Topological Complexity of Sets Defined by Automata and Formulas

Extended Summary of PhD Dissertation

Szczepan Hummel
University of Warsaw

May 10, 2017

In this thesis we consider languages of infinite words or trees defined by automata of various types or formulas of various logics. We are interested in the highest possible position in the Borel or the projective hierarchy inhabited by sets defined in a given formalism.

Introduction

The theory of definability, as stated by J.W. Addison [Add04], "studies the complexity of concepts by looking at the grammatical complexity of their definitions". Measuring such a complexity is the main interest of this thesis. The concepts that we consider are properties of infinite words or infinite labeled trees, while as "grammars" we take automata models, logics or descriptive set theoretical constructions.

While infinite objects occur naturally in mathematics, it may require explanation why should they be considered on the ground of computer science. Infinite words are used e.g. in formal verification to model potentially infinite runs of systems, like an operating system or a server interacting with clients (see e.g. [JGP99]). In such a context system can be modeled as an automaton, while a specification of a system can be expressed in an appropriate logic. Infinite trees occur e.g. when we need to reason about all possible computations of a nondeterministic system (see e.g. branching time concept [Eme90]).

By considering infinite structures we gain a variety of complexity measures that come from descriptive set theory. A set of all words or all trees over a finite alphabet is, from the topological point of view, equivalent to the Cantor space—one of the two main spaces of interest of descriptive set theory (the other being the Baire space). The measures provided by this discipline include: the Borel hierarchy, the projective hierarchy, and the Wadge hierarchy. A hierarchy is here understood as an increasing, with respect to inclusion, sequence of classes of sets. Hierarchies implement the following way of measuring complexity. The higher a set is located in the hierarchy, the more complex the set is.

Hierarchies can also be used to measure complexity of formalisms. We ask how the class of all sets definable in a given formalism, e.g. a model of automata, embeds into the Borel or the projective hierarchy (see Figure 1). Such a complexity analysis usually boils down to answering one of the two following questions.

Question 1 (topological complexity) *What level of the Borel and the projective hierarchies do the sets definable in a given formalism reach?*

Or a more detailed one:

Question 2 *What levels of the Borel and the projective hierarchies are inhabited by the sets definable in a given formalism?*

The first researchers to investigate the questions in the context of formal language theory were Büchi and Landweber [BL69]. They have shown that all sets of ω-words definable in monadic second order (MSO) logic are boolean combinations of $\mathbf{\Pi}^0_2$ (or G_δ) Borel sets. The class of MSO-definable sets of ω-words, called **ω-regular languages**, gained interest several years earlier, when Büchi [Büc62] introduced a certain model of finite automata, later called Büchi automata, that were able to recognize exactly languages from this class. Since this model, as many other automata models, came with a straightforward algorithm deciding emptiness, it was used to prove decidability of MSO logic over ω-words. Similar method—through introduction of an appropriate model of automata—was used by Rabin [Rab69] to prove that MSO logic is decidable also over infinite binary labeled trees. Sets of trees definable in MSO logic are called **regular tree languages**.

Until now several models of automata equivalent to the Rabin's one were developed. In this thesis, we concentrate on one of them: parity tree automata. We also consider parity automata on ω-words. This is because those are the ones that give rise to a hierarchy, namely so called **Rabin-Mostowski index hierarchy**—another descriptive complexity measure.

The class of ω-regular languages is already well understood: as noted above, topologically, it reaches the class of boolean combinations of $\mathbf{\Pi}^0_2$ sets

