A Note on Genericity and Bi-Immunity *

José L. Balcázar

Dept. L. S. I.
Univ. Politècnica de Catalunya
08028 Barcelona, Spain
balqui@lsi.upc.es

Elvira Mayordomo

Dept. Informática
Univ. de Zaragoza
50015 Zaragoza, Spain
elvira@prometeo.cps.unizar.es

Abstract

Generic sets have all properties (from among a suitably chosen class) that can be enforced by finite extension arguments. In particular, p-generic sets are known to be P-bi-immune. We try to clarify the precise relationship between genericity and bi-immunity by proposing an extended notion of bi-immunity that exactly characterizes the p-generic sets.

1 Introduction and definitions

The notion of genericity is playing a central role in several research areas of Structural Complexity. Many open problems relating relativization did only yield to genericity as a tool. There are many specific forms of genericity. We focus here on n^c-generic sets, introduced in [1] and [2], which can be shown to exist in the class E corresponding to deterministic exponential time.

Such n^c-generic sets are, intuitively, those that enjoy all n^c-decidable properties that can be enforced by finite extension arguments. We give some additional explanation of this; in the following all sets are subsets of $\{0,1\}^*$.

Let us say that a set A has the property defined by the set C if some finite initial segment of its characteristic function belongs to C. That is, imagining that A is being constructed one bit at a time, once the property C is ever met, it remains met forever, no matter how A is eventually extended. Put it yet another way: there is a finite prefix of A such that every set coinciding with A at this prefix must have that property. In the following we identify the set of binary words C with the property it defines in this sense.

It remains to explain what do we mean by "enforcing" a property. It may be the case that some properties offer only finitely many opportunities to be met along a bitwise extension construction, and that all of them are lost if the construction pursues some other additional aims. A property that can be enforced is one that offers infinitely many opportunities to be met along the construction of the set.

Now n^c-generic sets are those that have *all* the enforceable properties that can be tested in time n^c. More formally, say that A meets C if there is a prefix z of the characteristic function of A that belongs to C; and say that C is dense along A if there are infinitely many prefixes z of the characteristic function of A such that either $z0$ or $z1$ is in C. This formalizes the condition of "offering infinitely many opportunities" to met property C. Then A is n^c-generic if and only if it meets all properties that are dense along A.

For instance, given an infinite polynomial time decidable set B, the property of A not having B as a subset is dense along any A, since a prefix $z0$ of the characteristic function of A, where the 0 corresponds to one of the infinitely many strings in B, implies that B is not a subset of A. Since this happens for all infinite polynomial time decidable sets B, it follows that all n^c-generic sets are P-immune. (See definitions below, if necessary.)

It is worth to point out here that, in that example, the set C corresponding to the dense property that has to be met is decidable in linear (even, with the appropriate computational model, sublinear) time, independently of the polynomial that bounds the computation time for B. The reason is that a prefix, say $z0$, where the 0 corresponds to a word x in B of length $|x|$, has a length $|z0|$ which is already exponential in $|x|$. The polynomial time bound on B is measured on $|x|$, while the bound n^c on the decidability of C, formed by the words $z0$, is measured on $n = |z0|$.

2 Preliminaries

All our notions and notations are standard; see [5]. As indicated, all our sets are subsets of $\{0,1\}^*$. Let

*Work supported by the EC through the Esprit BRA Program (project 7141, ALCOM II) and through the HCM Program (project CHRX-CT93-0415, COLORET Network).

$s_0, s_1, s_2, \ldots$ be the standard enumeration of $\{0,1\}^*$, which is linearly ordered by lengths and lexicographically within each length. In this way we can identify each set with its characteristic function, and this in turn with an infinite binary sequence. We denote $\mathrm{DTIME}(f)$ the complexity class of all sets decidable in time $O(f)$ on a multitape Turing machine; similar notation is used for families $\mathcal{F}$ of running time bounds. Then $\mathrm{P} = \mathrm{DTIME}(n^{O(1)})$. Oracle Turing machines are used in the standard manner to define relativized computation. Given a language A, an OTM M and a string x, $M^A(x)$ denotes the output of machine M with oracle A on input x. In the same way, given strings x, w and OTM M, $M^w(x)$ denotes the output of machine M with oracle $\{s_i \mid w[i] = 1\}$ on input x.

