma 12, this latter inequality can be checked in $\mathrm{FP}^{\mathrm{NP}}$.

We claim that p is a certificate that x is a self-enforcing agar. Let $\vec{v} \in V = \langle n_1\vec{e}_1, \ldots, n_d\vec{e}_d \rangle$ be arbitrary. We compute

$$\hat{h}([x]) = \sigma_{\vec{v}}(\hat{h}([x])) \geq \sigma_{\vec{v}}(\hat{h}([p])) \geq \sigma_{\vec{v}}([x|B]) = [x|\vec{v} + B],$$

and since $\mathbb{Z}^d = \bigcup_{\vec{v} \in V} (\vec{v} + B)$, this implies $\hat{h}([x]) = [x]$. ◀

▶ Remark 14. The semi-algorithm described in the proof is not very practical: given an agar, we have no information about how large the certificate could be, so for each agar we either need to guess some certificate size, or we have to keep trying increasingly large certificates. Our implementation runs in parallel a search for other periodic preimages for the agar – if such a preimage exists, then clearly the agar does not enforce itself, and we can stop looking for a certificate. We omit the pseudocode.

Most agars in the range we searched were either self-enforcing or had another periodic preimage. There exist two-dimensional cellular automata whose set of self-enforcing agars is not computable (by a relatively simple reduction from the tiling problem of Wang tiles [2], which we omit), but we do not know whether this is the case for Game of Life.

Say a pattern $p \subset S^M$ is *locally fixed* for the CA f if there exists a pattern $q \in S^{M+N}$ (where N is the minimal neighborhood of f) such that $p = q|M = f(q)|M$.

▶ **Lemma 15.** *For every CA f on $S^{\mathbb{Z}^d}$, every locally fixed pattern $p \in S^M$ admits a unique maximal self-enforcing subpattern. For a fixed CA g, given p, a vector $\vec{v} \in \mathbb{Z}^d$ and $n \geq 1$ in unary, it can be computed in $\mathrm{FP}^{\mathrm{NP}}$ for the CA $f = \sigma_{\vec{v}} \circ g^n$.*

Proof. Since p has finitely many subpatterns and the empty pattern is trivially self-enforcing, p admits at least one maximal self-enforcing subpattern. If $D, D' \subset M$ satisfy $\hat{f}([p|D]) \geq [p|D]$ and $\hat{f}([p|D']) \geq [p|D']$, then $\hat{f}([p|D \cup D']) \geq [p|D \cup D']$ by monotonicity of $\hat{f}$. Thus $q = p|E$ for $E = \bigcup \{D \subset M \mid \hat{f}([p|D]) \geq [p|D]\}$ is the unique self-enforcing subpattern.

Then fix g, and let p, $\vec{v}$ and n be given. We apply Algorithm 2 to the CA $f = \sigma_{\vec{v}} \circ g^n$. On each iteration of the loop, the algorithm replaces p with the maximal subpattern forced by p (here we use the fact that p is locally fixed). Since q is a subpattern forced by itself, by monotonicity it is also forced by each of these subpatterns, and thus remains a subpattern on each iteration. Since q is maximal and p has finitely many subpatterns, the algorithm eventually converges on q.

Finally, Algorithm 2 is in $\mathrm{FP}^{\mathrm{NP}}$, since the number of iterations of the loop is at most $|M|$, and $\textsc{HatCA}(f, p)$ is in $\mathrm{FP}^{\mathrm{NP}}$ with respect to these parameters. ◀

To conclude this section, we show that in the formulation of the Generalized grandfather problem (which we prove as Theorem 1), it makes no difference whether we consider unrestricted configurations, finite configurations or finite patterns. This is well-known in cellular automata theory.

▶ **Lemma 16.** *Let $f : S^{\mathbb{Z}^d} \to S^{\mathbb{Z}^d}$ be a cellular automaton with a quiescent state $0 \in S$, and $n \in \mathbb{N}$. The following are equivalent:*

Algorithm 2 Finding the maximal self-enforcing subpattern of a locally fixed pattern $p \in S^M$.

```
function SelfEnforcingSubpattern(p)
    loop
        Let q ← HatCA(f, p)|M.
        if q = p then
            return q
        else
            Let p ← q.
```

1. *There exists a finite configuration $x \in S^{\mathbb{Z}^d}$ such that $f^{-n}(x)$ contains a finite configuration, and $f^{-(n+1)}(x) = \emptyset$.*
2. *There exists $x \in S^{\mathbb{Z}^d}$ such that $f^{-n}(x) \neq \emptyset$ and $f^{-(n+1)}(x) = \emptyset$.*
3. *There exists a finite pattern p such that $f^{-n}(p) \neq \emptyset$ and $f^{-(n+1)}(p) = \emptyset$.*

Proof. The implication $1 \implies 2$ is clear, and $2 \implies 3$ is the classical compactness argument that we used to prove Lemma 10.

