THE LOGIC IN COMPUTER SCIENCE COLUMN

by

Yuri GUREVICH
Microsoft Research, Redmond, WA 98052, USA

A New Zero-One Law and Strong Extension Axioms

Andreas Blass[1] and Yuri Gurevich

Abstract

One of the previous articles in this column was devoted to the zero-one laws for a number of logics playing prominent role in finite model theory: first-order logic FO, the extension FO+LFP of first-order logic with the least fixed-point operator, and the infinitary logic $L^{\omega}_{\infty,\omega}$. Recently Shelah proved a new, powerful, and surprising zero-one law. His proof uses so-called strong extension axioms. Here we formulate Shelah's zero-one law and prove a few facts about these axioms. In the process we give a simple proof for a "large deviation" inequality à la Chernoff.

1 Shelah's Zero-One Law

Quisani: What are you doing, guys?

Author: We[2] are proving a zero-one law which is due to Shelah.

Q: Didn't Shelah prove the law?

A: Oh yes, he proved it all right, and even wrote it down [14].

Q: So what is the problem? Can't you read his proof?

[1]Mathematics, University of Michigan, Ann Arbor, MI 48109-1109, USA; partially supported by a grant from Microsoft Research.

[2]The record of the conversation was simplified by blending the two authors into one who prefers "we" to "I".

A: Reading Shelah's proofs may be research in its own right. His great mathematical talent is not matched by his talent of exposition.

Q: I suspect that you don't limit yourself to reproving Shelah's theorem.

A: We have proved some related results [1].

Q: Tell me about this zero-one law which is exciting enough to divert your attention from abstract state machines.

A: Actually the law is related to ASMs. It is about the BGS model of computation [3] which is based on ASMs. The model was defined with the intention of modeling computation with arbitrary finite relational structures as inputs, with essentially arbitrary data types, with parallelism, but without arbitrary choices.

Q: What arbitrary choices?

A: Here's an example of what we mean. Recall the Bipartite Matching Problem: Given a relation A between some number of boys and the same number of girls, find a subset of A that constitutes a one-to-one correspondence between the boys and the girls. The problem is solvable in polynomial time in the usual computation models where the input bipartite graph is given by means of some presentation such as the adjacency matrix; see for example the book [9]. Having such a presentation, the algorithm may start with the first boy. A BGS algorithm may be unable do that. The reason is that, contrary to the usual computation models, the bipartite graph may be given directly and not via some representation. In particular there may be no notion of "first." And there may be no way to choose an arbitrary boy.

Q: With parallelism, the lack of a choice mechanism makes no difference. An algorithm can produce all possible linear orderings of its input. If the input has size n, you have $n!$ independent subcomputations, each using one of the orderings to make whatever choices are needed. I guess the output is supposed to depend only on the isomorphism type of the input. So all these computations produce the same output.

A: That's right, if you have unlimited resources. We require the total computation time (summed over all parallel subprocesses) to be polynomially bounded. So there isn't time to construct all the linear orderings. The inability to make arbitrary choices really matters. According to [3], choiceless polynomial time, $\tilde{\text{C}}$PTime, the complexity class defined by BGS programs subject to a polynomial time bound, does not contain the Bipartite Matching Problem. In fact, $\tilde{\text{C}}$PTime does not contain even the parity problem: Given a set, determine whether its cardinality is even. The proofs build on symmetry considerations, i.e., on automorphisms of the input structure.

Subsequently, Shelah [14] proved a zero-one law for $\tilde{\text{C}}$PTime properties of graphs and similar structures. Notice a crucial difference from the earlier results in [3]: Almost all finite graphs have no non-trivial automorphisms, so symmetry considerations cannot be applied to them. Shelah's proof therefore depends on a more subtle concept of partial symmetry.

Q: We spoke once about zero-one laws [10]. I remember that one distinguishes, at least in principle, between labeled and unlabeled structures. In the case of labeled structures, one assumes that the base set of an n-element structure is the set $\{1, 2, \ldots, n\}$. The labels $1, \ldots, n$ help us in counting but they do not belong to the vocabulary of the structure. More importantly, a zero-one law requires a probability distribution on structures.

A: For simplicity, let us restrict our attention to labeled structures and the uniform probability distribution. Define the *probability* of a class (assumed closed under graph isomorphisms) of n-vertex graphs by considering all graphs with vertex set $\{1, 2, \ldots, n\}$ to be equally probable. This probability measure can also be defined by saying that, for each potential edge, i.e., each set of two distinct vertices, we flip a fair coin to decide whether to include the edge in our graph. It is presumed that the coin flips are independent.

The *asymptotic probability* of a class of graphs is defined as the limit, as $n \to \infty$, of the probability of its intersection with the class of n-vertex graphs with vertex set $\{1, 2, \ldots, n\}$. We sometimes refer to the probability of a property of graphs, meaning the probability of the class of graphs that have that property. In general, a zero-one law says that definable classes have asymptotic probability 0 or 1, but, as we shall see, some care is needed in formulating the zero-one law for $\tilde{\text{C}}$PTime.

Q: Explain $\tilde{\text{C}}$PTime.

A: To this end, we need to describe the BGS model of computation [3]. It is a version of the abstract state machine (ASM) paradigm [11]. The input to a computation is an arbitrary finite relational structure. For simplicity, let us restrict attention to the case when the input is an undirected, loopless graph $\langle V, A\rangle$ where V is the set of vertices and A is the adjacency relation. A state of the computation is a structure whose base set is the set $\text{HF}(V)$ which consists of the set V together with all hereditarily finite sets over it. $\text{HF}(V)$ is the smallest set containing all vertices in V (which are assumed to be atoms, not sets) and all finite subsets of itself. In other words, it is the union of the sets $\mathcal{P}_n(V)$ defined inductively by

$$\begin{aligned} \mathcal{P}_0(V) &= V \\ \mathcal{P}_{n+1}(V) &= V \cup \mathcal{P}_{\text{fin}}(\mathcal{P}_n(V)), \end{aligned}$$

where $\mathcal{P}_{\text{fin}}(X)$ means the set of all finite subsets of X.