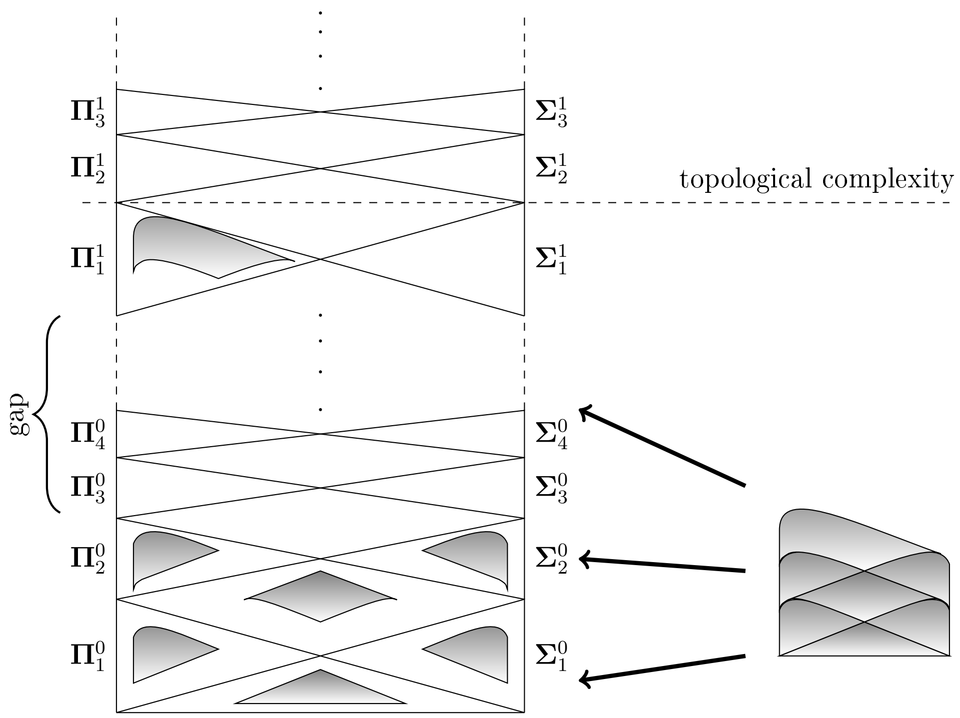

Figure 1: Embedding of the class of sets definable in a given formalism into the Borel and the projective hierarchy.

and not beyond [BL69]; the Rabin-Mostowski index hierarchy for nondeterministic automata collapses on the level of $(1, 2)$, i.e. each ω-regular language is recognized by a Büchi automaton [Büc62]; the Rabin-Mostowski index hierarchy for deterministic automata is known to be strict and for a given language one can effectively compute its position in the index hierarchy [NW98].

Robustness and tractability of the class of ω-regular languages have encouraged Bojańczyk and Colcombet [Boj04, BC06, Boj11, Boj10] to look for extension of this class of sets that would maintain some of the good properties. One of the approaches [Boj04] was to add to MSO logic a quantifier, called **unbounding quantifier**, that allows for expressing properties like "the length of blocks of consecutive letters a in a word is unbounded". The extended logic is called MSO+U. Formally, the quantifier is defined so that the formula $\mathsf{U}_X\varphi(X)$ is equivalent to writing:

"$\varphi(X)$ is satisfied by finite sets X of arbitrarily large cardinality"

Similar quantitative extensions of Büchi automata were considered [BC06]. Automata with bounding condition, called ωB-automata, are able to recognize languages like

$$L_B = \{a^{n_0}ba^{n_1}ba^{n_2}b\ldots \mid \limsup n_i < \infty\},$$

while automata with strongly unbounding condition, called ωS-automata, are able to recognize languages like

$$L_S = \{a^{n_0}ba^{n_1}ba^{n_2}b\ldots \mid \liminf n_i = \infty\}.$$

Automata combining features of the two above models are called ωBS-automata. In the thesis we calculate the exact topological complexity of ωB-automata, ωS-automata, ωBS-automata and MSO+U logic, what is described in details in the next section.

On the other hand, the class of regular languages of infinite trees is still a bit mysterious. Its topological complexity is non-Borel, namely regular tree languages reach the level $\mathbf{\Delta}^1_2$ of the projective hierarchy. The computability of nondeterministic Rabin-Mostowski index is proven only if as an input one provides a language recognized by a deterministic automaton, called a deterministic language [NW05], or a language recognized by a so-called game automaton [FMS16]. Also the shape of the embedding into the Borel and projective hierarchies (see Question 2) is better understood for deterministic than for all regular languages. Niwiński and Walukiewicz [NW03] have shown that each deterministic language is either $\mathbf{\Pi}^1_1$-complete or is at most at the level $\mathbf{\Pi}^0_3$ of the Borel hierarchy. It is clear, for the cardinality reasons, that some gap of this type exists also for the full class of regular languages, but the question what levels does the gap embrace remains open. Skurczyński [Sku93], Duparc and Murlak [DM07] have given an answer to Question 2 for another subclass of regular tree languages. Namely, they have shown that weakly recognized languages inhabit exactly all finite levels of the Borel hierarchy.