We prove $3 \implies 1$. Take an arbitrary $q \in f^{-n}(p)$, and complete it into a finite configuration $y \in S^{\mathbb{Z}^d}$ by setting $y_{\vec{v}} = 0$ for all $\vec{v} \in \mathbb{Z}^d \setminus \mathrm{dom}(q)$. Then $x = f^n(y)$ satisfies the conditions of item 1: $f^{-n}(x)$ contains the finite configuration y, while $f^{-(n+1)}(x) = \emptyset$ since x contains an occurrence of p. ◀

3.1 Köynnös

We begin by studying köynnös, which we recall is obtained from the 6×3 pattern

$$P = \begin{matrix} 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{matrix}$$

by repeating P horizontally and vertically to define an infinite 6×3-periodic configuration $x^P \in \{0,1\}^{\mathbb{Z}^2}$. Observe that every 0 in x^P is surrounded by exactly four 1s, and every 1 by exactly three 1s. Thus we have $g(x^P) = x^P$, so that x^P is indeed an agar. Moreover, we claim that x^P has no other predecessors than itself: $g^{-1}(x^P) = \{x^P\}$. This is due to the following lemma.

▶ **Lemma 17.** *Let x be in the spatial orbit of köynnös. Then $\hat{g}(x|[-12,17] \times [-12,14]) \geq x|[-8,13] \times [-9,10]$.*

Proof. Applying Algorithm 1 to $\sigma_{\vec{v}}(x^P)|[-12,17] \times [-12,14]$ for all $\vec{v} \in [0,5] \times [0,2]$ gives the result. The intersection of the domains of the patterns $\hat{g}(x|[-12,17] \times [-12,14])$ for such $x = \sigma_{\vec{v}}(x^P)$ is shown in Figure 3, and clearly contains the rectangle $[-8,13] \times [-9,10]$. ◀

Put concretely, the lemma states that if R is a periodic continuation of P of size 30×27 and Q is its predecessor, then P must occur at the center of Q (and indeed many more cells are forced, even beyond what we state in the lemma). This is indeed a certificate for x^P being a self-enforcing agar, as in the proof of Lemma 13: for any predecessor $y \in g^{-1}(x^P)$ and cell $\vec{v} \in \mathbb{Z}^d$, Lemma 17 gives $\sigma_{\vec{v}}(y)|[-8,13] \times [-9,10] = \sigma_{\vec{v}}(x^P)|[-8,13] \times [-9,10]$, so in particular $y_{\vec{v}} = x^P_{\vec{v}}$.

As a corollary of Lemma 17, finite perturbations of x^P can never be erased by g. We prove a stronger claim: all finite perturbations spread to the left and right at a speed of

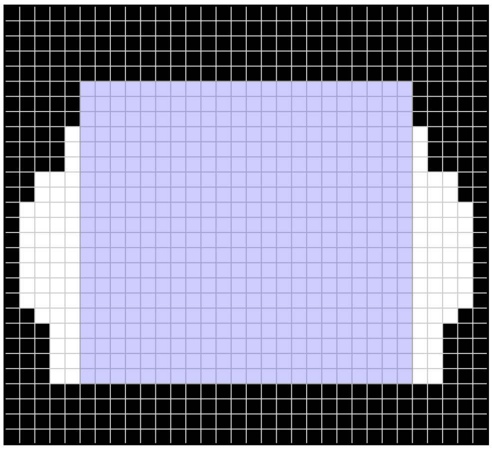


Figure 3 The intersection of the domains of $\hat{g}(x|[-12,17]\times[-12,14])$ for x in the spatial orbit of köynnös, drawn in white inside $[-13,18]\times[-13,15]$. The area $[-8,13]\times[-9,10]$ is highlighted in blue.

one column per time step. In particular, köynnös cannot be stabilized from the inside. We note that, as the agar kynnös studied in the next section does not possess this property, we cannot use it to prove Theorem 1.1, at least with the same method.

▶ **Lemma 18.** *Let $R = [-n_W, n_E]\times[-n_S, n_N]$ be a rectangle and $A = \left[\begin{smallmatrix} n_W+1 & n_E+1 & n_S+1 & n_N+1 \\ n_W & n_E & n_S & n_N \end{smallmatrix}\right]$ the surrounding annulus of thickness 1. Let p be a pattern such that the domain of $g(p)$ contains $A \cup R$, and suppose $p|A = g(p)|A = x^P|A$. If $\mathrm{diff}(x^P, g(p)) \cap R \subset [a,b]\times\mathbb{Z}$, then $\mathrm{diff}(x^P, p) \cap R \subset [a+1, b-1]\times\mathbb{Z}$.*

Note that we may have $b-a \leq 1$, in which case the conclusion becomes $\mathrm{diff}(x^P,p)\cap R = \emptyset$, or equivalently, $x^P|R = p|R$.

Proof. Since the orbit of köynnös and g are left-right symmetric, it is enough to prove that $\mathrm{diff}(x^P,p) \cap R \subset (-\infty, b-1]$. We prove the contrapositive: suppose there exists $(i,j) \in \mathrm{diff}(x^P,p)\cap R$ for some $i \geq b$, and let i be maximal. We split into cases based on the congruence class of i modulo 6, that is, the column of P that i lies in. Note that the bottom left cell of P is at the origin in x^P, and the domain of P is the rectangle $[0,5]\times[0,2]$. If $i \in \{0,2,3\} + 6\mathbb{Z}$, we choose j as maximal, and otherwise we choose it as minimal.