The structure has the adjacency relation A, some set-theoretical apparatus (for example the membership relation $\in$), and some dynamic functions. The computation proceeds in stages, always modifying the dynamic functions in accordance with the program of the computation. The dynamic functions are initially constant with value $\emptyset$ and they change at only finitely many arguments at each step. So, although $\mathrm{HF}(V)$ is infinite, only a finite part of it is involved in the computation at any stage. The computation ends when and if a specific dynamic 0-ary function Halt acquires the value $\texttt{true} = \{0\}$, and the result of the computation is then the value of another dynamic 0-ary function Output.

Q: Why $\texttt{true} = \{0\}$?

A: We have adopted the convention that the truth values are identified with the first two von Neumann ordinals, $\texttt{false} = 0 = \emptyset$ and $\texttt{true} = 1 = \{0\}$. Recall that the finite von-Neumann ordinals represent the natural numbers by identifying n with the set $\{0, 1, \ldots, n-1\}$.

This model was used to define choiceless polynomial time $\tilde{\mathrm{C}}$PTime by requiring a computation to take only polynomially many (relative to the size of the input structure $\langle V, A\rangle$) steps and to have only polynomially many active elements.

Q: Which elements are active?

A: Roughly speaking, an element of $\mathrm{HF}(V)$ is active if it participates in the updating of some dynamic function at some stage.

Further, Output was restricted to have Boolean values, so the result of a computation could only be true, or false, or undecided.

Q: I guess the "undecided" situation arises if the computation exhausts the allowed number of steps or the allowed number of active elements without Halt becoming `true`.

A: Precisely. We shall use the phrase *polynomial time BGS program* to refer to a BGS program, with Boolean Output, together with polynomial bounds on the number of steps and the number of active elements.

Two classes $\mathcal{K}_0$ and $\mathcal{K}_1$ of graphs are $\tilde{\mathrm{C}}$PTime-*separable* if there is a polynomial time BGS program Π such that, for all input structures from $\mathcal{K}_0$ (resp. $\mathcal{K}_1$), Π halts with output `false` (resp. `true`) without exceeding the polynomial bounds. It doesn't matter what Π does when the input is in neither $\mathcal{K}_0$ nor $\mathcal{K}_1$.

Theorem 1.1 (Shelah's Zero-One Law) *If $\mathcal{K}_0$ and $\mathcal{K}_1$ are $\tilde{C}PTime$-separable classes of undirected graphs, then at least one of $\mathcal{K}_0$ and $\mathcal{K}_1$ has asymptotic probability zero.*

An equivalent formulation of this is that, for any given polynomial time BGS program, either almost all graphs produce output true or undecided or else almost all graphs produce output false or undecided.

Q: Is it true that either almost all graphs produce true, or almost all produce false, or almost all produce undecided?

A: No, we can give you a counterexample, but this would require a more thorough review of the definition of BGS programs.

The theorem was, however, strengthened considerably in another direction in [14]. It turns out that the number of steps in a halting computation is almost independent of the input.

Theorem 1.2 *Let a BGS program* Π *with Boolean output and a polynomial bound for the number of active elements be given. There exist a number* m*, an output value* v*, and a class* $\mathcal{C}$ *of undirected graphs, such that* $\mathcal{C}$ *has asymptotic probability one and such that, for each* $\langle V, A\rangle \in \mathcal{C}$*, either*

- Π *on input* $\langle V, A\rangle$ *halts after exactly* m *steps with output value* v *and without exceeding the given bound on active elements, or*
- Π *on input* $\langle V, A\rangle$ *either never halts or exceeds the bound on active elements.*

The proof of the theorem gives a somewhat more precise result. If there is even one input $\langle V, A\rangle \in \mathcal{C}$ for which Π eventually halts, say at step m, without exceeding the bound on active elements, then in the second alternative in the theorem the computation will exceed the bound on active elements at or before step m.

Notice that this theorem does not assume a polynomial bound on the number of steps. It is part of the conclusion that the number of steps is not only polynomially bounded but constant as long as the input is in $\mathcal{C}$ and the number of active elements obeys its bound.

Intuitively, bounding the number of active elements, without bounding the number of computation steps, amounts to a restriction on space, rather than time. Thus, Theorem 1.2 can be viewed as a zero-one law for choiceless polynomial space computation.

Q: Is the BGS logic more powerful than the extension FO+LFP of first-order logic with the least fixed-point operator?

A: Yes, every property definable in FO+LFP is $\tilde{\text{C}}$PTime, and there are $\tilde{\text{C}}$PTime properties that are not definable in FO+LFP. The BGS logic is quite expressive [4].

Q: In the case of FO+LFP, the almost sure theory, that is the set of sentences with asymptotic probability 1, is decidable. What about the almost sure theory of the BGS logic?

A: It is undecidable.

Proposition 1.3 *The class of almost surely accepting polynomially bounded BGS programs and the class of almost surely rejecting polynomially bounded BGS programs are recursively inseparable.*

The proof is easy. We can sketch it for you. Consider Turing machines with two halting states h_1 and h_2. For $i = 1, 2$, let H_i be the collection of Turing machines that eventually halt in state h_i if started on the empty input tape. It is well-known that H_1 and H_2 are recursively inseparable. Associate to each Turing machine T a polynomial time BGS program as follows. The program Π ignores its input graph and simulates T on empty input tape (working exclusively with pure sets). Π outputs `true` (resp. `false`) if T halts in state h_1 (resp. h_2). The polynomial bounds on steps and active elements are both twice the number of atoms. Then if $T \in H_1$ (resp. $T \in H_2$) our polynomial time BGS program will accept (resp. reject) all sufficiently large inputs.