The complexity of the class of regular tree languages encourages one to look for subclasses that would be simpler to tackle. Since the best understood subclass corresponds to deterministic automata, in the thesis we consider a natural extension of this model—**unambiguous automata**. An important characteristic of deterministic automata is that they admit only one run on each input. Unambiguous automata, in a way, preserve this characteristic by having at most one accepting run on each input while syntactically preserving nondeterminism. It is known that the class of tree languages recognized by unambiguous automata is a strict extension of the class of deterministic languages and a strict subclass of the class of regular languages [NW96, CL07, CLNW10].

The class of unambiguous tree languages is not well understood yet. For instance, it is not known if the following question is decidable. Given a regular language of infinite trees, is the language recognized by some unambiguous automaton? Before results obtained by the author of this thesis, it was not even known whether topological complexity of this class is any greater than the complexity of deterministic languages, which, as noted above, are known to be co-analytic (i.e. $\mathbf{\Pi}^1_1$). It occurred to be greater [Hum12]. The thesis presents results that lift lower topological complexity bound for unambiguous languages, but no upper bound is given. In particular, it is still not known whether for each regular language there exists an unambiguous language that is topologically harder.

Results for Languages of ω-words

Languages recognized by ωBS-automata (respectively ωB-automata, ωS-automata) are called ω**BS-regular** (respectively ω**B-regular**, ω**S-regular**).

Theorem 3 *The topological complexity of the class of ωB-regular languages is $\mathbf{\Sigma}^0_3$. The topological complexity of the class of ωS-regular languages is $\mathbf{\Pi}^0_3$.*

More specifically, the first of the above statements is equivalent to saying that:

1. each ωB-regular language is in $\mathbf{\Sigma}_3^0$ (the **upper topological complexity bound**), and
2. each set from $\mathbf{\Sigma}_3^0$ is continuously reducible to a ωB-regular language (the **lower topological complexity bound**).

The second statement in Theorem 3 is interpreted analogously.

Theorem 4 *The topological complexity of the class of ωBS-regular languages is $\mathbf{\Sigma}_4^0$.*

Next we consider an alternating variant of ωBS-automata, and prove the following.

Theorem 5 *For each $n < \omega$, there is an alternating ωBS-automaton recognizing a $\mathbf{\Pi}_{2n}^0$-hard language.*

The following corollary provides a negative answer to the question stated by Bojańczyk and Colcombet [BC06, Chapter 6], whether nondeterministic ωBS-automata are equivalent to their alternating form.

Corollary 6 *Alternating ωBS-automata are more expressive than boolean combinations of nondeterministic ωBS-automata.*

Languages L_n hard for respective finite levels of the Borel hierarchy used in the proof of Theorem 5 are also proven to be definable in MSO+U. However, the results for the logic go much higher.

Theorem 7 *For every $i > 0$, there exists an* MSO+U *formula φ_i such that the language $L(\varphi_i)$ is $\mathbf{\Sigma}_i^1$-hard.*

Since it is straightforward that each language of ω-words or infinite trees definable in MSO+U is in the projective hierarchy, we get the following.

Theorem 8 *The topological complexity of* MSO+U *logic over ω-words or infinite trees is the class of* ***all*** *projective sets.*

Corollary 9 *There is no model of deterministic, nondeterministic, alternating, or nested automata with an accepting condition on a fixed level of the projective hierarchy that can capture the whole expressive power of* MSO+U *on ω-words.*

All the mentioned results on the topological complexity of ωB-automata, ωS-automata, ωBS-automata, and MSO+U were published [HMT10, HS12].

Results for Languages of Infinite Trees

A language L is **bi-unambiguous** if L is unambiguous and $\overline{L}$ is unambiguous.

Theorem 10 *There is a bi-unambiguous language of infinite trees that is $\boldsymbol{\Sigma}^1_1$-complete (analytic-complete).*

Corollary 11 *Unambiguous languages are topologically more complex than the deterministic ones.*

The language G used to prove Theorem 10 is recognized by a nondeterministic Büchi automaton.

Corollary 12 *There is a language of non-Borel topological complexity that is on one hand unambiguous, and on the other hand recognized by a nondeterministic Büchi automaton.*

The above may be considered interesting because Finkel and Simonnet [FS09, Corollary 4.14] have proven that each tree language recognized by an unambiguous Büchi automaton is Borel.

