Pseudorandom Generators, Measure Theory, and Natural Proofs

Kenneth W. Regan [*] D. Sivakumar [†] Jin-yi Cai [‡]
Department of Computer Science,
State University of New York at Buffalo, Buffalo, NY 14260.
Email:{regan,sivak-d,cai}@cs.buffalo.edu

Abstract

*We prove that if strong pseudorandom number generators exist, then the class of languages that have polynomial-sized circuits (*P/poly*) is not measurable within exponential time, in terms of the resource-bounded measure theory of Lutz. We prove our result by showing that if* P/poly *has measure zero in exponential time, then there is a* natural proof *against* P/poly*, in the terminology of Razborov and Rudich [25]. We also provide a partial converse of this result.*

1 Introduction

The theory of resource-bounded measure, initiated by Lutz [13], provides a useful framework that links many central problems in complexity theory. Given a measure defined on a large complexity class, such as EXP = DTIME$[2^{n^{O(1)}}]$, and a subclass $\mathcal{C}$ such as P, NP, or PSPACE, one tries to determine whether $\mathcal{C}$ has measure zero, has measure one, or perhaps is not measurable at all.

For example, P has measure zero in EXP. In fact, for any fixed $c > 0$, DTIME$[2^{n^c}]$ has measure zero in EXP. The class of P-bi-immune sets in EXP has measure one [19]. Lutz [13, 15] has advanced the hypothesis that NP does not have measure zero, which implies NP $\neq$ P. Indeed, the hypothesis implies that NP has P-bi-immune sets, and that for every $c > 0$, there are languages in NP that require deterministic time more than 2^{n^c}. Lutz and Mayordomo [17] showed another plausible implication: there would be NP-complete sets under Turing (Cook) reductions that are not complete under many-one (Karp) reductions. In view of this, it is important to seek techniques for proving that certain subclasses do not have measure zero, or are non-measurable. This paper provides a new technique of this kind, using the theory of pseudorandom generators (PSRGs).

The meaning of a class $\mathcal{C}$ having measure zero in EXP is, roughly speaking, that there is a *single* exponential time deterministic Turing Machine M that can "predict" *every* language in $\mathcal{C} \cap$ EXP reasonably well. This M embodies a strong and quantitative form of diagonalization. A prime motivation of the theory is that the notion of measure should connect to quantitative notions of cryptographic hardness and randomness and information content that are important in other areas of complexity. Lutz and Mayordomo [16] showed that, for any fixed c, the class of languages that "appear random" to all 2^{n^c} time-bounded machines has measure one in EXP. Lutz [15] showed a measure-one class in which every member is a pseudorandom source for BPP, and Allender and Strauss [1] extended this for measures on DTIME$[2^{n^{\epsilon}}]$, for every $\epsilon > 0$. Related work is [14, 18, 11, 12, 20, 31]. Our main theorem relates a measure question directly to PSRGs and the class P/poly of languages having polynomial-sized circuits:

Theorem 1 *If strong PSRGs exist, then* P/poly *is not measurable in* EXP.

Here *strong* means that there exists some $\epsilon > 0$ such that the PSRG is secure against $2^{n^{\epsilon}}$-sized circuits. There are PSRGs based on the discrete logarithm problem that are widely believed to be strong, indeed with ϵ approaching $1/2$.

We first prove that if strong PSRGs exist, then for any fixed exponential time deterministic TM M, there is a large collection of "pseudorandom" languages in P/poly that cannot be predicted by M. For the *non*-measurability, we show that if P/poly $\neq$ EXP, as implied by the existence of strong PSRGs, then P/poly cannot have measure *one* in EXP. Our proof of this shows that NP and many other classes cannot have measure one in EXP unless they equal EXP. This answers a question left open by Lutz for NP in [15].

What is interesting about Theorem 1 is that ordinarily P/poly is considered a "feasible" class, intuitively smaller than a "hard" class like NP. What our result really brings out is the quantitative role of *nonuniformity*. We treat this issue and the question of security of PSRGs against *uniform* adversaries in Section 4.

Razborov and Rudich [25] introduced the notion of "natural proofs," and showed that if there is a proof that is "P/poly natural against P/poly," then strong PSRGs do not exist. We prove Theorem 1 by showing that a machine M witnessing measure-zero for P/poly$\cap$EXP yields such a natural proof, in fact one of exponentially greater size than what

[*] Research supported in part by NSF grant CCR-9409104.
[†] Research supported in part by NSF grant CCR-9409104.
[‡] Research supported in part by NSF grants CCR-9057486 and CCR-9319093, and by an Alfred P. Sloan Fellowship.

suffices for their theorem. The second half of this paper takes a closer look at the nature of statistical tests, and at the specific size and strength of natural proofs in the Razborov-Rudich framework. We show a partial converse to theorem 1: if there is a $\mathcal{D}$-natural proof of sufficient density and strength against a class $\mathcal{C}$, then $\mathcal{C}$ has measure zero in nonuniform-$\mathcal{D}$.

We also prove unconditionally that nonuniform AC^0 plus parity does not have measure zero under either of the measures defined by Allender and Strauss [1]. Our results give strong reasons to investigate further both the measure theory and the natural-proofs theory, promising progress on important problems in complexity.

2 Preliminaries

The notation and conventions we use are essentially standard. All languages and functions are assumed to be defined over the finite alphabet $\Sigma = \{0,1\}$. The empty string is denoted by λ. We denote by F_n the set of all Boolean functions in n variables. A Boolean function $f_n \in F_n$ can be thought of as a binary string of length 2^n that represents the *truth table* of f_n. For readability we often write N for 2^n. We identify a language A with its characteristic sequence χ_A, and regard the latter also as a member of the set $\{0,1\}^\omega$ of infinite binary strings. For all $n \geq 0$ we also identify $A^{=n}$ with the segment u_n of χ_A of length 2^n that represents the membership or nonmembership in A of all strings of length n, and likewise identify $A^{\leq n}$ with $u_0 u_1 \cdots u_n$. Note that each u_n belongs to F_n. Then the *cylinder* $C_w = \{z \in \{0,1\}^\omega : w \sqsubseteq z\}$ contains A and all languages that agree with A on the membership of strings up to the last one indexed by w, under the standard ordering of Σ^*.

Unless specified otherwise, all Boolean circuits are over the basis $\{\wedge, \vee, \neg\}$. P/poly denotes the class of languages that have polynomial-sized circuit families. QP stands for DTIME$[2^{\text{polylog}\, n}]$, which is often called *quasipolynomial time*. QP/qpoly stands for the class of languages that have quasipolynomial-sized circuit families. This is analogous to P/poly but for quasipolynomial bounds. AC^0 denotes the class of languages that have polynomial size, constant depth circuit families, and $AC^0[\oplus]$ denotes the class of languages that have polynomial size, constant depth circuit families over the basis $\{\wedge, \vee, \neg, \oplus\}$, where $\oplus$ denotes parity. All logarithms in this paper are to the base 2.

A PSRG is formally a sequence $\{G_n\}$, where each G_n is a function from $\{0,1\}^n$ to $\{0,1\}^{\ell(n)}$, and $\ell(n) > n$. Intuitively, G_n "stretches" a sequence of n truly random bits into a longer sequence of bits that appear random to resource-bounded adversaries.