Q: A question about the class $\mathcal{C}$ in Theorem 1.2. Do you just prove the existence of it?

A: The class $\mathcal{C}$ has a fairly simple description. It consists of the graphs that satisfy the so-called strong extension axioms for up to some number n of variables. The parameter n can be easily computed when the program Π and the polynomial bound on the number of active elements are specified.

2 Strong Extension Axioms

Q: What are strong extension axioms?

A: Do you remember the ordinary extension axioms?

Q: It would be good to review them. I remember though that the extension axioms played a key role in the zero-one law for first-order logic, that they explained the mystery of that law.

A: The ordinary extension axioms (for graphs) assert the existence of vertices in any possible "configuration" relative to finitely many given vertices; strong extension axioms assert not only existence but plentitude.

More precisely, a k-parameter *type* is a formula $\tau(y, x_1, \ldots, x_k)$ of the form

$$\bigwedge_{1 \le i \le k} (y \neq x_i \wedge \pm(yAx_i)).$$

Here $\pm$ before a formula means that the formula may or may not be negated. So τ specifies the adjacency and non-adjacency relationships between y and the k parameters x_i; in addition, it says that y is distinct from the x_i's (which is redundant when yAx_i is not negated, since the adjacency relation A is irreflexive). The *extension axiom* $\mathrm{EA}(\tau)$ associated to a type τ is

$$\forall x_1, \ldots, x_k \left(\left(\bigwedge_{1 \le i < j \le k} x_i \neq x_j \right) \rightarrow \exists y \, \tau(y, x_1, \ldots, x_k) \right).$$

For a fixed k, there are 2^k of these extension axioms, because of the k choices for the $\pm$ signs in τ. We write EA_k for their conjunction together with the statement that there are at least k vertices (so that the $\mathrm{EA}(\tau)$'s aren't vacuous). Thus, EA_k says that there exist at least k vertices and that every possible configuration for a vertex y, relative to k distinct, given vertices, is realized at least once.

We say that a graph satisfies the *strong extension axiom* $\mathrm{SEA}(\tau)$ if, for every k distinct vertices $x_1, \ldots, x_k$, there are at least $\frac{1}{2}n/2^k$ vertices y satisfying $\tau(y, x_1, \ldots, x_k)$. Unlike the extension axioms, strong extension axioms are not first-order formulas. We write SEA_k for the conjunction of all 2^k of the strong extension axioms $\mathrm{SEA}(\tau)$ as τ ranges over all the k-parameter types, together with the statement that there are at least k vertices. Thus, SEA_k says that there exist at least k vertices and that each possible configuration of y relative to k distinct x_i's is realized not just once (as EA_k says) but fairly often, $\frac{1}{2}n/2^k$ times.

Q: I remember seeing somewhere a stronger version of extension axioms.

A: Phokion Kolaitis and Moshe Vardi introduced a version of strong extension axioms with $\sqrt{n}$ instead of $\frac{1}{2}n/2^k$. They proved that their extension axioms were almost surely true and then used the axioms to derive a zero-one law for a fragment of second-order logic [12].

Q: Where did that number $\frac{1}{2}n/2^k$ come from?

A: Consider any particular k-parameter type with fixed values for the k parameters, say $\tau(y, a_1, \ldots, a_k)$. On the average, how many vertices would you expect to realize this type?

Q: Well, since τ says that y is distinct from all the a_i, there are $n - k$ vertices that could conceivably realize τ, and each of them has probability $1/2^k$ of realizing τ. So the expected number of realizers is $(n - k)/2^k$.

A: Right. Since k is fixed and we are interested in asymptotics for large n, this expected number is very nearly $n/2^k$.

Q: And the strong extension axiom says that the type has at least half the expected number of realizers. That sounds plausible, but why "half"?

A: It doesn't matter. We could use any constant strictly smaller than 1, and $\frac{1}{2}$ is the simplest choice. The analogous axiom with a constant α in place of $\frac{1}{2}$ will be denoted by SEA_k^α.

Q: I said the strong extension axiom sounds plausible, but now I think that it is definitely true.

A: You are right.

Proposition 2.1 *For each k, the asymptotic probability of SEA_k is 1.*

Q: I see how to prove it, using the central limit theorem. Here's the idea. First, I can ignore the distinction between $n-k$ and n as you suggested, because it can be compensated for by slightly increasing the $\frac{1}{2}$. So let me pretend there are n rather than $n-k$ vertices that could conceivably realize $\tau(y, a_1, \ldots, a_k)$. The number of these vertices that actually realize it is a Bernoulli random variable X with mean $n/2^k$ and standard deviation proportional to $\sqrt{n}$. I don't remember the constant of proportionality, but I don't think it'll matter. So for $\tau(y, a_1, \ldots, a_k)$ to be realized fewer than the desired number of times would mean that X differs from its mean by an amount linear in n, and that's more than some constant times $\sqrt{n}$ standard deviations. That probability can be estimated, for large n, by the central limit theorem, and it decreases exponentially (or at least like $\exp(-\sqrt{n})$ — again I don't remember exactly but it won't matter). Now there are only polynomially many (at most n^k) choices for $a_1, \ldots, a_k$, so the probability that *some* choice of the a_i's has fewer than the desired number of y's realizing τ still approaches zero.

A: This sounds good, but unfortunately the central limit theorem doesn't quite provide the information you need.

Q: What's the problem?

A: Well, the central limit theorem says that, as n approaches infinity, the probability that a Bernoulli random variable is more than β standard deviations below its mean approaches

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{-\beta} \exp\left(-\frac{1}{2}t^2\right)\, dt.$$

But it says this for each fixed β, not for a β that depends on n. And you needed a β proportional to $\sqrt{n}$.