All the above results on the topological complexity of unambiguous tree languages were published [Hum12] together with a note that building on the language G one can construct an unambiguous language that is topologically harder than any countable boolean combination of analytic sets.

The results to be presented in the sequel are not published yet.

We introduce operation σ on tree languages, with the following properties.

Theorem 13 *Let $L \subseteq T_A$. If L is hard for a topological complexity class $\boldsymbol{K}$, then $\sigma(L)$ is hard for the sigma-algebra[1] $\sigma(\boldsymbol{K})$.*

Theorem 14 *If a language $L \subseteq T_A$ is bi-unambiguous then $\sigma(L)$ is bi-unambiguous.*

Definition 15 ([AN07]) *Let (X, d) be a metric space. A language $L \subseteq X$ is* ***stretchable*** *if for each sequence $\{a_n\}$ of natural numbers, there is a stretching of L with respect to $\{a_n\}$, i.e. a mapping $s : X \to X$ that reduces L to itself and satisfies for each $k \geqslant 0$ and each $t_1, t_2 \in X$:*

$$d(t_1, t_2) \leqslant 2^{-k} \implies d(s(t_1), s(t_2)) \leqslant 2^{-a_k} \qquad (1)$$

[1] Extension of the notion of the sigma-algebra generated by $\mathbf{K}$, to $\mathbf{K}$ being a class instead of a set is explained formally in the thesis.

We note that many example tree languages considered in literature are stretchable. This includes game tree languages $W_{(\iota,\kappa)}$ [AN07].

Lemma 16 *If A is a finite nonempty alphabet and a language $L \subseteq T_A$ is stretchable then $\sigma(L)$ is also stretchable.*

Below, symbol $<_W$ stands for strict Wadge quasiorder, i.e. $L <_W M$ means that a set L is continuously reducible to a set M, while M is not continuously reducible to L.

Theorem 17 *If A is a finite nonempty alphabet and $L \subseteq T_A$ is stretchable then, for each $n \geqslant 0$, $\sigma^n(L) <_W \sigma^{n+1}(L)$.*

We also construct another operation, σ^ω, that is intended to play a role of ω'th iteration of application of operation σ.

Theorem 18 *If A is a finite nonempty alphabet and $L \subseteq T_A$ is stretchable then, for each $n \geqslant 0$, $\sigma^n(L) <_W \sigma^\omega(L)$.*

Theorem 19 *If A is a finite nonempty alphabet and a language $L \subseteq T_A$ is stretchable then $\sigma^\omega(L)$ is also stretchable.*

Theorem 20 *If a language L is bi-unambiguous then the language $\sigma^\omega(L)$ is bi-unambiguous.*

Using the two operations we can define the following sequence of languages.

$$
\begin{aligned}
&\emptyset <_W \sigma(\emptyset) <_W \sigma^2(\emptyset) <_W \sigma^3(\emptyset) <_W \dots \\
<_W\, &\sigma^\omega(\emptyset) <_W \sigma(\sigma^\omega(\emptyset)) <_W \sigma^2(\sigma^\omega(\emptyset)) <_W \sigma^3(\sigma^\omega(\emptyset)) <_W \dots \\
<_W\, &\sigma^\omega(\sigma^\omega(\emptyset)) <_W \sigma(\sigma^\omega(\sigma^\omega(\emptyset))) <_W \sigma^2(\sigma^\omega(\sigma^\omega(\emptyset))) <_W \dots \\
&\vdots \\
<_W\, &(\sigma^\omega)^k(\emptyset) <_W \sigma\left((\sigma^\omega)^k(\emptyset)\right) <_W \sigma^2\left((\sigma^\omega)^k(\emptyset)\right) <_W \dots \\
<_W\, &(\sigma^\omega)^{k+1}(\emptyset) <_W \dots \\
&\vdots
\end{aligned}
\tag{2}
$$

The length of the above sequence is ω^2. All languages in the sequence are bi-unambiguous, recognized by unambiguous parity tree automata of index $\{(0,1),(1,2)\}$. It means that each strongly connected component of states in such an automaton uses only 2 priorities. On the other hand, no language in the sequence, except the first two, is recognized by an alternating automaton using 2 priorities.