Definition 1 *Given a PSRG $G = \{G_n\}$ and a circuit C with $\ell(n)$-many input gates, say that the **bias achieved by** C is the quantity*

$$\left| \Pr_{y \in \{0,1\}^{\ell(n)}} [C(y) = 1] - \Pr_{x \in \{0,1\}^n} [C(G_n(x)) = 1] \right|.$$

*Similarly, we define the bias achieved **at length** n by a fixed probabilistic Turing machine M in place of C. The **hardness** of G **at** n, denoted by $H(G_n)$, is the largest integer $S(n)$ such that every $\ell(n)$-input circuit C of size at most $S(n)$ achieves bias at most $1/S(n)$. Then we say:*

*(a) G has **hardness at least** $h(\cdot)$ **against nonuniform adversaries** if for all but finitely many n, $H(G_n) \geq h(n)$.*

*(b) G is $h(\cdot)$**-hard against uniform adversaries** if for every $h(n)$-time bounded probabilistic TM, M, and all but finitely many n, M achieves bias at most $1/h(n)$ on G_n.*

A well-known "robustness" theorem (see [5, 9]) states that so long as $\ell(n) = n^{O(1)}$, $H(G_n)$ is invariant up to constant factors. As Razborov and Rudich do, we work with PSRGs that stretch n bits to $2n$ bits. We use "secure against" interchangeably with "hard against."

2.1 Resource-bounded measure

The resource-bounded measure theory of Lutz [13, 15] is developed along the lines of classical measure theory (see [22, 6, 23]). Languages are regarded as points in the topological space whose basic open sets are the cylinders C_w, one for each $w \in \{0,1\}^*$, and complexity classes are point sets. The general form of Lutz's theory, expounded recently by Mayordomo [20], defines conditions for a class $\mathcal{C}$ to be *measurable* by a function class Δ, and to have *measure e*, written $\mu_\Delta(\mathcal{C}) = e$, where $0 \leq e \leq 1$. Since all complexity classes we discuss are closed under finite variations, and by a form of the *Kolmogorov zero-one law* proved in [20] have measure zero or one, we need only discuss conditions for classes to have measure zero. References [13, 15, 20] show that these measurability conditions can be defined in terms of *martingales* of the kind studied earlier by Schnorr [28, 29, 30]. A *martingale* is a function d from $\{0,1\}^*$ into the nonnegative reals that satisfies the following "exact average law": for all $w \in \{0,1\}^*$,

$$d(w) = \frac{d(w0) + d(w1)}{2}. \quad (1)$$

Let $\mathbb{D}$ stand for the nonnegative dyadic rationals; i.e., those numbers of the form $n/2^r$ for integers $n, r \geq 0$.

Definition 2 (compare [13, 20]) *Let Δ be a complexity class of functions. A class $\mathcal{C}$ of languages is **Δ-measurable and has Δ-measure zero**, written $\mu_\Delta(\mathcal{C}) = 0$, if there is a martingale $d : \{0,1\}^* \to \mathbb{D}$ computable in Δ that **succeeds on** $\mathcal{C}$, in the sense that $\mathcal{C} \subseteq S^\infty[d]$ where*

$$S^\infty[d] = \{A : \lim_{w \sqsubseteq A} d(w) = +\infty\}.$$

Put another way, the *success class* $S^\infty[d]$ is the class of languages A that satisfy

$$(\forall K > 0)(\exists N > 0)(\forall w \sqsubseteq A)[|w| \geq N \Rightarrow d(w) \geq K]. \quad (2)$$

Intuitively, the martingale d is a "betting strategy" that starts with a capital sum $d(\lambda) > 0$ and makes infinite profit along the characteristic strings of every $A \in S^\infty[d]$. The purpose of the theory is to analyze the complexity required for a martingale to succeed on every language in certain subclasses $\mathcal{C}$ of a given class $\mathcal{D}$. This provides a tool for analyzing the internal structure of $\mathcal{D}$.

If $\mathcal{D}$ is defined by a collection $\mathcal{R}$ of resource bounds that is closed under squaring, then Lutz defines $\Delta(\mathcal{D})$ to be the class of martingales computable within the bound $r(\log N)$ for some function $r(N) \in \mathcal{R}$. For any class $\mathcal{C}$, Lutz writes $\mu(\mathcal{C}|\mathcal{D}) = 0$, read "$\mathcal{C}$ has measure zero within $\mathcal{D}$," if $\mu_{\Delta(\mathcal{D})}(\mathcal{C} \cap \mathcal{D}) = 0$. Two instances of particular importance are:

$$\begin{array}{ll} \mathcal{D} = \mathrm{E}, & \Delta = \mathrm{P}, \\ \mathcal{D} = \mathrm{EXP}, & \Delta = \mathrm{QP}. \end{array}$$

If $\mu(\mathcal{C}|\mathcal{D}) = 0$, then $\mathcal{C} \cap \mathcal{D}$ is intuitively "small" as a subclass of $\mathcal{D}$. The classical time-hierarchy theorems carry over to measure; in particular, P and QP have measure zero in E, and E itself, indeed DTIME$[2^{n^c}]$ for any fixed c, has measure zero in EXP. It is shown in [13, 15, 1] that classes of measure zero behave very much like null-sets in classical measure theory. The complement (in $\mathcal{D}$) of a measure-zero subclass $\mathcal{C}$ has $\Delta(\mathcal{D})$ measure 1 (this is a definition in [13, 15] and a theorem in [20]). Finite unions, and also "$\Delta(\mathcal{D})$-bounded" countable unions, of measure-zero classes have measure zero.

2.2 Natural Proofs

The technical concept at the heart of the paper by Razborov and Rudich [25] is the following. Define a *combinatorial property* to be a sequence $\Pi = [\Pi_n]_{n=0}^\infty$, where each Π_n is a subset of the set F_n of all n-variable Boolean functions. A language A is *drawn from* Π if for all n, the Boolean function given by $A^{=n}$ belongs to Π_n. The property Π *diagonalizes over* a class $\mathcal{C}$ of languages, or "is useful against" $\mathcal{C}$[1], if no language drawn from Π belongs to $\mathcal{C}$. When $\mathcal{C}$ is closed under finite variations, this is equivalent to diagonalizing *i.o.* against $\mathcal{C}$:

$$(\forall B \in \mathcal{C})(\exists^\infty n)\, B^{=n} \notin \Pi_n. \quad (3)$$

We remark that all of the natural properties constructed in [25] satisfy the stronger condition

$$(\forall B \in \mathcal{C})(\forall^\infty n)\, B^{=n} \notin \Pi_n, \quad (4)$$

which was adopted in [26]. We call this *diagonalizing a.e. against* $\mathcal{C}$.

The complexity of Π is the complexity of the decision problem: given a Boolean function $f_n \in F_n$, is $f_n \in \Pi_n$? Finally, define the *density* of Π_n by $\rho(\Pi_n) = \frac{\|\Pi_n\|}{2^N}$. The property is *large* if there exists a polynomial p such that for all but finitely many n,

$$\rho(\Pi_n) \geq \frac{1}{p(N)}. \quad (5)$$

Put another way, the Boolean functions in Π_n have *non-negligible* density in the space of all Boolean functions.

Definition 3 (cf. [25]) *Let $\mathcal{C}$ and $\mathcal{D}$ be complexity classes of languages. A combinatorial property Π is **$\mathcal{D}$-natural against** $\mathcal{C}$ if Π is large, belongs to $\mathcal{D}$, and diagonalizes over $\mathcal{C}$.*

Razborov and Rudich show that several important separation results in complexity theory use techniques that construct natural properties. Their main theorem points out limitations of such techniques. The following improvement of their theorem from polynomial to quasipolynomial size bounds for $\mathcal{D}$ was noted by Razborov [24]:

Theorem 2 *If there exists a combinatorial property that is* QP/qpoly-*natural against* P/poly*, then PSRGs of exponential hardness against nonuniform adversaries do not exist.*

(*Remarks:* In their conference version [25], Razborov and Rudich used the "i.o." definition of natural proof, which suffices for Theorem 2. All of their examples, however, satisfy the stronger "a.e." definition, and they have adopted it in the later version [26]. In Appendix 1, we sketch the additions needed for Theorem 2 that do not appear in [24, 25, 26], and give details of the proof in [25] for later reference in the proof of Theorem 17.)