Q: I guess you're right. So is there a better version of the central limit theorem that salvages the argument?

A: Actually, there's an extensive theory of so-called large deviations, designed to handle just this sort of thing. To establish that the strong extension axioms have asymptotic probability one, we need only a little bit of that theory — a version of Chernoff's inequality [7].

Lemma 2.2 *Fix numbers β, r in the open interval $(0,1)$. There is a constant c, also in $(0,1)$, such that the following is true for every positive integer m. Let X be the number of successes in m independent trials, each trial having probability r of success. Then Prob $[X \leq \beta mr] \leq c^m$.*

That is, the probability that the number of successes (X) is smaller than the expected number (mr) by at least the (fixed) factor β decreases exponentially as a function of the number m of trials. The proof of this proposition depends on a "large deviation" inequality of the sort given in Chernoff's paper [7] and Loève's book [13, Section 18]. The references we have found prove stronger results than we need and therefore give more complicated proofs than we need. But here is a simple proof of an inequality strong enough for our purposes.

Proof We begin with the well-known observation that, if Z is a non-negative random variable and q is a positive real number, then

$$\text{Prob}\,[Z \geq q] \leq \frac{E(Z)}{q},$$

where E means "expectation." Indeed,

$$\begin{aligned} E(Z) &= E(Z|Z \geq q)\text{Prob}\,[Z \geq q] + E(Z|Z < q)\text{Prob}\,[Z < q] \\ &\geq E(Z|Z \geq q)\text{Prob}\,[Z \geq q] \\ &\geq q \cdot \text{Prob}\,[Z \geq q]\,. \end{aligned}$$

We apply this with $Z = \exp[t(mr - X)]$, where t is a positive parameter to be chosen later. Thus, we have

$$\begin{aligned} \text{Prob}\,[X \leq \beta mr] &= \text{Prob}\,[Z \geq \exp[t(mr - \beta mr)]] \\ &= \text{Prob}\,[Z \geq \exp[tmr(1-\beta)]] \\ &\leq \frac{E(Z)}{\exp[tmr(1-\beta)]}. \end{aligned}$$

We continue by computing $E(Z)$. The random variable $mr - X$ can be viewed as the sum, over all m trials, of $r - S$, where S is 1 if the trial is a success and 0 if not. Thus, Z is the product over all trials of $\exp[t(r-S)]$. But the trials are independent, so the expectation of this product is the product of the individual expectations. For each individual trial, we have

$$\begin{aligned} E(\exp[t(r-S)]) &= r \cdot \exp[t(r-1)] + (1-r) \cdot \exp[t(r-0)] \\ &= \exp[tr](r\exp[-t] + 1 - r). \end{aligned}$$

Therefore,

$$E(Z) = \exp[tmr](r\exp[-t] + 1 - r)^m.$$

Substituting this into the inequality for $\text{Prob}\,[X \leq \beta mr]$, we find

$$\begin{aligned} \text{Prob}\,[X \leq \beta mr] &\leq \left(\frac{\exp[tr](r\exp[-t] + 1 - r)}{\exp[tr(1-\beta)]}\right)^m \\ &= [\exp[t\beta r] \cdot (r\exp[-t] + 1 - r)]^m. \end{aligned}$$

So the lemma will be proved if we can find a positive t for which the value of

$$f(t) = \exp[t\beta r] \cdot (r\exp[-t] + 1 - r)$$

is in the open interval $(0, 1)$, for then this value can serve as the required c.

Notice that $f(0) = 1$ and that

$$f'(t) = \beta r \exp[t\beta r] \cdot (r\exp[-t] + 1 - r) + \exp[t\beta r] \cdot (-r)\exp[-t].$$

Thus, $f'(0) = \beta r - r < 0$ (because $\beta < 1$ and $r > 0$). Therefore, any sufficiently small positive t will give $0 < f(t) < 1$ as required. □

Q: The best, i.e., smallest value of c obtainable by the preceding argument is the minimum value of $f(t)$.

A: Right. A routine calculation, setting $f'(t) = 0$, shows that this minimum is

$$\left(\frac{1-r}{1-\beta r}\right)^{1-\beta r} \cdot \left(\frac{1}{\beta}\right)^{\beta r}.$$

Notice that this is the weighted geometric mean of two quantities whose correspondingly weighted arithmetic mean is

$$\left(\frac{1-r}{1-\beta r}\right) \cdot (1-\beta r) + \left(\frac{1}{\beta}\right) \cdot (\beta r) = 1.$$

Since the two quantities are not equal to 1 (as $\beta < 1$), the arithmetic-geometric mean inequality shows again that the optimal c is smaller than 1.

Q: Now prove the proposition.

A: All right.

Proof of Proposition 2.1 We shall show that, for each fixed k-parameter type τ, the probability that SEA(τ) fails, in a random graph on vertex set $\{1, 2, \ldots, n\}$, approaches 0 as $n \to \infty$. Then, as SEA_k is the conjunction of a fixed number 2^k (independent of n) of SEA(τ)'s, its probability of failure also approaches 0, as required.

So we concentrate henceforth on a single τ. Temporarily, also concentrate on k specific, distinct vertices $a_1, \ldots, a_k \in \{1, 2, \ldots, n\}$. Let X be the number of vertices b satisfying $\tau(b, a_1, \ldots, a_k)$. In a random graph, each of the $n - k$ vertices other than $a_1, \ldots, a_k$ has probability $1/2^k$ of satisfying τ, and these $n - k$ trials are independent. So, applying the lemma with $m = n - k$, with $r = 1/2^k$, and with some β in the interval $(\frac{1}{2}, 1)$, and noting that, as $\beta > \frac{1}{2}$, we have $\beta \cdot (n - k) \geq \frac{1}{2}n$ for large n, we obtain some $c \in (0, 1)$ such that

$$\begin{aligned} \text{Prob}\left[X < \frac{1}{2}n/2^k\right] &\leq \text{Prob}\left[X < \beta(n-k)/2^k\right] \\ &\leq c^{n-k}. \end{aligned}$$

This bounds the probability that our specific choice of $a_1, \ldots, a_k$ is a counterexample to SEA(τ).