We handle the case $i \in 2 + 6\mathbb{Z}$, the others being similar or easier. If $j \in 3\mathbb{Z}$, then $p_{(i+1,j+1)} = x^P_{(i+1,j+1)} = 1$ has four other 1s in its neighborhood, and becomes 0 in $g(p)$. If $j \in 1+3\mathbb{Z}$, then $p_{(i+1,j)} = x^P_{(i+1,j+1)} = 1$ has four or five other 1s in its neighborhood, and becomes 0 in $g(p)$. If $j \in 2+3\mathbb{Z}$, then $p_{(i+1,j+1)} = x^P_{(i+1,j+1)} = 0$ has three 1s in its neighborhood, and becomes 1 in $g(p)$. In each case $\mathrm{diff}(x^P, g(p))$ intersects $\{i+1\}\times\mathbb{Z}$. ◀

For any $D \subset \mathbb{Z}^2$, the subpattern of köynnös of shape $D + [0,29]\times[0,26]$ forces the subpattern of shape $D + [4,25]\times[3,22]$ to occur in its g-preimage, by Lemma 17. By Lemma 18, we force more: a non-köynnös area inside a hollow patch of köynnös expands horizontally under g, so under $\hat{g}$ the horizontal extent of the hole must shrink. We do not give a precise statement for this general fact, and only apply the lemma in the case of annuli.

▶ **Lemma 19.** *Let x be in the orbit of köynnös. Suppose the following inequalities hold:*

$$m_W - n_W \geq 30, m_E - n_E \geq 30, m_S - n_S \geq 27, m_N - n_N \geq 27.$$

Denote $Q = x|\left[\begin{smallmatrix} m_W & m_E & m_S & m_N \\ n_W & n_E & n_S & n_N \end{smallmatrix}\right]$. *If* $n_E + n_W \geq 2$, *then*

$$\hat{g}(Q) \geq x|\begin{bmatrix} m_W - 4 & m_E - 4 & m_S - 3 & m_N - 4 \\ n_W - 1 & n_E - 1 & n_S + 4 & n_N + 3 \end{bmatrix}$$

while if $n_E + n_W \in \{0, 1\}$ *we have*

$$\hat{g}(Q) \geq x|[-(m_W - 4), m_E - 4] \times [-(m_S - 3), m_N - 4]$$

One may consider the latter case a special case of the former: there too, the hole shrinks horizontally by two steps, and since its width is at most two it disappears.

Proof. The inequalities simply state that the annulus Q is thick enough that each of its cells is part of a 30×27 rectangle contained in Q. From Lemma 17, we get $\hat{g}(Q) \geq x|A$, where $A = \left[\begin{smallmatrix} m_W-4 & m_E-4 & m_S-3 & m_N-4 \\ n_W+4 & n_E+4 & n_S+4 & n_N+3 \end{smallmatrix}\right]$ is a slightly thinner annulus. If there is no g-preimage for Q, then $\hat{g}(Q) = \top$ and we are done. Suppose then that it has a preimage R. Since $R \geq \hat{g}(Q) \geq x|A$, both Q and R agree with x on the thickness-1 annulus $\left[\begin{smallmatrix} n_W+5 & n_E+5 & n_S+5 & n_N+4 \\ n_W+4 & n_E+4 & n_S+4 & n_N+3 \end{smallmatrix}\right] \subset A$. Lemma 18 implies that R agrees with x on $\left[\begin{smallmatrix} m_W-4 & m_E-4 & m_S-3 & m_N-4 \\ n_W-1 & n_E-1 & n_S+4 & n_N+3 \end{smallmatrix}\right]$, as claimed. ◀

We now prove Theorem 1.1, and thus give the first proof of Theorem 1. In fact, we give a simple formula that produces configurations that have an nth preimage, but no $(n+1)$st one.

▶ **Lemma 20.** *Let* x *be in the orbit of köynnös, and suppose* $\emptyset \neq \mathrm{diff}(y, x) \subset B = [0, a] \times [0, n]$ *where* $a \in \{0, 1\}$. *Then*

$$p = g^k(y)|[-30 - 6k, 30 + a + 6k] \times [-27 - 8k, 27 + n + 8k]$$

appears in the k*th image of* g, *but not in the* $(k+1)$*st.*

Proof. By definition, p appears in the kth image of g. It suffices to show its $\hat{g}^{k+1}$-image is $\top$. Namely, we then have $\widehat{g^{k+1}}(p) \geq \hat{g}^{k+1}(p) = \top$ by Lemma 11, which means precisely that p has no g^{k+1}-preimage.

Let q be the restriction of p to

$$\begin{bmatrix} 30 + 6k & 30 + a + 6k & 27 + 8k & 27 + n + 8k \\ k & a + k & k & n + k \end{bmatrix}.$$