By a proof analogous to that of Theorem 2, and exploiting the fact that there *is* a pseudorandom generator, based on the parity function, that is of exponential hardness against AC^0 [21], Razborov and Rudich show:

Theorem 3 *There does not exist a combinatorial property that is* qAC^0*-natural against* $\mathrm{AC}^0[\oplus]$*, where* qAC^0 *denotes the class of languages accepted by a quasipolynomial size circuit family of constant depth.*

3 Main Results

To prove our main theorem, we show that if $\mu(\mathrm{P/poly}|\mathrm{EXP}) = 0$, then one can build a natural property that diagonalizes over P/poly. Our first lemma follows by an elementary counting argument, using the fact that $\sum_{v\in\{0,1\}^\ell} d(uv) = 2^\ell \cdot d(u)$.

[1] Since $\mathcal{C}$ may be an uncountable class like P/poly, this is not necessarily a "diagonalization" in the classical sense—hence the term "useful" in [25]. But we prefer to retain it.

Lemma 4 *Let d be a martingale. For any string u and any $\ell \in \mathbb{N}$, $b \in \mathbb{R}$,*

$$\left\| \left\{ v \in \{0,1\}^{\ell} : d(uv) \leq \left(1 + \frac{1}{b}\right) d(u) \right\} \right\| \geq 2^{\ell} \left(\frac{1}{b+1} \right).$$

Our key lemma has the idea that given a martingale d that succeeds on P/poly, we can build a combinatorial property that captures those Boolean functions on $\{0,1\}^n$ along which d makes too little income to succeed. This property then diagonalizes i.o. against the success class of the martingale, which contains P/poly. Since $\Pi_n(1+1/n^2)$ converges, we can say that a return on capital of $1/n^2$, let alone losing money along a branch, is "too little income" for d. Lemma 4 will guarantee that the density of these poor branches is at least $1/n^2 = 1/(\log^2 N)$, a notably greater density than that called "large" in Equation (5).

Lemma 5 *If a* QP *martingale d succeeds on* P/poly $\cap$ EXP *then for every polynomial q, there exist infinitely many n and circuits C_i of size at most $q(i)$, for $0 \leq i < n$, such that for all circuits C_n of size at most $q(n)$,*

$$d(u_0 \ldots u_n) > \left(1 + \frac{1}{n^2}\right) d(u_0 \ldots u_{n-1}),$$

where u_i is the 2^i-bit binary "characteristic string" that indicates the membership in $L(C_i)$ of $\{0,1\}^i$.

Proof. Suppose not. Then there is a polynomial q and constant $n_0 \in \mathbb{N}$ such that for all $n \geq n_0$, for every sequence of circuits C_i of size at most $q(i)$, for $0 \leq i < n$, there exists a circuit C_n of size at most $q(n)$ such that $d(u_0 \ldots u_n) \leq \left(1 + \frac{1}{n^2}\right) d(u_0 \ldots u_{n-1})$, where the u_i's have the same meaning as in the statement of the lemma.

We will build a language L as follows: for strings of length less than n_0, membership in L will be an arbitrary but fixed sequence. Let $\alpha = d(u_0 \ldots u_{n_0 - 1})$. Clearly $\alpha < \infty$. For $n \geq n_0$, we define $L^{=n}$ inductively. Let $u_0, \ldots, u_{n-1}$ be the result of the recursively applying the construction to obtain $L^{<n}$; that is, $u_i = L^{=i}$. By assumption, there exists a circuit C_n of size at most $q(n)$ such that $d(u_0 \ldots u_n) \leq \left(1 + \frac{1}{n^2}\right) d(u_0 \ldots u_{n-1})$. Set $u_n = L(C^*)^{=n}$, where C^* is the lexicographically first C_n that satisfies this inequality (under some fixed encoding of circuits of size at most $q(n)$).

Clearly $L \in$ P/poly, since it can be accepted by the circuit family $[C_n]_{n=0}^{\infty}$. That $L \in$ EXP is immediate from the fact that finding the lexicographically first C_n takes time at most $2^{q(n)+p(n)}$, where the running time to compute the martingale d determines $p(n)$. Finally,

$$\lim_{n \to \infty} d(L^{\leq n}) \leq \alpha \prod (1 + 1/n^2) < \infty,$$

so d does not succeed on L, a contradiction. $\square$

The remaining technical problem is to weave together the constructions in Lemma 5 for all polynomial bounds q. We do not know of a *uniform* way to choose the circuits $C_0, C_1, \ldots, C_{n-1}$ promised by Lemma 5 over all q and the infinitely-many n for each q, and this is where nonuniformity enters into our results.

Lemma 6 *If $\mu(\mathrm{P/poly}|\mathrm{EXP}) = 0$, then there is a* QP/poly*-natural property against* P/poly.

Proof. For each k, let T_k be the infinite set of numbers n promised by Lemma 5 for the bound $q(n) = n^k$. Set $T := \cup_k T_k$. For all $n \in T$, take the largest number $k \leq n$ such that $n \in T_k$, take the lexicographically first $C_0, \ldots, C_{n-1}$ that works in Lemma 5, and define U_{n-1} to be the concatenation of the corresponding $u_0, \ldots, u_{n-1}$. For $n \notin T$, make some arbitrary choice such as $U_{n-1} = 0^{2^n - 1}$. Finally, for all n define

$$\Pi_n := \left\{ f_n : d(U_{n-1} f_n) \leq \left(1 + \frac{1}{n^2}\right) d(U_{n-1}) \right\}.$$

Now, by Lemma 4, the property $\Pi = \{\Pi_n\}$ is large; in fact, it has density $1/\mathrm{poly}(n)$, not just $1/\mathrm{poly}(2^n)$. By the computability of the martingale d, Π_n can be recognized in quasi-polynomial time in 2^n, given the U_{n-1}'s as advice. Equivalently, there is a family of circuits of size quasi-polynomial in 2^n that recognizes Π_n. Let L be an arbitrary language in P/poly, and let n^k be a bound on the size of a family of circuits to recognize L. Clearly, for all $n \in T_k$, $L^{=n} \notin \Pi_n$. Therefore, property Π diagonalizes i.o. over P/poly. $\square$

Theorem 7 *If $\mu(\mathrm{P/poly}|\mathrm{EXP}) = 0$, then for every family of pseudorandom generators $G = \{G_k : \{0,1\}^k \to \{0,1\}^{2k}\}$ computable in* P/poly*, for every $\varepsilon > 0$, for sufficiently large values of k, $H(G_k) \leq 2^{k^{\varepsilon}}$.*

This follows from the above three lemmas and Theorem 2. From the known equivalence of strong PSRGs and strong one-way functions (see [10, 8, 9, 25]), we also have:

Corollary 8 *If for some $\gamma > 0$ there exists a one-way function of security $2^{n^{\gamma}}$, then* P/poly *does not have measure zero in* EXP. $\square$

3.1 Non-measurability of P/poly

We strengthen the conclusion of our main result from "not measure zero" to "not measurable at all," after observing that Lutz's measures are invariant under "affine translations."

Lemma 9 *Let $\mathcal{C}$ be a proper subclass of* EXP *that is closed under symmetric difference. Then $\mathcal{C}$ does not have measure one in* EXP.