We will mention also the *indexing* Turing machine; it is the deterministic analog of the indexing alternating Turing machines of [7]. The only difference is that the input tape is not read sequentially by a tape head, but via direct access through a separate indexing tape. When a specific "read" state is entered, with number m written in the indexing tape, the m^{th} input symbol is transfered to some specific tape head. A special worktape symbol is reserved for the case that m is larger than the input length. A standard algorithm (unbounded binary search, or exploring consecutive powers of two followed by a standard binary search) allows for such a machine to find out the input length in logarithmic time. Note that this model defines the same class P as the standard one, but that classes such as $\mathrm{DTIME}(n^c)$ for constant c cannot be defined in this way.

We denote by $A^{<x}$ the initial segment of the characteristic function of A, up to and not including the bit corresponding to the word x.

A set A is *immune* with respect to a class $\mathcal{C}$ if it contains no infinite subset in $\mathcal{C}$. A set is *co-immune* if its complement is immune, and is *bi-immune* if it is simultaneously immune and co-immune.

Definition 1. [2] A $t(n)$-*condition* is a set $C \in \mathrm{DTIME}(t(n))$. A language A *meets the condition* C if for some string x, $A^{<x} \in C$. C is *dense along* A if

$$\exists^\infty x \in \{0,1\}^* \, \exists i \in \{0,1\} (A^{<x} i \in C);$$

and A is $t(n)$-*generic* if A meets every $t(n)$-condition which is dense along A.

Definition 2. A set A is p-generic if A is n^c-generic, for every $c > 0$.

For each c, n^c-generic sets are known to exist in linear exponential time; and p-generic sets exist in polynomial exponential time [1].

3 The main result

This section introduces the bi-immunity notion that characterizes n^c-generic sets. A technical notion needed for the definition is that of a *moderate* oracle Turing machine:

Definition 3. An oracle Turing machine is *moderate* under oracle A if it obeys the following condition: on input x, all queries to A are either smaller in length than x, or of the same length as x and lexicographically smaller than it.

That condition was used in [4] to define an extended notion of self-reducibility which could hold, in principle, for sets outside PSPACE. Notice that a machine could be moderate under some oracles but not under all of them. Notice also that a machine that is moderate under A can be simulated by a moderate machine under $\overline{A}$.

Recall that, according to the classical recursion-theoretic notion, an infinite set is immune if and only if it has no infinite recursively enumerable subset. Definitions that parallel this one, for instance using the class P instead of the class of the recursively enumerable sets, have been used in complexity theory. In particular, the notion of an infinite, co-infinite set such that neither it nor its complement has an infinite P subset, has been used in complexity theory under the name of P-*bi-immune set* [6], [9].

In order to characterize genericity, we employ the following generalizations of P-bi-immunity:

Definition 4. A set A is *c-self-immune* if it is infinite, and every subset $B \subseteq A$ accepted by a moderate machine with oracle A in time 2^{cn} is actually finite. A set A is *c-self-bi-immune* if both A and $\overline{A}$ are c-self-immune.

Definition 5. A set A is *self-bi-immune* if A is c-self-bi-immune for every $c > 0$.

As defined in [4], a set A is wdq-self-reducible if there exists a polynomial time moderate Turing machine M such that $A = L(M, A)$. Notice that a c-self-bi-immune set is not wdq-self-reducible, what is more, is very far from being so since a moderate machine can only decide A correctly for a finite number of inputs.

Since every polynomial time machine without oracle can be trivially simulated by a moderate machine in exponential time, even with a sublinear exponent, it follows immediately that every c-self-bi-immune set is P-bi-immune for all c. Our main result is now as follows:

Theorem 6. Fix $c > 1$.

- a/ If a set A is n^{c+1}-generic then it is c-self-bi-immune.

- b/ If a set A is c-self-bi-immune then it is n^c-generic.

Proof .

a/ Assume now that A is not c-self-bi-immune. We will find a set of conditions $C \in \mathrm{DTIME}(n^{c+1})$ that is dense along A but not met. Take two sets B_0 and B_1 such that both can be decided by moderate machines M_0 and M_1 under oracle A in time 2^{cn}, are subsets respectively of A and $\overline{A}$, and at least one of them is infinite; clearly this is possible if A is not c-self-bi-immune. Define C as follows:

$$C = \{w0 \mid M_0^w(s_{|w|}) = 1\} \cup \{w1 \mid M_1^w(s_{|w|}) = 0\}$$

Since the input to M_0 or M_1 has length logarithmic in $|w|$, it is possible to simulate them in time n^{c+1}, using the w part of the input to answer the oracle queries of the moderate machines. Thus indeed $C \in \mathrm{DTIME}(n^{c+1})$. The definition also ensures that C is not met by A, since C requires a 0 in the characteristic function of A precisely on all inputs from $B_0 \subseteq A$; and similarly with the ones. On the other hand, each member of $B_0 \cup B_1$ ensures a position at which either $A^{<x}0 \in C$ or $A^{<x}1 \in C$, so that this happens infinitely often. Thus C is dense along A, and hence A is not n^{c+1}-generic.