Observe that q agrees with x because g has radius 1, so by Lemma 19 and induction, we can deduce that

$$\hat{g}^j(q) \geq x|\begin{bmatrix} 30 + 6k - 4j & 30 + a + 6k - 4j & 27 + 8k - 3j & 27 + n + 8k - 4j \\ k - j & a + k - j & k + 4j & n + k + 3j \end{bmatrix}$$

for all $j \leq k$. This is because for $j \leq k - 1$ we have

$$30 + 6k - 4j - (k - j) \geq 30, \qquad 30 + a + 6k - 4j - (a + k - j) \geq 30,$$
$$27 + 8k - 3j - (k + 4j) \geq 27, \qquad 27 + n + 8k - 4j - (n + k + 3j) \geq 27,$$

and thus we can inductively apply the lemma. But in

$$\hat{g}^k(q) \geq x|\begin{bmatrix} 30 + 2k & 30 + a + 2k & 27 + 5k & 27 + n + 4k \\ 0 & a & 5k & n + 4k \end{bmatrix}$$

the annulus still has sufficient thickness (i.e. the inequalities still hold for $j = k$), so we can apply the second case of the lemma to get

$$\hat{g}^{k+1}(q) \geq x|[-(26 + 2k), 26 + 2k] \times [-(24 + 5k), 23 + n + 4k] = r.$$

By Lemma 11 we have $\hat{g}^{k+1}(p) \geq \hat{g}^{k+1}(q) \geq r$, and by the same lemma we either have $\hat{g}^{k+1}(p) = \top$ (as desired), or

$$g^{k+1}(\hat{g}^{k+1}(p)) \leq g^{k+1}(\widehat{g^{k+1}}(p)) \parallel p.$$

But since the speed of light is 1 and x is a fixed point, we have

$$g^{k+1}(r) \geq x|[-(25+k), 25+k] \times [-(23+4k), 22+n+3k] \geq x|[-k, a+k] \times [-k, n+k]$$

and $x|[-k, a+k] \times [-k, n+k] \parallel p$. But Lemma 18 applied k times to y implies that $\mathrm{diff}(x, g^k(y))$, and thus $\mathrm{diff}(x, p)$, intersects $[-k, a+k] \times [-k, n+k]$, a contradiction. Thus we indeed must have $\hat{g}^{k+1}(p) = \top$. ◀

3.2 Kynnös

Denote by

$$Q = \begin{matrix} 0 & 0 & 1 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 & 1 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 \end{matrix}$$

the fundamental domain of kynnös, and by $x^Q \in \{0,1\}^{\mathbb{Z}^2}$ the associated 6×6-periodic configuration with $g(x^Q) = x^Q$. The following lemma states that it contains a self-enforcing patch (it is essentially a more precise stement of Theorem 4).

▶ **Lemma 21.** *There is a finite set $D \subset \mathbb{Z}^2$ such that $p = x^Q|D$ satisfies $\hat{g}(p) = p$. Furthermore, there is a finite-support configuration $x \in [p]$ with $g(x) = x$.*

The patch p is shaped like a 22×28 rectangle with 8 cells missing from each corner. It is depicted in Figure 4, together with the still life x containing it. The patch was found by simply applying the function of Algorithm 2 to the 70×70-patches of the agars we found during our searches. Kynnös was the first configuration that yielded a nonempty self-enforcing patch, which we then optimized to its current size. This lemma directly implies Theorem 4, and almost directly Theorem 5.

Proof of Theorem 5. Let y be the finite-support configuration obtained by taking x from the previous lemma and adding a glider that is just about to hit the kynnös patch. It can be checked by simulation that the patch can be annihilated this way. Observe that y is in the limit set $\Omega(g)$: simply shoot the glider from infinity. If $\epsilon > 0$ is very small, in any ϵ-chain starting from y we see the patch destroyed. It is impossible to reinstate it, as the existence of a first step in the chain where it appears again contradicts Lemma 21. ◀

As stated, kynnös can be stabilized from both inside and outside. Figure 5 shows a still life configuration containing a "ring" of kynnös with a hole of 0-cells inside it. From the figure it is easy to deduce the existence of such rings of arbitrary size and thickness.

Note that if the ring is at least 22 cells thick, then its interior is completely surrounded by a ring-shaped self-enforcing pattern consisting of translated, rotated and partially overlapping copies of the 28×22 self-enforcing patch, through which no information can pass without destroying it forever. If we then replace the empty cells inside the ring with an arbitrary

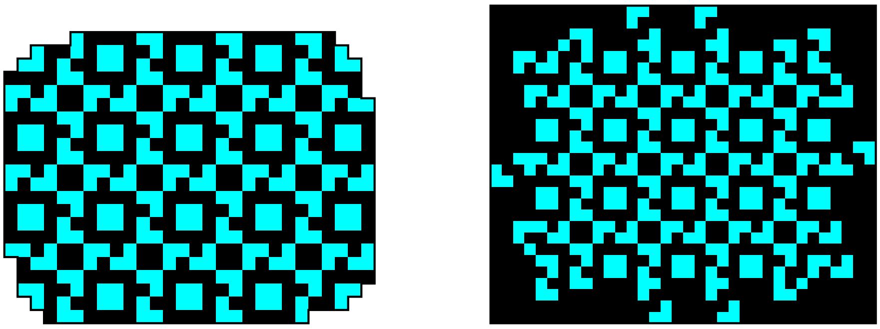


Figure 4 A self-enforcing patch of kynnös and a still life containing it. The still life has the minimal number of live cells, 306, of any still life containing the patch. The number was minimized by Oscar Cunningham. [7]

pattern, the resulting finite pattern P occurs in the limit set $\Omega(g)$ if and only if the interior pattern evolves periodically under g. Namely, if the pattern occurs in $\Omega(g)$, then it has an infinite sequence of preimages, each of which must contain the self-enforcing kynnös ring. The interior has a finite number of possible contents (2^m for an interior of m cells), so it must evolve into a periodic cycle, of which P is part. From this idea, and some engineering with gadgets found by other researchers and Life enthusiasts, we will obtain Theorems 1.2, 2 and 3. The first one was essentially proved by Adam Goucher [16]. Note the difference between these rings and the köynnös annuli of Section 3.1: the latter force strictly smaller versions of themselves in their preimages, and do not admit nontrivial periodically evolving interiors.