Proof. Suppose $\mathcal{C}$ does have measure one in EXP, that is, EXP $\setminus \mathcal{C}$ has measure zero in EXP. Let d be a martingale that succeeds on EXP $\setminus \mathcal{C}$.

Let $A \in \mathrm{EXP} \setminus \mathcal{C}$. Define $\mathcal{C}_A = \{L \triangle A \mid L \in \mathcal{C}\}$. We claim that if EXP $\setminus \mathcal{C}$ has measure zero, then so does $\mathcal{C}_A$. To see this, first note that $\mathcal{C}_A \subseteq \mathrm{EXP} \setminus \mathcal{C}$ since $\mathcal{C}$ and EXP are closed under symmetric difference. Now define a QP martingale d' that, on input w, outputs $d(w \oplus v)$, where v is the prefix of χ_A of length $|w|$, and where $\oplus$ denotes bitwise exclusive-or. Since $A \in \mathrm{EXP}$, it is easy for d' to compute χ_A. Finally observe that for every $L \in \mathcal{C}$, $L \triangle A$ belongs to $\mathcal{C}_A$, so d succeeds on $L \triangle A$, and thus d' succeeds on L. $\square$

Since the existence of a secure PSRG implies EXP $\not\subseteq$ P/poly, Lemma 9 implies:

Theorem 10 *If there exists a PSRG of hardness 2^{n^γ}, for some constant $\gamma > 0$, then* P/poly *is not measurable in* EXP.

Proposition 11 *Let $\mathcal{C}$ be a proper subclass of* EXP *that is closed under finite union and intersection. Then $\mathcal{C}$ does not have measure one in* EXP.

Proof. If $\mathcal{C}$ has measure one in EXP, then so does co-$\mathcal{C}$, and hence $\mathcal{C} \cap$ co-$\mathcal{C}$. If $\mathcal{C}$ is closed under finite union and intersection, so is co-$\mathcal{C}$. Therefore, $\mathcal{C} \cap$ co-$\mathcal{C}$ is closed under symmetric difference, and Lemma 9 does the rest. $\square$

Corollary 12 *Let $\mathcal{C}$ denote any of* NP, coNP, Σ_k^p, Π_k^p, P/poly, *nonuniform* NC, BPP, PP, *or* PSPACE. *Then* $\mu(\mathcal{C}|\mathrm{EXP}) = 1 \iff \mathcal{C} = \mathrm{EXP} \iff \mathcal{C} \cap co\text{-}\mathcal{C} = \mathrm{EXP}$. *In particular,* NP *has measure one in* EXP *iff* NP = EXP.

3.2 Measure of $\mathrm{AC}^0[\oplus]$

Allender and Strauss [1] have defined measures on the class $\mathcal{D} = \mathrm{P}$, imposing a restriction on the corresponding martingale class that becomes vacuous for $\mathcal{D} = \mathrm{E}$ or $\mathcal{D} = \mathrm{EXP}$, and that can be described as follows: Rather than give the Turing machines M computing martingale values $d(w)$ the string w as input, give them $N = |w|$ in binary notation on their input tape, and let them query individual bits of w. (Then M is formally the same as the machines used to define the PCP classes in [4, 3].) Measure time bounds in terms of $n = \lceil \log_2 N \rceil = |N|$ rather than N. Then the function $d(\cdot)$ belongs to $\Gamma(\mathrm{P})$ as defined in [1] if M runs in time $n^{O(1)}$, and if every node N in the directed "dependency graph," defined to have an edge (m, N) if M on input N queries bit m of *some* w, has $n^{O(1)}$ predecessors. They write $\mu(\mathcal{C}|\mathrm{P}) = 0$ if there is a $\Gamma(\mathrm{P})$ martingale that succeeds on $\mathcal{C} \cap \mathrm{P}$.

Allender and Strauss note that their measure is robust under either one of the following relaxations, but that relaxing *both* yields a different measure: allowing $d(w) \geq (d(w0) + d(w1))/2$ in place of (1), and using the "lim sup" condition of success in place of (2). We write $\mu_2(\mathcal{C}|\mathrm{P}) = 0$ to signify that $\mathcal{C}$ is one of the strictly-larger family of null classes in their second measure. They show that the class of sparse sets in P is null in the latter but not the former, and in particular that (P-uniform) AC^0 is not $\Gamma(\mathrm{P})$-measurable. But whether $\mu_2(\mathrm{AC}^0|\mathrm{P}) = 0$ is open. We show:

Theorem 13 *Nonuniform* $\mathrm{AC}^0[\oplus]$ *does not have μ_2 measure zero.*

Proof Sketch. The main idea is that owing to the dependency-set restriction in defining $\Gamma_2(\mathrm{P})$, the hypothesis $\mu_2(\mathrm{AC}^0[\oplus]) = 0$ yields a qAC^0-natural property against $\mathrm{AC}^0[\oplus]$. To handle the fact that the notion of μ_2 measure is defined using lim sup rather than the limit, we use the following stronger versions of Lemmas 4 and 5. Theorem 3 then yields a contradiction. $\square$

Lemma 14 *Let d be a martingale. For any string u and any $\ell \in \mathbb{N}$, $b \in \mathbb{R}$, the quantity*

$$\left\| \left\{ v \in \{0,1\}^\ell : (\forall w \sqsubseteq v) d(uw) \leq \left(1 + \frac{1}{b}\right) d(u) \right\} \right\|$$

is at least $2^\ell/(b+1)$.

Lemma 15 *If a* $\Gamma(\mathrm{P})$ *martingale d succeeds on* $\mathrm{AC}^0[\oplus]$, *then for every polynomial q and constant h, there exist infinitely many n and $\{\wedge, \vee, \neg, \oplus\}$-circuits C_i of size at most $q(i)$ and depth at most h, for $0 \leq i < n$, such that for all $\{\wedge, \vee, \neg, \oplus\}$-circuits C_n of size at most $q(n)$ and depth at most h,*

$$(\exists u \sqsubseteq u_n) \left[d(u_0 \ldots u) > \left(1 + \frac{1}{n^2}\right) d(u_0 \ldots u_{n-1}) \right],$$

where u_i is the 2^i-bit binary "characteristic string" that indicates the membership in $L(C_i)$ of $\{0,1\}^i$.

We have not been able to strengthen this theorem to read: $\mathrm{AC}^0[\oplus]$ does not have measure zero *in* P, that is, no $\Gamma_2(\mathrm{P})$ martingale succeeds on $\mathrm{AC}^0[\oplus] \cap \mathrm{P}$. What we have is that no $\Gamma_2(\mathrm{P})$ martingale can succeed on all of (nonuniform) $\mathrm{AC}^0[\oplus]$. Another open question concerning the measure of $\mathrm{AC}^0[\oplus]$ is whether the converse to our main theorem, obtained below in Theorem 18 carries over to this case. The obstacle is that $\mathrm{AC}^0[\oplus]$ is known to be incapable of computing "majority," which is important in converting the randomized betting strategy into a nonuniform martingale.

4 The Uniform Case and Honest Martingales

The next interesting question is whether Theorem 7 can be made to work under the hypothesis that for some $\gamma > 0$ there is a one-way function of security 2^{n^γ} *against uniform adversaries*. The main problem is that the natural property we construct in Proposition 6 is nonuniform, and this nonuniformity carries over to the statistical test constructed in the theorem of Razborov and Rudich, drawing on [7]. That is, the property belongs to QP/poly. We have not been able to obtain a QP-natural property under the hypothesis $\mu(\text{P/poly}|\text{EXP}) = 0$—the sticking point is that we have not been able to enforce any "consistency" among the characteristic prefixes $u_0, \ldots, u_{n-1}$ obtained in applications of Lemma 5 to build the Π_k that are interleaved in the proof of Lemma 6.