b/ Assume that A is not n^c-generic. By definition, there must be a set C in $\mathrm{DTIME}(n^c)$ that is dense along A, but such that A never meets C. Therefore, for all x, $A^{<x} \notin C$, but there are infinitely many x for which either $A^{<x}0 \in C$ or $A^{<x}1 \in C$.

Consider the two sets $B_0 = \{x \mid A^{<x}0 \in C\}$ and $B_1 = \{x \mid A^{<x}1 \in C\}$. Since A never meets C, we have $B_0 \subseteq A$ and $B_1 \subseteq \overline{A}$; and since C is dense along A, we have that $B_0 \cup B_1$ is infinite, so that at least one of them is. It only remains to see that both are decidable in linear exponential time with oracle A by a moderate machine, to infer that A is not c-self-bi-immune.

But this is an easy task: on input x of length n, just query all the previous words to construct $A^{<x}$ in time $n2^n < 2^{cn}$ append either a 0 or a 1 for each of B_0 or B_1, and compute C on the obtained string in time $(2^n)^c = 2^{cn}$. ∎

As a direct corollary, we have a characterization of p-genericity in terms of self-bi-immunity.

Corollary 7. A set is p-generic if and only if it is self-bi-immune.

Further examination of the proof reveals that the "slack" of 1 in part a/ is due to the fact that the simulating machine is provided the necessary part of the oracle as input, so that it lacks the "direct access" capability of the oracle machine. Here, the indexing machine provides a neater characterization. Let us say that a set is n^c-i-generic if is is generic with respect to conditions that can be decided in time n^c on an indexing machine. Then the characterization works level by level, since the simulation in part a/ does not need anymore the extra time to move around the input tape looking for oracle answers.

Corollary 8. (Of the proof.) Fix $c > 1$. A set A is c-self-bi-immune if and only if it is n^c-i-generic.

A stronger notion of bi-immunity was proposed in [6] in order to capture previous works on the density of complete sets. Such strong bi-immunity, called also incompressibility in [8], holds for a given set if all the functions that can act as polynomial time reductions from that set to any other set are almost 1-1. We extend that notion here to functions computed in exponential time with a moderate use of the set itself as oracle. More formally:

Definition 9. A set A is ***strongly c-self-bi-immune*** if and only if, for every function f computable in time 2^{cn} by a moderate Turing machine, either the "collision set" of pairs x, y such that $f(x) = f(y)$ is finite, or there are $x \in A$ and $y \in \overline{A}$ with $f(x) = f(y)$.

Definition 10. A set A is ***strongly self-bi-immune*** if A is strongly c-self-bi-immune for every $c > 0$.

We see now that strong self-bi-immunity also characterizes exactly p-genericity, that is, self-bi-immunity and strong self-bi-immunity coincide. First we note:

Proposition 11. If A is strongly c-self-bi-immune then A is c-self-bi-immune.

The proof in [6] that every strong P-bi-immune is P-bi-immune translates trivially to prove this fact. Next we prove that p-generic sets are strongly self-bi-immune (this can also be shown directly from the definition of genericity):

Proposition 12. A set is p-generic if and only if it is strongly self-bi-immune.

Proof . Assume that A is not strongly self-$(c-1)$-bi-immune. We will see that A is not c-self-bi-immune; by the previous results, this suffices. By assumption, there exists f a reduction from A that is computable in $\mathrm{DTIME}(2^{(c-1)n})(A^<)$ such that its "collision set" of pairs with the same image is infinite. Define:

$$B_0 = \{y | \exists x < y, f(x) = f(y), x \in A\}$$

and similarly

$$B_1 = \{y | \exists x < y, f(x) = f(y), x \in \overline{A}\}$$

Then $B_0 \in \mathrm{DTIME}(2^{(c-1)n}2^{n+1})(A^<)$, as well as B_1, and $B_0 \subseteq A$ while $B_1 \subseteq \overline{A}$. Their union is infinite, so that one of them is, and therefore A is not c-self-bi-immune. ∎

4 Discussion

Ambos-Spies, Neis and Terwijn have shown in [3] that for every c, the class of n^c-generic sets has measure 1 in E (in the resource-bounded measure sense of Lutz [10]). They also remark that this result simplifies a number of measure 1 results, that is, in order to show that a class X has measure 1 in E, it is often simpler to show that every n^c-generic set is in X. With our characterization of n^c-genericity we have thus given an intuitive limit to this technique, namely that it can be used to show that X has measure 1 in E when X can be defined in terms of (bi-)immunity and/or self-reducibility properties.