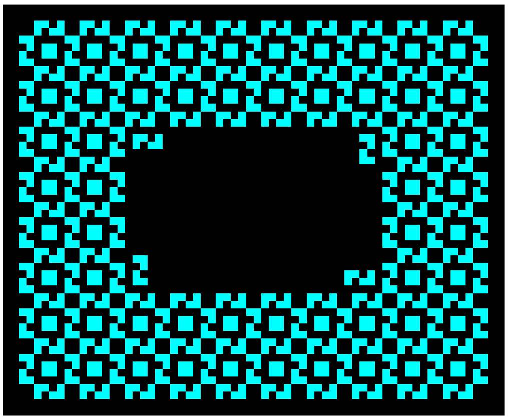


Figure 5 A stable ring of kynnös.

Proof of Theorem 1.2. Given integers $k, m \geq 1$ with k odd, we construct a configuration $x \in \{0,1\}^{\mathbb{Z}^2}$ such that the support of $g^n(x)$ is contained in $[0, 32k+73] \times [0, 46m]$ for all $n \geq 0$, and $g^{48 \cdot 4^{(2k+1)m}}(x)$ is not g-periodic. When the support of $g^{48 \cdot 4^{(2k+1)m}}(x)$ is surrounded by a kynnös ring of width 22, the resulting pattern has a $48 \cdot 4^{(2k+1)m}$th preimage, but not arbitrarily old preimages. If we choose $k = 23s$ and $m = 16s$ for some $s \geq 0$, the resulting pattern has size $(736s + O(1)) \times (736s + O(1))$, and the chain of preimages has length $48 \cdot 4^{736s^2+16s}$. Choosing $s = n/736 - O(1)$ yields the lower bound, and the upper bound is the trivial one (even ignoring the fact we do not modify the boundaries).

The main components of x are the *period-48 glider gun* [27], which produces one glider every 48 time steps, and the *quadri-snark* [29], which emits one glider at a 90 degree angle for every 4 gliders it receives. The support of x consists of a single period-48 gun aimed at a sequence of $(2k+1)m$ quadri-snarks, each of which receives the gliders the previous one emits. They effectively implement a quaternary counter with values in $[0, 4^{(2k+1)m} - 1]$. The glider emitted by the final quasi-snark will collide with the kynnös ring, ensuring that the pattern right before the impact does not occur in the limit set $\Omega(g)$, but has a chain of preimages of length at least $48 \cdot 4^{(2k+1)m}$. An example pattern and a schematic for $m = n = 2$ are given in Figure 6. It is easy to extrapolate to arbitrary $m, n \geq 1$ from the figure. ◀

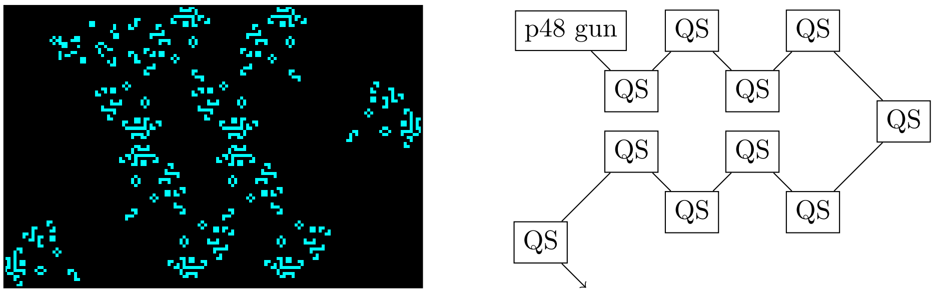


Figure 6 A configuration corresponding to $k = m = 2$ in the proof of Theorem 1.2.

To implement more complex Life patterns with desired properties, we use the fact that Life is *intrinsically universal*, that is, capable of simulating all $\mathbb{Z}^2$ cellular automata. Formally, for any other CA $f : \Sigma^{\mathbb{Z}^2} \to \Sigma^{\mathbb{Z}^2}$, there are numbers $K, T \geq 1$ and an injective function $\tau : \Sigma \to \{0,1\}^{K \times K}$ such that for all configurations $x \in \Sigma^{\mathbb{Z}^2}$ we have $g^T(\tau(x)) = \tau(f(x))$, where τ is applied cellwise in the natural way. We use the simulation technique of [10], which allow us to easily simulate patterns with *fixed boundary conditions*. This means that any rectangular pattern $R \in \Sigma^{a \times b}$ can be simulated by a finite-support configuration of g in such a way that simulated cells whose f-neighborhood is not completely contained in the rectangle $[0, a-1] \times [0, b-1]$ are forced to retain their value.