Interest in this problem led us to define the following "prefix-invariance" restriction on martingales, which also comes up naturally in the next section. We begin by formalizing the associated concept of a *betting strategy*.

Definition 4 *A **betting strategy** is any function $b(\cdot)$ from $\{0,1\}^*$ to the closed interval $[-1\ldots+1]$. The **martingale** d_b **derived from** b is defined by $d_b(\lambda) = 1$, and for all $w \in \{0,1\}^*$, $d_b(w1) = d_b(w)(1+b(w))$, $d_b(w0) = d_b(w)(1-b(w))$.*

For all w, let x_w stand for the string indexed by the bit c in wc, and let n_w be the length of x_w; i.e., $n_w = \lfloor \log_2(|w|+1) \rfloor$. Intuitively, $b(w)$ is the signed proportion of current capital bet on the event that x_w belongs to a given language L. A negative value of $b(w)$ indicates a bet that $x_w \notin L$. Given a martingale d, one can regard the function $b_d(w) := (d(w1) - d(w))/d(w)$ as the associated betting strategy. Henceforth we take "betting strategy" as the fundamental concept, and "martingale" as the derived one.

Definition 5 *A martingale $d : \{0,1\}^* \to \mathbb{R}$ is **honest** if it is derived from a betting strategy $b : \{0,1\}^* \to \mathbb{R}$, such that for all $w \in \{0,1\}^*$, the computation of $b(w)$ depends only on those parts of w that index strings of length n_w.*

With a few exceptions, most of the martingales implicitly constructed by Lutz et al. are honest, and this condition deserves further investigation. For honest martingales we note the following stronger form of Lemma 5:

Lemma 16 *If an honest* QP *martingale d succeeds on* $\text{P/poly} \cap \text{EXP}$ *then for every polynomial q, there exist infinitely many n such that for all circuits C_n of size at most $q(n)$, and all characteristic prefix strings $w \in \{0,1\}^{2^n-1}$, $d(wu_n) \geq \left(1+\frac{1}{n^2}\right) d(w)$, where u_n is the binary characteristic string of length 2^n that represents the strings accepted and rejected by C_n.*

Theorem 17 *If a honest* QP*-martingale succeeds on* $\text{P/poly} \cap \text{EXP}$*, then for all $\gamma > 0$, pseudorandom generators (and one-way functions) of security 2^{k^γ} against uniform adversaries do not exist.*

Proof Sketch. Given a honest QP-computable martingale d, for all n, let w_n be some characteristic prefix of length $N-1$ such that $d(w_n) > 0$. For all n, define

$$\Pi_n = \left\{ u \in \{0,1\}^N : d(w_n u) \leq \left(1+\frac{1}{n^2}\right) d(w_n) \right\}.$$

The corresponding property $\Pi = \{\Pi_n\}$ is large and belongs to QP. By Lemma 16, and with the step of fixing "$u_0, \ldots, u_{n-1}$" in the proof of Lemma 6 now rendered unnecessary, it follows that Π diagonalizes i.o. over P/poly. From the details of the proof due to Razborov and Rudich [25], which we have supplied in Appendix 1, it can be verified that the statistical test constructed from Π by Razborov and Rudich is computable by a probabilistic Turing machine in time less than 2^{k^γ}. □

Theorem 17 strengthens Theorem 7 as well as the main theorem of Razborov and Rudich: if there is a uniform P-natural or even QP-natural proof against $\text{P/poly} \cap \text{EXP}$, not against all of P/poly, then there are no PSRGs of hardness 2^{n^γ} against uniform adversaries. This leads to a sensitive and interesting point about the interplay between uniformity and nonuniformity. A QP/qpoly *martingale* is a martingale computed by circuits of quasipolynomial size; we also consider nonuniform martingales in the next section. The QP and QP/qpoly bounds, and the security bound, are tacit below:

(1) A nonuniform martingale that succeeds on P/poly yields a nonuniform natural proof against P/poly.

(2) A uniform martingale that succeeds on $\text{P/poly} \cap \text{EXP}$ also yields a nonuniform natural proof against P/poly.

(3) An honest uniform martingale that succeeds on $\text{P/poly} \cap \text{EXP}$ yields a uniform natural proof against P/poly.

(4) A uniform natural proof against $\text{P/poly} \cap \text{EXP}$ suffices to disprove the existence of PSRGs secure against uniform adversaries.

(5) A nonuniform natural proof against $\text{P/poly} \cap \text{EXP}$ does nothing, because one exists—even diagonalizing a.e. against all r.e. sets. Given an enumeration $[M_i]$ of TMs, define for all n,

$$\Pi_n = \{w \in F_n : (\forall i \leq n) w \neq L(M_i)^{=n}\},$$

Then $\Pi \in$ P/poly because for strings of length n, i.e. for w of length $N = 2^n$, we can "hard-wire" the n-many characteristic sequences of how machines $Q_1, \ldots, Q_n$ behave at length n. Also each Π_n has density $1 - n/2^N$, which is huge.

The last point indicates that much care is needed when using the natural proofs theory to talk about separations from *uniform* classes, whereas the measure theory is already tailor-made for uniformity. We ask, however, whether the theories are equivalent in the nonuniform case; i.e., whether every natural proof Π yields a ("randomized" or otherwise nonuniform) martingale that covers the class that Π diagonalizes against.

5 Are martingales and natural properties equivalent?

Say a class $\mathcal{D}$ is *nice* if it is closed under parallel evaluation of polynomially many functions in $\mathcal{D}$, under finite composition, and under the operation of finding "majority." Clearly P/poly is a nice circuit class. Recall $n = \log N$, and that density $1/2^{O(n)}$ equals "large" in [25].

Theorem 18 *Let $\mathcal{D}$ be a nice nonuniform class, and let $\mathcal{C}$ be any class of languages. Then:*

(a) *If there is a natural property $\Pi \in \mathcal{D}$ of density $1/n$ that diagonalizes a.e. against $\mathcal{C}$, then there is a martingale computable in $\mathcal{D}$ that succeeds on $\mathcal{C}$.*

(b) *If there is a natural property $\Pi \in \mathcal{D}$ of density $(1 - 1/n^{1+\varepsilon})$ that diagonalizes i.o. against $\mathcal{C}$, then there is a $\mathcal{D}$-martingale that succeeds on $\mathcal{C}$.*

(c) *If $\mathcal{D}$ is uniform, then the martingale is computed by a "randomized" $\mathcal{D}$-machine with negligible bounded error.*

Proof Sketch. Suppose we have a $\mathcal{D}$-natural property Π that diagonalizes a.e. over $\mathcal{C}$, and let $A = \{A_n\}$ denote the algorithm (family of circuits) that decides Π. For every n, consider the full binary tree T_n of depth $N = 2^n$ that has 2^N leaves in one-to-one correspondence with the members of F_n. Let $\Upsilon_n = F_n \setminus \Pi_n$, and when n is fixed or understood, let $\sigma = ||\Upsilon_n||/2^N$ denote the density of Υ_n.

For each n, the property $\Pi_n \subseteq F_n$ identifies a large subset of the leaves that are "avoided" by languages in $\mathcal{C}$. By the a.e. diagonalization condition, this means that for every $L \in \mathcal{C}$, and all but finitely many n, L goes through a branch in Υ_n at length n. This is the only property of $\mathcal{C}$ that is used in the proof; the martingale works only with the information about Π_n versus Υ_n. Given unit capital at the root of T_n, the martingale we construct will adopt the following simple strategy: try to make profit along the paths to all leaves in Υ_n, avoiding the leaves in Π_n. By the restriction on information, we allow that there may be no way for the martingale to distinguish among the leaves in Υ_n, so the best it can achieve is to amass a capital of $2^N/||\Upsilon_n|| = 1/\sigma$ at every leaf in Υ_n.