For another illustration of the topological complexity of the languages in the sequence we note that already $\sigma(\emptyset)$ is $\mathbf{\Pi}^1_1$-complete.

Note that an increasing sequence of bi-unambiguous languages similar to the one shown in (2) can be built starting from any stretchable bi-unambiguous language.

Applications of the Results

The topological complexity of MSO+U was used by Bojańczyk, Gogacz, Michalewski and Skrzypczak [BGMS14] to prove that the logic is "almost undecidabile". Precisely, they have shown that no algorithm which decides the MSO+U theory of the full binary tree has a correctness proof in ZFC. The technique used in this proof became a model example of utilization of relation between topological complexity and decidability. One year later Bojańczyk, Parys and Toruńczyk [BPT16] gave undecidability proof that did not use topology. However, the topological results have given a direction towards a solution of the decidability question that was open for a decade.

While working on the topological complexity of unambiguous languages, the author have consulted Duparc to find out how operations introduced by Wadge and Duparc, and then transferred to trees by Murlak [Wad72, Wad84, Dup01, Mur06, DM07], fit the unambiguous world. The cooperation resulted in a paper [DFH15], coauthored also by Fournier, about the Wadge hierarchy of unambiguous languages. The result used language G mentioned above. The result is not included in this thesis.

Credits

The lower complexity bounds stated in Theorems 3 and 4 were proven using the example languages provided by the authors of ωB-automata and ωS-automata models, Mikołaj Bojańczyk and Thomas Colcombet [BC06]. The upper topological complexity bound for ωBS-automata stated in Theorem 4 was proven by Szymon Toruńczyk using the proof of Theorem 3 by the author of the thesis.

A sequence of MSO+U definable languages L_n hard for respective finite levels of the Borel hierarchy was given by Michał Skrzypczak. The author of this thesis have proven that the languages are recognized by an alternating variant of ωBS-automata (Theorem 5).

The results concerning projective topological complexity of MSO+U were obtained together with Skrzypczak (Theorem 7). The presentation of one of the parts of the proof was improved in cooperation with Toruńczyk.

At an early stage of the construction of the operation σ the author have obtained support from Skrzypczak, that made the operation simpler, clearer and more general.

The notion of stretchability (Definition 15) and the way how it can be used to prove that a set is not continuously reducible to itself, was borrowed from André Arnold and Damian Niwiński [AN07].

References

[Add04] J.W. Addison. Tarski's theory of definability: common themes in descriptive set theory, recursive function theory, classical pure logic, and finite-universe logic. *Annals of Pure and Applied Logic*, 126(1–3):77–92, 2004. Provinces of logic determined. Essays in the memory of Alfred Tarski. Parts I, {II} and {III}.

[AN07] André Arnold and Damian Niwiński. Continuous separation of game languages. *Fundamenta Informaticae*, 81(1-3):19–28, 2007.

[BC06] Mikołaj Bojańczyk and Thomas Colcombet. Bounds in ω-regularity. In *LICS*, pages 285–296, 2006.

[BGMS14] Mikołaj Bojańczyk, Tomasz Gogacz, Henryk Michalewski, and Michał Skrzypczak. On the decidability of MSO+U on infinite trees. In *Automata, Languages, and Programming - 41st International Colloquium, ICALP 2014, Proceedings, Part II*, pages 50–61, Copenhagen, Denmark, 2014.

[BL69] J. Richard Büchi and Lawrence H. Landweber. Definability in the monadic second-order theory of successor. *J. Symb. Log.*, 34(2):166–170, 1969.

[Boj04] Mikołaj Bojańczyk. A bounding quantifier. In *Computer Science Logic*, volume 3210 of *Lecture Notes in Computer Science*, pages 41–55, 2004.

[Boj10] Mikołaj Bojańczyk. Beyond ω-regular languages. In *STACS*, pages 11–16, 2010.

[Boj11] Mikołaj Bojańczyk. Weak MSO with the unbounding quantifier. *Theory Comput. Syst.*, 48(3):554–576, 2011.

[BPT16] Mikołaj Bojańczyk, Pawel Parys, and Szymon Toruńczyk. The MSO+U theory of (N, <) is undecidable. In *33rd Symposium on Theoretical Aspects of Computer Science, STACS 2016*, pages 21:1–21:8, Orléans, France, 2016.