Proof of Theorem 2. Let $L \subset \{0,1\}^*$ be a PSPACE-hard language decidable in linear space, such as TQBF. Define a Turing machine M as follows. Given input $w \in \{0,1\}^*$, M determines whether $w \in L$ using at most $|w|$ additional tape cells and without modifying w. If $w \in L$, then it erases the additional tape cells and returns to its initial state, thus looping forever. If $w \notin L$, then M stays in a rejecting state forever. We simulate M by a cellular automaton f in a standard way: each cell is either empty, or contains a tape symbol and possibly the state of the computation head.

Next, we simulate the CA f by g as described above. Given a word $w \in \{0,1\}^*$, let $P(w)$ be the pattern corresponding to a simulated initial configuration of M on input w with $|w|$ additional tape cells and fixed boundary conditions, surrounded by a kynnös ring. If $w \in L$, then $P(w)$ occurs in the limit set $\Omega(g)$, since it can be completed into a g-periodic configuration in which the simulated M repeatedly computes $w \in L$. If $w \notin L$, then $P(w)$ does not occur in $\Omega(g)$, since the interior of the ring eventually evolves into a simulated configuration with M in a rejecting state, never returning to $P(w)$. ◀

Extending $P(w)$ by zeroes on all sides (resp. repeating it periodically), we obtain that it is PSPACE-hard whether a given finite-support configuration (resp. periodic configuration) appears in the limit set.

Proof of Theorem 6. Let L be as in the previous proof, and let M be a Turing machine that, on input $w \in \{0,1\}^*$, decides $w \in L$ using no additional tape cells. Then M erases the entire tape and enters an accepting or rejecting state depending on the result of the computation. We simulate M by g as in the previous proof. Given $w \in \{0,1\}^*$, let p be the pattern corresponding to a tape of M containing w and an initial state, and q the one corresponding to $|w|$ blank tape cells and an accepting state of M, both surrounded by a ring of kynnös of the same dimensions. Then q is reachable from p if and only if $w \in L$: if q is to be reached, the ring of p must stay intact, enclosing a correct simulation of M. ◀

Of course, again by extending the resulting patterns by 0-cells (resp. repeating them periodically), we obtain PSPACE-hardness of reachability between two given finite-support (resp. periodic) configurations, i.e. given the full descriptions of two configurations $x, y \in \{0,1\}^{\mathbb{Z}^2}$, the question of whether $g^n(x) = y$ for some $n \geq 0$. However, this reachability problem is in fact even Σ_1^0-complete (resp. PSPACE-complete) directly by intrinsic universality. For the case of finite configurations, one needs a variant of intrinsic universality where the zero state of an arbitrary cellular automaton is represented by an all-zero pattern; such a variant was proved in [15].

Proof of Theorem 3. Let M be a two-dimensional Turing machine whose tape alphabet has two distinguished values, denoted a and b. When M is initialized on a rectangular tape containing only as and bs, it repeatedly checks whether its left and right halves are equal, destroying the tape if they are not. We again simulate M by a CA f, and then f by g. Then a simulated rectangular tape with the head of M in its initial state, surrounded by a kynnös ring, is in $\Omega(g)$ if and only if the two halves of the tape are equal.

It was proved in [21] that for all sofic shifts $X \subset S^{\mathbb{Z}^2}$ there exists an integer $C > 1$ with the following property. For all $n \geq 1$ and configurations $x^1, \ldots, x^{C^n} \in X$, there exist $i \neq j$ such that the configuration $y = (x^i|[0, n-1]^2) \sqcup (x^j|\mathbb{Z}^2 \setminus [0, n-1]^2)$ is in X. Assuming for a contradiction that $\Omega(g)$ is sofic, consider the configurations $x(P) \in \Omega(g)$ for $P \in \{a, b\}^{n \times n}$ that contain a kynnös ring and a simulated tape of M with two identical P-halves. Based on the above, when n is large enough that $2^{n^2} > C^{Kn}$, we can swap the right half of one $x(P)$ with that of another to obtain a configuration $y \in \Omega(g)$ containing a simulated tape of M with unequal halves inside a kynnös ring, a contradiction. ◀

We remark that a weaker version of Theorem 1.2 (where 1/368 is replaced by a much smaller, or even implicit, constant) could also be proved by intrinsic universality.

3.3 The marching band

Let $h = g^2$. Denote by

$$R = \begin{matrix} 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \end{matrix}$$

the fundamental domain of the marching band, and by $x^R \in \{0,1\}^{\mathbb{Z}^2}$ the associated 8×4-periodic configuration with $h(x^R) = x^R$. The following is proved just like Lemma 17. Note that the forced region extends outside the original pattern.

▶ **Lemma 22.** *Let x be in the spatial orbit of the marching band. Then $\hat{h}(x|[0, 47] \times [0, 43]) \geq x|[10, 29] \times [-1, 44]$.*

Proof of Theorem 7. Let x be in the orbit of x^R, and let $p = x|[-a, b] \times [-c, d]$. By the previous lemma, as long as $a + b \geq 48$ and $c + d \geq 44$ we have $\hat{h}(p) = x|[-(a-10), b-18] \times [-(c+1), d+1]$. Iterating this we get

$$\hat{h}^n(x|[-10n, 18n+47] \times [0, 43]) \geq x|[0, 47] \times [-n, n+43].$$

Denote $S = [-10n, 18n+47] \times [0, 43]$. Let $P = x|S$ and $Q = \sigma_{\vec{u}}(x)|\vec{v} + S$ for some $\vec{u} \in \mathbb{Z}^2$ and $\vec{v} = (0, 2n)$. Both patterns appear in the limit set of g, since they are extracted from a fixed point of $h = g^2$. Observe that since the domains of $\hat{h}^n(P)$ and $\hat{h}^n(Q)$ intersect, we can pick the shift $\vec{u}$ so that one of the forced bits is different in some position in $\hat{h}^n(P)$ and $\hat{h}^n(Q)$, which clearly means $\hat{h}^n(P \sqcup Q) = \top$.