Now un-fix $a_1, \ldots, a_k$. Since the number of choices for this k-tuple is $\leq n^k$, the probability that at least one choice gives a counterexample, i.e., the probability that SEA(τ) fails, is at most

$$n^k c^{n-k}.$$

Since $0 < c < 1$, this bound approaches 0 as $n \to \infty$. □

The same proof can be used to show that, for each k and each $\alpha \in (0, 1)$, the axiom SEA_k^α has asymptotic probability 1.

3 Inadequacy of Extension Axioms

Q: You needed strong extension axioms to define the class $\mathcal{C}$ in Theorem 1.2. Might the ordinary extension axioms suffice to define $\mathcal{C}$? Maybe the use of strong extension axioms is just an artifact of the proof.

A: Ordinary extension axioms are too weak to support the zero-one law for $\tilde{\text{C}}$PTime. We will show you a particular polynomial-time BGS program that separates structures satisfying arbitrarily many extension axioms. So strong extension axioms are really needed for the $\tilde{\text{C}}$PTime zero-one law.

Though our general policy has been, for expository purposes, to concentrate on algorithms whose inputs are graphs, this program will use as input a graph together with a single distinguished vertex, i.e., a rooted graph. That is, we add a constant symbol d to the vocabulary $\{A\}$ of graphs. We expect that a similar example could be given without introducing the constant symbol.

Notice that the zero-one laws for the logics that you mentioned above, e.g. FO+LFP, continue to hold and to follow from the extension axioms, in the presence of a distinguished vertex.

Q: They fail when there are two distinguished vertices, because these two are adjacent with probability $\frac{1}{2}$.

A: Right. But one distinguished vertex is benign. Instead of a distinguished vertex, we could add a unary relation R to the vocabulary and modify the extension axioms to specify, in addition to adjacency information, whether y should satisfy R. In that version of the construction, R would play the role played in our proof by the set of neighbors of d.

Proposition 3.1 *There is a polynomial time BGS program* Π *such that, for any given* k*, there are two rooted graphs, both satisfying* EA_k*, such that* Π *produces output* `true` *on one of them and* `false` *on the other.*

Proof We begin by exhibiting the BGS program Π; the polynomial bounds on the number of steps and the number of active elements will be n and $2n + 3$, respectively. The program Π computes the parity of the maximum size of a clique containing the distinguished vertex d. It does this by building up the collection of all i-element subsets of $\{x : xAd\}$ for $i = 0, 1, \ldots$, checking at each step whether any cliques remain. One essential ingredient of the proof will be that d has so few neighbors in our graphs that the time used by this computation is polynomial relative to the sizes of these graphs.

Π uses four dynamic 0-ary function symbols: Halt, Output, Mode, and C. Recall that in the initial state of a computation these have the value $\emptyset =$ `false` $= 0$. The program Π is

```
do in parallel
   if Mode = 0 then
      do in parallel
         C := {∅}, Mode := 1
      enddo
   endif
   if Mode = 1 then
      do in parallel
         C := {x ∪ {y} : x ∈ C, y ∈ Atoms : yAd ∧ y ∉ x},
         Output := ¬Output, Mode := 2
      enddo
   endif
   if Mode = 2 then
      if (∃x ∈ C)(∀u, v ∈ x) uAv
      then Mode := 1
      else Halt := true
      endif
   endif
enddo.
```

After the first part of Π, with Mode $= 0$, has been executed, C is initialized to $\{\emptyset\}$, the family of 0-element subsets of the set R of neighbors of d. After i executions of the part with Mode $= 1$, C has become the family of i-element subsets of R. The part with Mode $= 2$ checks whether there are any cliques in C. If so, we return to Mode 1 to enlarge the sets in C; if not, then the common size i of the sets in C is one more than the maximum size of a clique included in R. Since Output reverses its truth value at each Mode 1 step and since it is initially `false`, we see that, the final value of Output is `true` if and only if the maximum clique size in R is even, if and only if the maximum size of a clique containing d is odd.

Let us estimate the number of steps and the number of active elements in a computation of our program. Writing n for the number of vertices in the input graph, r for the number of neighbors of d, and s for the maximum size of a clique among these neighbors, we find that the Mode 0 part of Π is executed once and the Mode 1 and Mode 2 parts are executed $s + 1$ times each. So the whole computation takes $2s + 3$ steps.

According to the definition in [3], the truth values, the atoms and the set of atoms, altogether $n + 3$ entities, are active already in the initial state. The elements that the computation activates are the number 2 (which is one of the values of Mode), the subsets of R of cardinality at most $s + 1$ and the $s + 2$ values taken by C. For non-trivial values of r and s, the number of activated elements is easily seen to be majorized by $(r + 1)^{s+1}$.

Q: But 0 is also among the subsets of R and $1 = \{0\}$ is also among the values of C.

A: So we have a slight overcount of activated elements. In any case, the total number of active elements in this computation is at most $n+3+(r+1)^{s+1}$. We shall design our graphs to have quite small r and s (relative to n), so that $n + 3 + (r + 1)^{s+1}$ is below the bound $2n + 3$ that we imposed on the number of active elements, and $2s + 3$ is below the bound n on the number of steps.

The graphs required in the proposition will be described in three steps. First, we give a general description depending on two parameters: the (large) number n of vertices and the (much smaller but still rather large) number r of neighbors of d. Second, leaving n arbitrary (but large), we prescribe two values for r, to produce the two graphs we want. Finally, we fix n so large that these graphs have all the required properties. Actually, the description in the first step involves randomization, and rather than fixing n in the last step we simply show that, for all sufficiently large n, the graphs have the required properties with high probability. This clearly suffices for the existence claim in the proposition.