Suppose the martingale is at an interior node v of T_n. Let $V_0 = \{w \in F_n \mid w \sqsupseteq v0\}$ and $V_1 = \{w \in F_n \mid w \sqsupseteq v1\}$ denote the set of leaves in the subtrees $v0$ and $v1$, respectively. Let $p_0(v) = ||V_0 \cap \Upsilon||/||V_0||$, $p_1(v) = ||V_1 \cap \Upsilon||/||V_1||$. If the martingale could calculate $p_0(v)$ and $p_1(v)$ *exactly*, then it could set $d(v0) = 2d(v)\left(\frac{p_0}{p_0+p_1}\right)$ and $d(v1) = 2d(v)\left(\frac{p_1}{p_0+p_1}\right)$. This would ensure that each leaf in Υ ends up with a capital of $1/\sigma$ (as per the "density systems" idea of Lutz [13]).

The problem is that a martingale that runs in time poly(N) cannot compute the membership in Υ_n of all the 2^N leaves. However, by taking polynomially many random samples at each interior node, a *randomized* machine M can (with high probability) *estimate* the values $p_0(v)$ and $p_1(v)$ to a high degree of accuracy. Then M can use these estimates in lieu of the actual values, and still obey the condition (1) that defines a martingale. This strategy is continued so long as the subtree below v has more than N^2 nodes; when the subtree has atmost N^2 nodes, an exhaustive examination of all leaves is done and most of the capital is diverted towards the leaves in Υ_n, leaving a tiny portion for the leaves in Π_n. This tiny amount is donated to ensure that leaves $z \in \Pi_n$ do not go to zero, so that the martingale may eventually succeed on languages $L \in \mathcal{C}$ with $z \sqsubseteq \chi_L$. To simplify the description of M and the calculations below, we assume that if M discovers that small subtree with N^2 nodes has *no* leaves that belongs to Υ_n, it chooses some leaf arbitrarily and directs profits toward it. This "wastage" does not matter much to the profits on leaves that actually do belong to Υ_n.

Let $q_0(v)$ and $q_1(v)$ denote, respectively, the estimates of $p_0(v)$ and $p_1(v)$ that are obtained by sampling. Via standard Chernoff-bound methods, one can show that upon taking poly(N)-many samples (for a suitably large polynomial), with probability $1 - \exp(-N)$, the estimates are within an additive term of $\delta = 1/\text{poly}(N)$ of the true values. The martingale will then adopt the policy that overestimation (by upto δ) is harmless, but underestimation is dangerous. More precisely, the martingale will pretend that $q_0(v)$ and $q_1(v)$ underestimate $p_0(v)$ and $p_1(v)$, and will therefore use $q_0(v) + \delta$ and $q_1(v) + \delta$ as safer approximations to the actual values. It follows that

$$\frac{d(v0)}{d(v)} = 2\frac{q_0(v)+\delta}{(q_0(v)+\delta)+(q_1(v)+\delta)},$$

$$\frac{d(v1)}{d(v)} = 2\frac{q_1(v)+\delta}{(q_0(v)+\delta)+(q_1(v)+\delta)},$$

and that $d(v0) + d(v1) = 2d(v)$.

Let $m = \lceil 2^N/N^2 \rceil$, let $\tau_1, \tau_2, \ldots, \tau_m$ denote the subtrees of T_n at height $2\log N$ that contain N^2 leaves each. For each i, let u_i denote the root of τ_i, and let p_i denote the probability $||\texttt{leaves}(\tau_i) \cap \Upsilon||/N^2$. Let ρ_i denote the density $||\texttt{leaves}(\tau_i) \cap \Upsilon||/||\Upsilon||$; it is easy to see that $\rho_i = \frac{p_i}{p_1+p_2+\ldots+p_m}$. The total value of $d(\cdot)$ at height $2\log N$ is exactly $2^{N-2\log N} = m$, and the strategy works if for each i, $d(u_i) = \Omega(\rho_i m)$. We show:

Claim. For every i, $d(u_i) \geq 0.99\rho_i m$ whp.

For any node u, let $\pi(u)$ denote the parent of u. Wlog. let $i = 1$, and focus on the first subtree τ_1 with N^2 leaves. Recall that by the simplifying assumption made above, for all i, $p_i \geq 1/N^2$. The worst case for τ_1 is the following: at every ancestor v of τ_1, the subtree of v containing τ_1 had an underestimated probability, and the other subtree of v had an overestimated probability. To wit: at the first level, p_1 is underestimated to be $p_1 - \delta$, and p_2 is overestimated to be $p_2 + \delta$; at the second level, $\frac{1}{2}(p_1 + p_2)$ is underestimated to be $\frac{1}{2}(p_1 + p_2) - \delta$, and $\frac{1}{2}(p_3 + p_4)$ is overestimated to be $\frac{1}{2}(p_3 + p_4) + \delta$, and so on. When this happens,

$$
\begin{aligned}
d(u_1) &\geq \frac{p_1 - \delta + \delta}{(p_1 - \delta + \delta) + (p_2 + \delta + \delta)} \cdot 2d(\pi(u_1)) \\
&= 2\frac{p_1}{p_1 + p_2 + 2\delta} \cdot d(\pi(u_1))
\end{aligned}
$$

Similarly,

$$
d(\pi(u_1)) \geq 2\frac{p_1 + p_2}{p_1 + p_2 + p_3 + p_4 + 4\delta} \cdot d(\pi(\pi(u_1)))
$$

Continuing in this fashion $\log m$ times, we have

$$
d(u_1) \geq m \prod_{\ell=1}^{\log m} \frac{\sum_{i=1}^{2^{\ell-1}} p_i}{\left(\sum_{i=1}^{2^\ell} p_i\right) + 2^\ell \delta}
$$

Multiplying and dividing the above by $(p_1 + \ldots + p_m)$, and regrouping the terms,

$$
\begin{aligned}
d(u_1) &\geq m \frac{p_1}{p_1 + \ldots + p_m} \prod_{\ell=1}^{\log m} \frac{\sum_{i=1}^{2^\ell} p_i}{\left(\sum_{i=1}^{2^\ell} p_i\right) + 2^\ell \delta} \\
&= m\rho_1 \prod_{\ell=1}^{\log m} \left(1 - \frac{2^\ell \delta}{\left(\sum_{i=1}^{2^\ell} p_i\right) + 2^\ell \delta}\right)
\end{aligned}
$$

Since $p_i \geq 1/N^2 = p$ for all i,

$$
\begin{aligned}
d(u_1) &\geq m\rho_1 \prod_{\ell=1}^{\log m} 1 - \frac{2^\ell \delta}{2^\ell p + 2^\ell \delta} \\
&= m\rho_1 \left(1 - \frac{\delta}{p + \delta}\right)^{\log m} \\
&\geq m\rho_1 \left(1 - \frac{1}{N^2}\right)^N \qquad \text{setting } \delta = 1/N^4 \\
&= m\rho_1 e^{-1/N} \\
&\geq 0.99 m\rho_1 \qquad \text{for } N \geq 100. \qquad \clubsuit
\end{aligned}
$$

By standard arguments about converting high-probability algorithms into nonuniform algorithms, this can be shown to give a $\mathcal{D}$ martingale that succeeds on $\mathcal{C}$.