Now, P and Q each fit inside a $29n \times 29n$ rectangle (if $n \geq 47$), and the patterns cannot be glued in the limit set with gluing distance at most $2n$, since the glued pattern should have an nth h-preimage. This gives the statement. ◀

4 Chaotic conclusions

There are several definitions of topological chaos. We refer the reader to [3] for a survey. Briefly, a system is called Auslander-Yorke chaotic if it is topologically transitive and is sensitive to initial conditions, and Devaney chaotic if it is Auslander-Yorke chaotic and additionally has dense periodic points. As far as we know, before our results it was open whether Game of Life exhibits these types of chaos on its limit set; the following corollary shows that it does not.

▶ **Theorem 23.** *The Game of Life restricted to its limit set is not topologically transitive, and does not have dense periodic points.*

Proof. Either of these properties clearly implies chain-nonwanderingness, contradicting Theorem 5. ◀

Two other standard notions of chaos are Li-Yorke chaos and positive entropy (we omit the definitions). Game of Life exhibits these trivially, since it admits a glider. More generally, intrinsic universality implies that it exhibits any property of spatiotemporal dynamics of cellular automata that is inherited from subsystems of finite-index subactions of the spacetime subshift. Sensitivity in itself is also sometimes considered a notion of chaos. This remains wide open.

▶ **Question 1.** *Is Game of Life sensitive to initial conditions?*

One can also ask about chaos on "typical configurations". For example, take the uniform Bernoulli measure (or some other distribution) as the starting point, and consider the trajectories of random configurations. We can say essentially nothing about this setting.

In our topological dynamical context, a natural way to formalize this problem is through the *generic limit set* as defined in [25]. It is a subset of the phase space of a dynamical system that captures the asymptotic behavior of topologically large subsets of the space. We omit the exact definition, but for a cellular automaton f, this is a nonempty subshift invariant under f [9]. It follows that the generic limit set is contained in the limit set, and that the language of the generic limit set of g contains the letter 0 (because the singleton subshift $\{1^{\mathbb{Z}^2}\}$ is not g-invariant).

We say a cellular automaton f on $S^{\mathbb{Z}^d}$ is *generically nilpotent* if its generic limit set contains only one configuration, which must then be the all-s configuration for a quiescent state $s \in S$. This is equivalent to the condition that every finite pattern can be extended into some larger pattern p such that for large enough $n \in \mathbb{N}$, we have $f^n(x)_{\vec{0}} = s$ for all $x \in [p]$. By the previous observation, if Game of Life were generically nilpotent, we would have $s = 0$. We strongly suspect that it is not generically nilpotent, i.e. the symbol 1 occurs in the generic limit set. However, we have been unable to show this.

▶ **Question 2.** *Is Game of Life generically nilpotent?*

Chaos is usually discussed for one-dimensional dynamical system, but we find its standard ingredients, such as topological transitivity and periodic points, quite interesting. We have been unable to resolve most of these.

▶ **Question 3.** *Is the limit set of Game of Life topologically transitive as a subshift?*

▶ **Question 4.** *Does the limit set of Game of Life have dense totally periodic points as a subshift?*

References

1 Nathalie Aubrun and Mathieu Sablik. Simulation of effective subshifts by two-dimensional subshifts of finite type. *Acta Appl. Math.*, 126(1):35–63, August 2013. URL: `http://dx.doi.org/10.1007/s10440-013-9808-5`, `doi:10.1007/s10440-013-9808-5`.

2 Robert Berger. The undecidability of the domino problem. *Mem. Amer. Math. Soc. No.*, 66, 1966. 72 pages.

3 François Blanchard. Topological chaos: what may this mean? *Journal of Difference Equations and Applications*, 15(1):23–46, 2009. `doi:10.1080/10236190802385355`.

4 Mike Boyle, Ronnie Pavlov, and Michael Schraudner. Multidimensional sofic shifts without separation and their factors. *Transactions of the American Mathematical Society*, 362(9):4617–4653, 2010. `doi:10.1090/s0002-9947-10-05003-8`.

5 John H. Conway. Email to Dean Hickerson. Private email group *LifeCA*, 1992. Provided by Dave Greene.

6 Karel Culik, II, Jan Pachl, and Sheng Yu. On the limit sets of cellular automata. *SIAM J. Comput.*, 18(4):831–842, 1989. `doi:10.1137/0218057`.

7 Oscar Cunningham. Response on ConwayLife forum (username macbi). `https://conwaylife.com/forums/viewtopic.php?f=7&t=3180&start=125#p140295`. Accessed: 2022-02-09.