Given n and r, we build a (random) graph $G(n, r)$ as follows. The vertex set consists of the distinguished vertex d and $n-1$ others which we denote by $1, 2, \ldots, n-1$.

Q: This introduces an ambiguity, since these vertices are atoms in $HF(I)$ and the same symbols denote natural numbers (finite von Neumann ordinals) which are sets.

A: Fortunately, the ambiguity never leads to a confusion. Let's proceed. d is adjacent to the vertices $1, 2, \ldots, r$ and no others. The rest of the adjacency relation is chosen at random; flip independent, fair coins for all potential edges to decide whether to include them in the graph. This completes the description of $G(n, r)$. Notice that the subgraph induced by the set $R = \{1, 2, \ldots, r\}$ of neighbors of d is a random graph (in the usual sense) on r vertices.

The next part of the proof, choosing r as a function of n, is the most delicate. We need r small enough so that Π stays within the bound on active elements, but if we take r too small then $G(n, r)$ will violate the extension axioms. We need the following result from [6, Section XI.1].

Lemma 3.2 *There is a function ρ from natural numbers to natural numbers with the following two properties. First,*

$$C_1 s 2^{s/2} \leq \rho(s) \leq C_2 s 2^{s/2}$$

for certain positive constants C_1 and C_2. Second, if $p(s)$ denotes the probability that a random graph on $\rho(s)$ vertices has maximum clique size exactly s, then $p(s) \to 1$ as $s \to \infty$.

Actually, Bollobás proves a far more precise result. The constants in the lemma can be taken to be any constants satisfying $C_1 < 1/(e\sqrt{2}) < C_2$ provided s is sufficiently large. All we shall need, however, is the lemma as stated.

Using this lemma, we associate to each (large) n two values of r as follows. Let s and s' be the two largest integers below $3 \log\log n$.

Q: What is the base of logarithm?

A: We use log to mean base 2 logarithm and ln to mean base e logarithm.

Let $r = \rho(s)$, $r' = \rho(s')$, $G = G(n, r)$, and $G' = G(n, r')$. According to the lemma, when n is large enough there is a very high probability that, among the neighbors of d, the largest clique in G has size s and similarly for G' and s'. In particular, since s and s' are consecutive integers, the program Π will (unless it runs out of time) produce output `true` for one of G and G' and `false` for the other.

We next address the question whether Π with these inputs G and G' succeeds in carrying out its computation within the bounds on the number of steps (n) and active elements ($2n + 3$). We already computed the number of steps and an upper bound for the number of active elements. In the present context, these are, with high probability for large n,

$$2s + 3 < 3 \log\log n + 3 < n$$

and

$$n + 3 + (r + 1)^{s+1} < n + 3 + \left(C_3 s 2^{s/2}\right)^{s+1}$$

for G (where C_3 is slightly larger than C_2 to compensate for changing from $r+1$ to r), and similarly for G' with s' and r' in place of s and r. So the bound on steps is satisfied with high probability for sufficiently large n. As for the bound on active elements, we must show that

$$\left(C_3 s 2^{s/2}\right)^{s+1} \leq n.$$

To this end, we first compute that, since $s < 3 \log\log n$,

$$C_3 s 2^{s/2} < C_3 \cdot 3 \log\log n \cdot (\log n)^{3/2} < (\log n)^2$$

for large n. The desired inequality follows because the logarithm of its left side is at most

$$(s + 1) \log\left((\log n)^2\right) < (3 \log\log n + 1) \cdot 2 \log\log n < \log n.$$

To complete the proof of the proposition, we must still verify that (with high probability, when n is large) G and G' satisfy EA_k. We give the argument for G; it applies equally well to G'. Recall that G was defined as $G(n, r)$ with n sufficiently

large and $r = \rho(s) \geq C_1 s 2^{s/2}$, where s is one of the largest two integers below $3 \log\log n$. In particular, $s > 2 \log\log n$ (for large n), and so

$$r \geq C_1 \cdot 2 \log\log n \cdot \log n = \gamma(n) \log n,$$

where all we need to know about $\gamma(n) = C_1 \cdot 2 \log\log n$ is that it tends to infinity with n. We can therefore complete the proof by showing that, for each fixed k-parameter type τ, the probability that $G(n, r)$ satisfies EA(τ) is close to 1 when n is large and $r \geq \gamma(n) \log n$.

Fix, therefore, an arbitrary k-parameter type τ, and temporarily fix values $a_1, \ldots, a_k$ for its parameters. There are three cases to consider, according to whether d is among the parameters and, if it is, whether τ says y should be adjacent or non-adjacent to d. Since d has so few neighbors (only r, compared with approximately $n/2$ for other vertices), the probability that $\tau(y, a_1, \ldots, a_k)$ holds is smallest in the case where some a_i is d and τ says y is adjacent to d. We calculate this case first and then indicate the changes for the other cases.

So suppose τ says y must be adjacent to d. Then the only candidates for values of y satisfying τ are the r neighbors $1, 2, \ldots, r$ of d. For any one of these neighbors b, distinct from the other parameters, the probability that it satisfies τ, i.e., that it satisfies $k-1$ additional adjacency or non-adjacency requirements each of which has probability $\frac{1}{2}$, is $1/2^{k-1}$. Since these probabilities are independent for different b's and since there are at least $r - k + 1$ available b's (the r neighbors of d minus at most $k-1$ that are among the other parameters),

$$\begin{aligned} \text{Prob}\left[\text{no } b \text{ satisfies } \tau(y, a_1, \ldots, a_k)\right] &\leq \left(1 - \frac{1}{2^{k-1}}\right)^{r-k+1} \\ &\leq e^{-(r-k+1)/2^{k-1}}, \end{aligned}$$

where we used the fact that $1 - t \leq e^{-t}$.