[Büc62] J. Richard Büchi. On a decision method in restricted second-order arithmetic. In *Proc. 1960 Int. Congr. for Logic, Methodology and Philosophy of Science*, pages 1–11, 1962.

[CL07] Arnaud Carayol and Christof Löding. MSO on the infinite binary tree: Choice and order. In *CSL*, pages 161–176, 2007.

[CLNW10] Arnaud Carayol, Christof Löding, Damian Niwiński, and Igor Walukiewicz. Choice functions and well-orderings over the infinite binary tree. *Central European Journal of Mathematics*, 8:662–682, 2010.

[DFH15] Jacques Duparc, Kevin Fournier, and Szczepan Hummel. On unambiguous regular tree languages of index (0, 2). In *24th EACSL Annual Conference on Computer Science Logic, CSL 2015*, pages 534–548, Berlin, Germany, 2015.

[DM07] Jacques Duparc and Filip Murlak. On the topological complexity of weakly recognizable tree languages. In *Fundamentals of Computation Theory, 16th International Symposium, FCT 2007, Proceedings*, pages 261–273, Budapest, Hungary, 2007.

[Dup01] Jacques Duparc. Wadge hierarchy and Veblen hierarchy part I: Borel sets of finite rank. *J. Symb. Log.*, 66(1):56–86, 2001.

[Eme90] E. Allen Emerson. Temporal and modal logic. In *Handbook of Theoretical Computer Science, Volume B: Formal Models and Sematics (B)*, pages 995–1072. 1990.

[FMS16] Alessandro Facchini, Filip Murlak, and Michał Skrzypczak. Index problems for game automata. *ACM Trans. Comput. Log.*, 17(4):24:1–24:38, 2016.

[FS09] Olivier Finkel and Pierre Simonnet. On recognizable tree languages beyond the Borel hierarchy. *Fundamenta Informaticae*, 95(2-3):287–303, 2009.

[HMT10] Szczepan Hummel, Michał Skrzypczak, and Szymon Toruńczyk. On the topological complexity of MSO+U and related automata models. In *Mathematical Foundations of Computer Science 2010, 35th International Symposium, MFCS 2010. Proceedings*, pages 429–440, Brno, Czech Republic, 2010.

[HS12] Szczepan Hummel and Michał Skrzypczak. The topological complexity of MSO+U and related automata models. *Fundamenta Informaticae*, 119(1):87–111, 2012.

[Hum12] Szczepan Hummel. Unambiguous tree languages are topologically harder than deterministic ones. In *GandALF*, pages 247–260, 2012.

[JGP99] Edmund M. Clarke Jr., Orna Grumberg, and Doron A. Peled. *Model Checking*. MIT Press, 1999.

[Mur06] Filip Murlak. The Wadge hierarchy of deterministic tree languages. In *Automata, Languages and Programming, 33rd International Colloquium, ICALP 2006, Proceedings, Part II*, pages 408–419, Venice, Italy, 2006.

[NW96] Damian Niwiński and Igor Walukiewicz. Ambiguity problem for automata on infinite trees. unpublished note, 1996.

[NW98] Damian Niwiński and Igor Walukiewicz. Relating hierarchies of word and tree automata. In *STACS 98, 15th Annual Symposium on Theoretical Aspects of Computer Science, Proceedings*, pages 320–331, Paris, France, 1998.

[NW03] Damian Niwiński and Igor Walukiewicz. A gap property of deterministic tree languages. *Theor. Comput. Sci.*, 1(303):215–231, 2003.

[NW05] Damian Niwiński and Igor Walukiewicz. Deciding nondeterministic hierarchy of deterministic tree automata. *Electr. Notes Theor. Comput. Sci.*, 123:195–208, 2005.

[Rab69] Michael O. Rabin. Decidability of second-order theories and automata on infinite trees. *Transactions of the AMS*, 141:1–23, 1969.

[Sku93] Jerzy Skurczyński. The Borel hierarchy is infinite in the class of regular sets of trees. *Theoretical Computer Science*, 112(2):413–418, 1993.

[Wad72] William W. Wadge. Degrees of complexity of subsets of the Baire space. *Notices of the American Mathematical Society*, 19:714–715, 1972.

[Wad84] William W. Wadge. *Reducibility and determinateness on the Baire space*. PhD thesis, University of California, Berkeley, 1984.