8 Alberto Dennunzio, Enrico Formenti, Darij Grinberg, and Luciano Margara. From Linear to Additive Cellular Automata. In Artur Czumaj, Anuj Dawar, and Emanuela Merelli, editors, *47th International Colloquium on Automata, Languages, and Programming (ICALP 2020)*, volume 168 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 125:1–125:13, Dagstuhl, Germany, 2020. Schloss Dagstuhl–Leibniz-Zentrum für Informatik. URL: `https://drops.dagstuhl.de/opus/volltexte/2020/12532`, `doi:10.4230/LIPIcs.ICALP.2020.125`.

9 Saliha Djenaoui and Pierre Guillon. The generic limit set of cellular automata. *Journal of Cellular Automata*, 14(5-6):435–477, 2019. URL: `https://www.oldcitypublishing.com/journals/jca-home/jca-issue-contents/jca-volume-14-number-5-6-2019/jca-14-5-6-p-435-477/`.

10 Bruno Durand and Zsuzsanna Róka. The game of life: universality revisited. In *Cellular automata (Saissac, 1996)*, volume 460 of *Math. Appl.*, pages 51–74. Kluwer Acad. Publ., Dordrecht, 1999.

11 Bruno Durand, Andrei Romashchenko, and Alexander Shen. Effective closed subshifts in 1D can be implemented in 2D. In *Fields of logic and computation*, volume 6300 of *Lecture Notes in Comput. Sci.*, pages 208–226. Springer, Berlin, 2010.

12 Noam D Elkies. The still-life density problem and its generalizations. *arXiv preprint math/9905194*, 1999.

13 Henryk Fukś. Explicit solution of the cauchy problem for cellular automaton rule. *J. Cell. Autom.*, 12(6):423–444, 2017. URL: `http://www.oldcitypublishing.com/journals/jca-home/jca-issue-contents/jca-volume-12-number-6-2017/jca-12-6-p-423-444/`.

14 Martin Gardner. Mathematical Games: The Fantastic Combinations of John Conway's New Solitaire Game "Life". *Scientific American*, 223(4):120–123, 1970.

15 Adam Goucher. Fully self-directed replication. `https://cp4space.hatsya.com/2018/11/12/fully-self-directed-replication/`. Accessed: 2022-04-20.

16 Adam Goucher. Response on ConwayLife forum (username calcyman). `https://conwaylife.com/forums/viewtopic.php?f=7&t=3180&start=100#p140273`. Accessed: 2022-02-09.

17 Gustav A. Hedlund. Endomorphisms and automorphisms of the shift dynamical system. *Math. Systems Theory*, 3:320–375, 1969.

18 Lyman P. Hurd. Formal language characterizations of cellular automaton limit sets. *Complex Systems*, 1(1):69–80, 1987.

19 N. Johnston and D. Greene. *Conway's Game of Life: Mathematics and Construction*. Lulu.com, 2022. URL: `https://books.google.fi/books?id=xSJlEAAAQBAJ`.

20 Jarkko Kari. Universal pattern generation by cellular automata. *Theoret. Comput. Sci.*, 429:180–184, 2012. `doi:10.1016/j.tcs.2011.12.037`.

21 Steve Kass and Kathleen Madden. A sufficient condition for non-soficness of higher-dimensional subshifts. *Proceedings of the American Mathematical Society*, 141(11):3803–3816, 2013. URL: `http://dx.doi.org/10.1090/S0002-9939-2013-11646-1`, `doi:10.1090/S0002-9939-2013-11646-1`.

22 Douglas Lind and Brian Marcus. *An introduction to symbolic dynamics and coding*. Cambridge University Press, Cambridge, 1995. URL: `http://dx.doi.org/10.1017/CBO9780511626302`, `doi:10.1017/CBO9780511626302`.

23 Ville Lukkarila. Sensitivity and topological mixing are undecidable for reversible one-dimensional cellular automata. *J. Cell. Autom.*, 5(3):241–272, 2010. URL: `http://www.oldcitypublishing.com/journals/jca-home/jca-issue-contents/jca-volume-5-number-3-2010/jca-5-3-p-241-272/`.

24 Alejandro Maass. On the sofic limit sets of cellular automata. *Ergodic Theory and Dynamical Systems*, 15, 1995. `doi:10.1017/S0143385700008609`.

25 John Milnor. On the concept of attractor. *Communications in Mathematical Physics*, 99(2):177–195, 1985. `doi:10.1007/BF01212280`.

26 MiniSat 2.2. `http://minisat.se/`. Accessed: 2022-02-09.

27 Period-48 glider gun – LifeWiki. `https://conwaylife.com/wiki/Period-48_glider_gun`. Accessed: 2022-02-09.

28 PySAT 0.1.7.dev15. `https://pysathq.github.io/`. Accessed: 2022-02-09.

29 Quadri-Snark – LifeWiki. `https://conwaylife.com/wiki/Quadri-Snark`. Accessed: 2022-02-09.

30 Ville Salo and Ilkka Törmä. Game of Life agars. `https://github.com/ilkka-torma/gol-agars`, 2022. GitHub repository.

31 Robert Wainwright. Lifeline vol. 6, 1972.

32 Hao Wang. Proving theorems by pattern recognition II. *Bell System Technical Journal*, 40:1–42, 1961.