If the only information used by the martingale is the fact that for every L in $\mathcal{C}$, $L^{=n} \in \{0,1\}^N \setminus \Pi_n$ (i.o./a.e.), then the factor of $1/\sigma = 1/(1 - \rho(\Pi_n))$ is the best possible in stage n. If Π is a.e. diagonalizing, then a density of $\Omega(1/n) = \Omega(1/\log N)$ for Π_n gives a factor of $\Omega(1 + 1/n)$ in stage n, which suffices for the martingale to succeed on $\mathcal{C}$.

If Π is merely i.o. diagonalizing, then the above factor seems insufficient. By a modification of the Borel-Cantelli lemma as applied to martingales [13] (see also [27]), it can be shown that if $\sum_n (1 - \rho(\Pi_n))$ converges, then a successful martingale of equivalent nonuniform complexity can be constructed. For example, an i.o.-natural property Π of density $1 - \frac{1}{n^{1+\epsilon}}$ for some $\epsilon > 0$ against $\mathcal{C}$ would give a nonuniform martingale that succeeds on $\mathcal{C}$. $\square$

This partial converse brings out the importance of the actual *density* of the natural proof, and whether the diagonalization is i.o. or a.e. These are somewhat submerged in [25, 26], but we note that all six of their examples diagonalize a.e., and the first four have density at least constant or $1 - o(1)$. The natural proof involved in the striking formal independence result of Razborov [24] has density at least 1/2. Hence there are reasons to investigate the effect of different densities.

A stronger converse question is whether the *non*-existence of strong PSRGs implies that P/poly *does* have measure zero in EXP. From the non-existence it follows that given any generator of "pseudorandom" functions on $\{0,1\}^n$, a relatively small statistical test T can distinguish them from truly random Boolean functions. However, T need not have the sharp "all-or-nothing" form of the statistical test given by a natural proof, and this lack also hampers efforts to apply our proof idea of Theorem 18. In any case, there can be no simple answer, because there are oracles relative to which EXP is contained in P/poly—these give no PSRGs but also P/poly has measure *one* in EXP!

5.1 Concluding Remarks

One of the original motivations for this research was to find a sufficient condition for Lutz's hypothesis $\neg\mu(\mathrm{NP}|\mathrm{EXP}) = 0$. We briefly analyze whether Theorem 7 can be made to work with NP in place of P/poly. Our proof works by taking a hard PSRG G and a *given* QP-computable martingale d, and constructing a language $L \in \mathrm{P/poly} \cap \mathrm{EXP}$ on which d does not succeed. The languages L involved are defined by nonuniform sequences of seeds x for the "iterated generator" $f_x = G_x(y)$ defined from G in [25, 7]. These seeds define the circuits C_n in our key Lemma 5. The selection of sequences C_n in Lemma 5 is nonuniform, however. Worse yet, the definition of L uses a predicate that involves d, which is only known to be computable in exponential time.

We have shown that there is much ground for a deeper investigation into details of the natural-proofs theory of [25], in terms of the *size* of the properties and whether the diagonalization is i.o. or a.e. This may have further ramifications for the connections to formal systems shown by Razborov [24]. Finally, the idea of "randomized martingales" used to prove Theorem 18, and that of "honest" martingales that bypass the nonuniformity problem, seem to merit further study in themselves.

Acknowledgments. We are grateful to Alexander Razborov and to Steven Rudich for personal communications about extensions of their work. We thank Razborov, Rudich, Eric Allender, Jack Lutz, Elvira Mayrodomo, Alan Selman, and Martin Strauss for helpful comments on earlier drafts of this paper. Ken Regan thanks Sam Buss for communicating Krajíček's new proof of the main independence result of Razborov [24], which greatly aided his understanding of it.

References

[1] E. Allender and M. Strauss. Measure on small complexity classes, with applications for BPP. In *Proc. 35th FOCS*, pages 807–818, 1994.

[2] N. Alon, L. Babai, and A. Itai. A fast and simple randomized parallel algorithm for the maximal independent set problem. *Journal of Algorithms*, 7:567–583, 1986.

[3] S. Arora, C. Lund, R. Motwani, M. Sudan, and M. Szegedy. Proof verification and hardness of approximation problems. In *Proc. 33rd FOCS*, pages 14–23, 1992.

[4] L. Babai, L. Fortnow, L. Levin, and M. Szegedy. Checking computations in polylogarithmic time. In *Proc. 23rd STOC*, pages 21–31, 1991.

[5] R. Boppana and R. Hirschfeld. Pseudorandom generators and complexity classes. In S. Micali, editor, *Randomness and Computation*, volume 5 of *Advances in Computing Research*, pages 1–26. JAI Press, Greenwich, CT, USA, 1989.

[6] J. Doob. *Measure Theory*. Springer Verlag, New York, 1991.

[7] O. Goldreich, S. Goldwasser, and S. Micali. How to construct random functions. *J. ACM*, 33:792–807, 1986.

[8] J. Hastad. Pseudorandom generation under uniform assumptions. In *Proc. 22nd STOC*, pages 395–404, 1990.

[9] J. Hastad, R. Impagliazzo, L. Levin, and M. Luby. Construction of a pseudo-random generator from any one-way function. Technical Report 91–68, International Computer Science Institute, Berkeley, 1991.

[10] R. Impagliazzo, L. Levin, and M. Luby. Pseudorandom generation from one-way functions (extended abstract. In *Proc. 21st STOC*, pages 12–24, 1989.

[11] D. Juedes and J. Lutz. The complexity and distribution of hard problems. In *Proc. 34th FOCS*, pages 177–185, 1993. SIAM J. Comput., to appear.

[12] S. Kautz and P. Miltersen. Relative to a random oracle, NP is not small. In *Proc. 9th Structures*, pages 162–174, 1994.

[13] J. Lutz. Almost everywhere high nonuniform complexity. *J. Comp. Sys. Sci.*, 44:220–258, 1992.

[14] J. Lutz. One-way functions and balanced NP, 1992. Unpublished manuscript.

[15] J. Lutz. The quantitative structure of exponential time. In *Proc. 8th Structures*, pages 158–175, 1993.

[16] J. Lutz and E. Mayordomo. Measure, stochasticity, and the density of hard languages. In *Proc. 10th STACS*, volume 665 of *Lect. Notes in Comp. Sci.*, pages 38–47. Springer Verlag, 1993.

[17] J. Lutz and E. Mayordomo. Cook versus Karp-Levin: Separating completeness notions if NP is not small. In *Proc. 11th STACS*, volume 775 of *Lect. Notes in Comp. Sci.*, pages 415–426. Springer Verlag, 1994.

[18] J. Lutz and W. Schmidt. Circuit size relative to pseudorandom oracles. *Theor. Comp. Sci.*, 107:95–120, 1993.

[19] E. Mayordomo. Almost every set in exponential time is P-bi-immune. In *Proc. 7th MFCS*, volume nnn of *Lect. Notes in Comp. Sci.*, pages 392–400. Springer Verlag, 1992. Theor. Comp. Sci., to appear.

[20] E. Mayordomo. *Contributions to the Study of Resource-Bounded Measure*. PhD thesis, Universidad Polytécnica de Catalunya, Barcelona, April 1994.

[21] N. Nisan. Pseudorandom bits for constant-depth circuits. *Combinatorica*, 11:63–70, 1991.

[22] J. Oxtoby. *Measure and Category*. Springer Verlag, New York, 2nd edition, 1980.

[23] K.R. Parthasarathy. *Introduction to Probability and Measure*. The Macmillan Company of India, Ltd., Madras, 1977.