In the case where d is one of the parameters but τ says that y is not adjacent to d, the computation works the same way but with $n - r - k$ in place of $r - k + 1$. In the case where d is not one of the parameters, the result is again similar but with $(n-k)/2$ in place of $r - k + 1$. In either of these cases, $r - k + 1$ has been replaced with something larger (when n is large), so the upper bound for the probability of failure is even smaller than in the first case. Summarizing, we have, for every choice of k distinct parameters, an upper bound of $e^{-(r-k+1)/2^{k-1}}$ for the probability that τ has no solution. Therefore, the probability that EA(τ) fails is at most

$$\binom{n}{k} e^{-(r-k+1)/2^{k-1}} \leq n^k e^{-(r-k+1)/2^{k-1}}.$$

To estimate this, we consider its natural logarithm, which is at most

$$k \ln n - \frac{r-k+1}{2^{k-1}} \leq -\frac{\gamma(n) \log n}{2^{k-1}} + k \ln n + \text{constant}.$$

Since $\gamma(n) \to \infty$ as $n \to \infty$, the right side of this formula tends to $-\infty$. So our upper bound for the probability that $\mathrm{EA}(\tau)$ fails tends to 0 as $n \to \infty$. □

It should perhaps be pointed out explicitly that the graphs constructed in the preceding proof violate SEA_1, for the number of neighbors of the special vertex d is far smaller than the $n/4$ that SEA_1 would require.

4 Rigidity and Hamiltonicity

Q: From the conversation [10] I remember that almost all finite graphs are rigid[3] but rigidity does not follow from the ordinary extension axioms. Does it follow from strong extension axioms?

A: No dice.

Proposition 4.1 *For any k, there exists a non-rigid graph satisfying SEA_k.*

Proof We use the same randomizing construction as in [5]. Let l be a large natural number, and let $G(l)$ have the integers from $-l$ through l as vertices. We require that, if a is adjacent to b, then $-a$ is adjacent to $-b$; this ensures that $G(l)$ is not rigid, for $a \mapsto -a$ is a non-trivial automorphism. Except for this symmetry requirement, $G(l)$ is random. That is, for each pair of corresponding potential edges $\{a, b\}$ and $\{-a, -b\}$, we decide whether to include both or neither in $G(l)$ by flipping a fair coin. We shall show that, for any fixed k, the probability that the graph $G(l)$ satisfies SEA_k tends to infinity with l.

Q: The pair of potential edges $\{a, b\}$ and $\{-a, -b\}$ is a single potential edge if $a = -b$, but I guess this does not change anything.

A: Right. As usual, it suffices to check the asymptotic probability for $\mathrm{SEA}(\tau)$ for every single k-parameter type τ. So let τ be given, and temporarily fix values $a_1, \ldots, a_k$ for the parameters. Let b range over positive vertices different from all the $\pm a_i$'s. For each such b, the probability that it satisfies τ with our fixed parameters is $1/2^k$, and for different b's these probabilities are independent. By Lemma 2.2, for any $\beta \in (0, 1)$, the probability that fewer than $\beta \cdot \frac{l-k}{2} \cdot 2^{-k}$ of these b's satisfy τ decreases exponentially with l. Of course the same goes for negative b's. Therefore, the same goes for the probability that fewer than $\beta \cdot (l-k) \cdot 2^{-k}$ vertices altogether satisfy τ.

Q: Why do you consider positive and negative b's separately?

[3] A graph is *rigid* if its only automorphism is the identity.

A: Because their behavior is not independent as required for application of Lemma 2.2.

Taking β slightly larger than $\frac{1}{2}$, to get $\beta \cdot (l-k) > \frac{1}{2} \cdot l$ for large l, we find that the probability that our fixed $a_1, \ldots, a_k$ constitute a counterexample to SEA(τ) decreases exponentially with l.

Now un-fix the parameters a_i. Notice that the number of choices of the parameters is bounded by a polynomial in l, namely $(2l+1)^k$. Therefore, the probability that SEA(τ) fails is small for large l. □

We remark that the proof shows that even the axioms SEA_k^α for arbitrary $\alpha \in (0,1)$ do not imply rigidity.

Q: What about hamiltonicity? From the sane conversation [10] I remember that almost all finite graphs are hamiltonian[4] but hamiltonicity does not follow from the ordinary extension axioms. Does it follow from strong extension axioms?

A: In the case of hamiltonicity, we have only a partial answer.

Proposition 4.2 *For any k and any $\alpha \in (0, \frac{1}{2})$, there is a graph that satisfies SEA_k^α but is not hamiltonian.*

Proof We simplify a randomizing construction from [5]. Let l be a large natural number, and let $G(l)$ be the graph produced as follows. The vertices are the natural numbers from 0 to $2l$; we call the first l of these vertices *friendly* and the remaining $l+1$ *unfriendly*. No two unfriendly vertices will be adjacent. For each potential edge subject to this constraint, i.e., for each two vertices of which at least one is friendly, flip a fair coin to decide whether to include that edge in $G(l)$.

No matter what happens in the randomization, this graph cannot be hamiltonian. Indeed, in any cycle, at most half the vertices can be unfriendly, since unfriendly vertices are not adjacent. But in the whole graph, more than half of the vertices are unfriendly.