There are a number of currently assumed "axioms" or working hypothesis in complexity theory, such as $\mathrm{P} = \mathrm{NP}$ or $\mathrm{P} \neq \mathrm{NP}$, used essentially to analyze the consequences of each and to try to gather intuition regarding the proved difficulty of the many elusive complexity-theoretic problems that are by now poorly understood. Among them, as Ambos-Spies has pointed out (verbal communication), some "largeness" axioms explore the consequences of assuming that not only is NP different from P, but that it is large enough to contain progressively more "not P" sets. For instance, assuming that NP does not have measure 0 as in [11] is equivalent to saying that it contains a p-random set; other works assume the less restrictive hypothesis that it contains an n^c-generic set or a P-bi-immune set. We believe that our characterization is useful in this context, in that it shows to what extent the differences between generic and bi-immune sets are of a quantitative or of a qualitative nature.

Indeed, the only qualitative difference is the presence of the moderately used oracle set, the only remaining difference being the time bound, which is clearly a quantitative aspect. We believe that, contrarily, the step from genericity to randomness is an inherently qualitative one, and that therefore genericity is, in some sense, closer to bi-immunity than to randomness; and, to some extent, closer to bi-immunity than some intuitions suggested.

5 Acknowledgements

Thanks to Jacobo Toran and Birgit Jenner, for lending hardware to the authors, and to Maria Luisa Bonet for a car ride to Radio Shack in a moment of great need.

Both authors want to thank the organizers of the Barcelona ALCOM summer school 1994, where a preliminary version of this work was presented.

References

[1] K. Ambos-Spies, H. Fleischhack, H. Huwig: Diagonalizations over Polynomial Time Computable Sets. *Theoretical Computer Science* **51** (1987), 177–204.

[2] K. Ambos-Spies, H. Fleischhack, H. Huwig: Diagonalizing over Deterministic Polynomial Time. *Proceedings of CSL'87,* Lecture Notes in Computer Science Vol. 329 (1988), 1–16.

[3] K. Ambos-Spies, H.-C. Neis, S.A. Terwijn: Genericity and Measure for Exponential Time. *Proceedings of the 19th International Symposium on Mathematical Foundations of Computer Science,* Lecture Notes in Computer Science Vol. 841 (1994), 221–232.

[4] J.L. Balcázar: Self-Reducibility. *Journal of Computer and System Sciences* **41** (1990), 367–388.

[5] J.L. Balcázar, J. Díaz, J. Gabarró: *Structural Complexity I.* EATCS Monographs on Theoretical Computer Science, Vol. 11, Springer-Verlag 1988.

[6] J.L. Balcázar, U. Schöning: Bi-Immune Sets for Complexity Classes. *Mathematical Systems Theory* **18** (1985), 1–10.

[7] A.K. Chandra, D. Kozen, L.J. Stockmeyer: Alternation. *Journal of the ACM* **28** (1981), 114–133.

[8] D.W. Juedes, J.H. Lutz: The Complexity and Distribution of Hard Problems. *Proceedings of the 34th Symposium on Foundations of Computer Science* (1993), 177–185. *SIAM Journal on Computing,* to appear.

[9] K. Ko, D. Moore: Completeness, Approximation and Density. *SIAM Journal on Computing* **10** (1981), 787–796.

[10] J.H. Lutz: The Quantitative Structure of Exponential Time. *Proceedings 8th Annual Conference on Structure in Complexity Theory* (1993), 158–175.

[11] J.H. Lutz, E. Mayordomo: Cook Versus Karp-Levin: Separating Reducibilities if NP is not Small. *Proceedings of the 11th Annual Symposium on Theoretical Aspects of Computer Science,* Lecture Notes in Computer Science Vol. 775 (1994), 415–426. *Theoretical Computer Science,* to appear.