[24] A. Razborov. Unprovability of lower bounds on circuit size in certain fragments of bounded arithmetic. *Izvestiya of the RAN*, 1994. To appear.

[25] A. Razborov and S. Rudich. Natural proofs. In *Proc. 26th STOC*, pages 204–213, 1994.

[26] A. Razborov and S. Rudich. Natural proofs, 1994. Update of STOC paper, November 1994.

[27] K. Regan and D. Sivakumar. Improved resource-bounded Borel-Cantelli and stochasticity theorems. Technical Report UB-CS-TR 95-08, Computer Science Dept., University at Buffalo, February 1995.

[28] C.P. Schnorr. A unified approach to the definition of random sequences. *Math. Sys. Thy.*, 5:246–258, 1971.

[29] C.P. Schnorr. *Zufälligkeit und Wahrscheinlichkeit*, volume 218 of *Lect. Notes in Math.* Springer Verlag, 1971.

[30] C.P. Schnorr. Process complexity and effective random tests. *J. Comp. Sys. Sci.*, 7:376–388, 1973.

[31] M. Strauss. Normal numbers and sources for BPP, 1994. DIMACS TR 94-17, 1994; appeared in the proceedings of STACS'95, pages 515-526, 1995.

Appendix 1
Proof sketch of Theorem 2.

(This is only to bridge the gap between the result stated in [24] and the proof of the weaker result given in [25].) We first note the following, which is implicit in [25].

Lemma 19 *If a natural property* Π *(of arbitrary complexity) diagonalizes over* P/poly*, then for every polynomial* q*, there exist infinitely many* n *such that for every circuit* C_n *of size at most* $q(n)$*,* $L(C_n)^{=n}$*, treated as a* 2^n*-bit string, does not belong to* Π_n*.*

Proof. Suppose to the contrary that for some polynomial q there exists $n_0 \geq 0$ such that for all $n \geq n_0$, there exists a circuit C_n of size $q(n)$ such that $L(C_n)^{=n} \in \Pi_n$. Define a language L by letting $L^{=n} = L(C_n^*)^{=n}$ for all $n \geq n_0$, where C_n^* denotes the lexicographically first circuit of size $p(n)$ that satisfies $L(C_n^*)^{=n} \in \Pi_n$. Clearly $L \in$ P/poly, yet Π does not diagonalize over L, a contradiction. □

Now for Theorem 2, let a PSRG G and an arbitrary $\varepsilon > 0$ be given. The goal is to show that for infinitely many k, $H(G_k) \leq 2^{k^\varepsilon}$. Let the natural property Π against P/poly be such that each Π_n has density $1/2^{(\log N)^c}$ and circuit size $2^{(\log N)^c} = 2^{n^c}$. For any n, set $k = n^{c/\varepsilon}$. Using G, one can build a *pseudorandom function generator* [7] f as follows: given a seed x of size k, a (pseudorandom) Boolean function $f_x : \{0,1\}^n \to \{0,1\}$ is defined such that there is a circuit of size $\mathrm{poly}(n^{c/\varepsilon}) = \mathrm{poly}(n)$ that computes $f_x(y)$ for all $y \in \{0,1\}^n$. Using this construction, every infinite sequence of seeds $\vec{x} = x_1, x_2, \ldots$ gives a language $L_{\vec{x}}$, and all such languages have circuit families of a fixed polynomial size, say $q(n)$.

Now by Lemma 19, there are infinitely many n such that for every seed x, $f_x \notin \Pi_n$. On the other hand, by the largeness of Π, it follows that a randomly chosen $f \in \{0,1\}^{2^n}$ belongs to Π_n with probability at least $1/2^{O(n^c)}$. This shows that a circuit for Π_n is a statistical test of size $2^{O(n^c)} = 2^{O(k^\varepsilon)}$ that distinguishes f_x from a truly random Boolean function f. The remaining details are the same as in [25] drawing on [7]: Using this statistical test, one can build a statistical test of the same size that distinguishes (with bias of the same order) the output of G_k from a truly random string of length $2k$. Since ε was chosen to be arbitrary, the result follows. For the sake of completeness, we show how this conversion is done.

Claim. Suppose there is a circuit C_n of size $2^{O(n^c)}$ that achieves a bias of $2^{-O(n^c)}$ in distinguishing between f_x when x is chosen randomly from $\{0,1\}^k, k = n^{c/\varepsilon}$ and a randomly chosen 2^n-bit string. Then there is a circuit D_k of size $2^{O(n^c)} = 2^{k^\varepsilon}$ that achieves a bias of $2^{-O(n^c)} = 2^{-k^\varepsilon}$ in distinguishing between $G(x)$ when x is chosen randomly from $\{0,1\}^k$, and a randomly chosen $2k$-bit string.

Proof of Claim. Consider the full binary tree T of height n. Label the internal nodes of T by $v_1, v_2, \ldots, v_{2^n-1}$ such that if v_i is a child of v_j then $i < j$. Note that T has 2^n leaves; we will associate the leaves in one-to-one correspondence with all strings of length n. Denote by T_i the union of subtrees of T consisting of the nodes $v_1, \ldots, v_i$, together with all leaves. For a leaf y of T let $v_i(y)$ be the root of the subtree in T_i containing y. For all leaves y, define $G_{0,y}$ to be the identity function, and let $G_{i,y}$ denote the composition $G_{y_n} \circ G_{y_{n-1}} \cdots G_{y_{n-h(i,y)+1}}$. Here $h(i,y)$ denotes the height of y in T_i, or the distance between $v_i(y)$ and y. To each internal node v of the tree T, assign a string x_v chosen uniformly at random from $\{0,1\}^k$. Next, define the random collection f_i to be the collection of functions $\{f_{i,x}\}$ described as follows. Let z be a leaf of the tree. Define $f_{i,x}(z)$ to be the first bit of $G_{i,z}(x_{v_i(z)})$. Note that f_0 is just a random boolean function on n variables, and f_{2^n-1} is just f_x defined above. We know that

$$|\Pr[C_n(f_0) = 1] - \Pr[C_n(f_x) = 1]| \geq 2^{-O(n^c)}.$$

Therefore, there must exist an index i such that

$$|\Pr[C_n(f_i) = 1] - \Pr[C_n(f_{i+1}) = 1]| \geq 2^{-O(n^c)}.$$

At this point, an averaging argument shows that we can fix all the random strings assigned to the nodes of T except the children of v_{i+1} while preserving the bias. (This might determine many of the bits of f_x.) Now there are two ways of assigning strings to the children of v_{i+1}: either assign them both independently chosen random strings from $\{0,1\}^k$, or assign a random string u to v_{i+1} and assign to its two children the strings $G_0(u)$ and $G_1(u)$ respectively. The crucial observation we make is that if these two nodes are assigned strings in the first way, then the resulting boolean function induced on the leaves is precisely f_i, and if they are assigned strings in the second way, then the resulting boolean function induced on the leaves is precisely f_{i+1}. To complete the proof, we will build a circuit D_n that takes a string in $\{0,1\}^{2k}$ and computes the resulting boolean function at the leaves (which one of f_i or f_{i+1}) as described, and feeds the result (f_i or f_{i+1}) to C_n. Note that computing f_i or f_{i+1} can be done in time $2^n \cdot \mathrm{poly}(n)$. Therefore, the size of D_n is bounded by $2^{O(n^c)}$. Now, C_n has an advantage of at least $2^{-O(n^c)}$ in distinguishing between f_i and f_{i+1}, whence it follows that $H(G_k)$ is bounded by $2^{O(n^c)} = 2^{O(k^\varepsilon)}$. □