To complete the proof, we show that, with high probability for large l, $G(l)$ satisfies SEA_k^α. As usual, it suffices to show, for each fixed k-parameter type τ and each choice of values $a_1, \ldots, a_n$, that the probability that fewer than $\alpha \cdot (2l+1) \cdot 2^{-k}$ vertices b satisfy τ (with the chosen parameters) decreases exponentially as a function of l. So let τ and the parameters be fixed, and let b range over friendly vertices distinct from all the parameters. So there are at least $l-k$ values for b, and each satisfies τ with probability 2^{-k}, these events being independent for different b's. We apply Lemma 2.2 with $\beta > 2\alpha$ to this situation. Since $\alpha < \frac{1}{2}$, we can find such a $\beta < 1$, so Lemma 2.2 is applicable. It gives us exponentially decreasing

[4] A graph is *hamiltonian* if it includes a cycle containing all its vertices.

probability for the event that fewer than $\beta \cdot (l-k) \cdot 2^{-k}$ friendly vertices satisfy τ with the chosen parameters. But since $\beta > 2\alpha$, we have $\beta \cdot (l-k) > \alpha \cdot (2l+1)$ once l is large enough. Thus, we also have exponentially decreasing probability for the event that fewer than $\alpha \cdot (2l+1)$ vertices satisfy τ with the chosen parameters. □

We do not know whether the preceding proposition can be extended to $\alpha \geq \frac{1}{2}$. Its proof made crucial use of the assumption that $\alpha < \frac{1}{2}$. The non-hamiltonicity of the graph depended on the presence of an independent set containing more than half the vertices (the unfriendly ones), and no such set can exist when $\alpha > \frac{1}{2}$. More precisely, we have the following proposition, which prevents any construction like the preceding one from working when $\alpha > \frac{1}{2}$. (The specific randomizing construction in the preceding proof doesn't work for $\alpha = \frac{1}{2}$ either.)

Proposition 4.3 *Let $\alpha > \frac{1}{2}$. There exists a number k (depending on α) such that, for every sufficiently large n and every n-vertex graph satisfying SEA_k^α, no independent set contains more than half of the vertices.*

Proof Given α, choose k so large that

$$\left(1 - \frac{1}{2^k}\right) \alpha > \frac{1}{2},$$

and let n be much larger yet. Consider an n-vertex graph satisfying SEA_k^α, and suppose it had an independent set U of cardinality at least $n/2$. Fix k distinct elements $a_1, \ldots, a_k \in U$, and consider the axioms $\mathrm{SEA}(\tau)$ applied to these k parameters, where τ ranges over all k-parameter types *except* the one that says y is adjacent to none of the parameters. Thus, we are considering $2^k - 1$ types, and each can be satisfied only by elements outside U. Thus, these types altogether have at most $n/2$ elements satisfying them. But, by SEA_k^α, each of them is realized by at least $\alpha \cdot n \cdot 2^{-k}$ vertices, so altogether they are realized by at least $\alpha \cdot n \cdot 2^{-k} \cdot (2^k - 1)$ vertices. But this number exceeds $n/2$ by our choice of k. □

Q: I guess you told me all that you know about strong extension axioms.

A: Actually, [1] has a little bit more information. By the way, most of Shelah's proof (of his zero-one law for $\tilde{\mathrm{C}}\mathrm{PTime}$) uses ordinary extension axioms. Only one combinatorial lemma requires strong extension axioms. But it looks like you had enough strong extension axioms.

Q: Indeed enough, for today anyway.

References

[1] Andreas Blass and Yuri Gurevich, *Strong Extension Axioms and Shelah's Zero-One Law for Choiceless Polynomial Time*, in preparation. See an abridged version of the paper, named *Choiceless Polynomial Time Computation and the Zero-One Law*, in Computer Science Logic 2000, editors Peter Clote and Helmut Schwichtenberg, Springer Lecture Notes in Computer Science 1862 (2000) 18–40.

[2] Andreas Blass, Yuri Gurevich, and Dexter Kozen, *A zero-one law for logic with a fixed-point operator*, Information and Control 67 (1985) 70–90.

[3] Andreas Blass, Yuri Gurevich, and Saharon Shelah, *Choiceless polynomial time*, Ann. Pure Applied Logic 100 (1999) 141–187.

[4] Andreas Blass, Yuri Gurevich and Jan Van den Bussche, *Abstract state machines and computationally complete query languages*, in "Abstract State Machines: Theory and Applications", editors Y. Gurevich et al., Springer Lecture Notes in Computer Science 1912, 2000.

[5] Andreas Blass and Frank Harary, *Properties of almost all graphs and complexes*, J. Graph Theory 3 (1979), 225–240.

[6] Béla Bollobás, Random graphs, Academic Press, 1985.

[7] H. Chernoff, *A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations*, Ann. Math. Statist. 23 (1952), 493–507.

[8] Heinz-Dieter Ebbinghaus and Jörg Flum, *Finite Model Theory*, Springer-Verlag, 1995.

[9] Shimon Even, *Graph Algorithms*, Computer Science Press, 1979.

[10] Yuri Gurevich, *Zero-One Laws*, Bulletin of the European Association for Theoretical Computer Science, No. 46 (Feb. 1992), 90–106. [Reprinted in G. Rozenberg and A. Salomaa, editors, *Current Trends in Theoretical Computer Science*, World Scientific, 1993, 293–309.

[11] Yuri Gurevich, *Evolving algebras 1993: Lipari guide*, in Specification and Validation Methods, ed. E. Börger, Oxford University Press (1995) pp. 9–36. See also the *May 1997 draft of the ASM guide*, Tech Report CSE-TR-336-97, EECS Dept., Univ. of Michigan, 1997. Found at `http://www.eecs.umich.edu/gasm/`.

[12] Phokion Kolaitis and Moshe Vardi, *0–1 laws and decision problems for fragments of second-order logic*, Information and Computation 87 (1990) 302–339.

[13] M. Loève, Probability Theory, Van Nostrand 1955.

[14] Saharon Shelah, *Choiceless polynomial time logic: inability to express* [paper number 634], in Computer Science Logic 2000, editors Peter Clote and Helmut Schwichtenberg, Springer Lecture Notes in Computer Science 1862 (2000) 72–125.