Eliminating Cardinality Quantifiers from MLO over Trees *

Vince Bárány
Oxford University
Computing Laboratory
Oxford, United Kingdom
vbarany@comlab.ox.ac.uk

Łukasz Kaiser
RWTH Aachen University
Fachgruppe Informatik
Aachen, Germany
kaiser@logic.rwth-aachen.de

Alexander Rabinovich
Tel Aviv University
School of Computer Science
Tel Aviv, Israel
rabinoa@post.tau.ac.il

Abstract

We study the extension of monadic second-order logic of order with the uncountability quantifier "there exist uncountably many sets".

We prove that over the class of finitely branching trees this extension with the uncountability quantifier is equally expressive as plain monadic second-order logic of order.

Additionally, it follows from our proof that the continuum hypothesis holds for classes of sets definable in monadic second-order logic over finitely branching trees, which is notable since not all of these classes are analytic.

Our method to eliminate the uncountability quantifier is based on Shelah's composition method and results from descriptive set theory, and it is constructive, yielding a decision procedure for the extended logic.

1 Introduction

This paper deals with the expressive power of the extension of monadic second-order logic of order (MLO) by cardinality quantifiers "there exists infinitely many subsets X such that", "there exists uncountably many subsets X such that" and "there exists continuum many subsets X such that".

MLO extends first-order logic by allowing quantification over *subsets* of the domain. The binary relation symbol $<$ and unary predicate symbols P_i are its only non-logical relation symbols. MLO plays a very important role in mathematical logic and computer science. The fundamental connection between MLO and automata was discovered independently by Büchi, Elgot and Trakhtenbrot [3, 4, 16] when the logic was proved to be decidable over the class of finite linear orders and over $(\omega, <)$. Moving away from linear orders, Rabin proved that monadic second-order theory of the full binary tree, S2S for short, is decidable [12]. This theorem, obtained using the notion of tree automata, is one of the most celebrated results in theoretical computer science, sometimes even called "the mother of all decidability results".

First-order cardinality quantifiers, counting the number of elements with a given property, are also known under the name of Magidor-Malitz quantifiers and have been widely investigated in mathematical logic with respect to both decidability and the possibility of elimination. The book [2] presents results on decidability and other properties of first-order logic extended with such cardinality quantifiers over various natural classes of structures.

Throughout the text we will use the formalism of monadic second-order logic extended with the second-order predicate $\mathrm{Inf}(X)$ expressing that the set X is infinite. This logic, denoted $\mathrm{MLO}(\mathrm{Inf})$, is expressively equivalent to the use of the first-order infinity quantifier $\exists^{\aleph_0} x$ inside monadic second-order formulas.

Second-order cardinality quantifiers $\exists^{\kappa} X$, which we study here, allow to express that there are *at least κ different sets* X such that a formula ψ holds. For a cardinal κ and a formula $\psi(X, \overline{Y})$ we write this statement as $\exists^{\kappa} X\ \psi(X, \overline{Y})$. We will denote by $\mathrm{MLO}(\exists^{\kappa})$ the extension of MLO with the quantifier $\exists^{\kappa}$.

This paper deals with $\mathrm{MLO}(\exists^{\aleph_0}, \exists^{\aleph_1}, \exists^{2^{\aleph_0}})$ over *simple trees*. These are finitely-branching trees every branch of which is either finite or of order type ω. The main results are summarized in the next two theorems.

Theorem 1 (Elimination of the uncountability quantifier)**.** *For every* $\mathrm{MLO}(\exists^{\aleph_0}, \exists^{\aleph_1}, \exists^{2^{\aleph_0}})$ *formula* $\varphi(\overline{Y})$ *there exists an* MLO *formula* $\psi(\overline{Y})$ *that is equivalent to* $\varphi(\overline{Y})$ *over the class of simple trees. Furthermore,* ψ *is computable from* φ.

In addition to the above, the reduction will show that over simple trees the quantifiers $\exists^{\aleph_1} X$ and $\exists^{2^{\aleph_0}} X$ are equivalent, i.e. that the continuum hypothesis holds for MLO-definable families of sets. This is notable, for it is

*This research was facilitated by the ESF project AutoMathA. The first author was partially supported by ANR-06-MDCA-05 (2007-2009), DocFlow.

known that in MLO one can define non-analytic classes of sets [11].

Theorem 2. *For every* MLO *formula* $\varphi(X,\overline{Y})$, $\exists^{\aleph_1} X\, \varphi(X,\overline{Y})$ *is equivalent to* $\exists^{2^{\aleph_0}} X\, \varphi(X,\overline{Y})$ *over simple trees.*

These results naturally extend to cardinality quantifiers $\exists^{\aleph_0}\overline{X}$, $\exists^{\aleph_1}\overline{X}$ and $\exists^{2^{\aleph_0}}\overline{X}$ counting (finite) tuples of sets. This follows from the basic fact that

$$\exists^{\kappa}(U,\overline{V})\,\varphi \;\equiv\; \exists^{\kappa}U\left(\exists\overline{V}\,\varphi\right) \;\vee\; \exists^{\kappa}\overline{V}\,(\exists U\,\varphi)$$

for any cardinal $\kappa \geq \aleph_0$.

1.1 Related work

In [10] Niwiński studied questions of cardinality of sets of infinite trees recognizable by finite tree automata. He proved that it is decidable whether there are uncountably many X which satisfy an MLO formula $\varphi(X)$. The theorem of Niwiński follows from the decidability of MLO over the full binary tree and the following simple parameterless instance of our theorem: for every $\varphi \in$ MLO the sentence $\exists^{\aleph_1} X\varphi(X)$ is equivalent to an MLO sentence computable from φ.

The case with parameters was considered over $(\omega,<)$ by Kuske and Lohrey, who proved in [6, 7] that cardinality quantifiers can be eliminated from monadic second-order formulas over $(\omega,<)$. Our result is thus a generalization of their theorem to all simple trees.

1.2 Organization

We begin by showing in Section 2 how $\mathrm{MLO}(\exists^{\aleph_0})$ collapses to $\mathrm{MLO}(\mathrm{Inf})$ over all structures, and consequently to plain MLO over simple trees. Next, Section 3 fix our notation and terminology for trees and recollects some essentials of Shelah's composition method for MLO. The rest of the paper is devoted to the proof of Theorems 1 and 2.

In Section 4 we set the agenda by reducing the question of the existence of uncountably many sets X satisfying a given MLO formula $\varphi(X,\overline{Y})$ with parameters $\overline{Y}$ over a simple tree to a disjunction of three (non-exclusive) conditions: A, B and C. Condition A deals with MLO-properties of antichains; Condition C deals with a simpler version of the uncountability quantifier, namely with the quantifier "there exist uncountably many branches". Condition B expresses that there are uncountably many subsets of a branch of the tree with a special MLO property.

The conditions are treated individually in separate succeeding sections showing that each one can be formulated in plain MLO (in the case of Condition B we can only show this under the assumption that neither A nor C holds) and that in fact each condition guarantees the existence of continuum many sets X satisfying $\varphi(X,\overline{Y})$.

The most straightforward of the three, Condition A, is swiftly dealt with in Section 5.

In Section 6, we show that Condition B can be significantly weakened assuming that conditions A and C are not satisfied. Relying on the elimination results on $(\omega,<)$ from [6, 7], we formalize this weakened form of Condition B in MLO and prove, that it guarantees the existence of continuum many sets satisfying φ.

In Section 7 we consider Condition C in the special case of the complete binary tree. The key theorem that we prove there, which might be of independent interest, is that MLO-definable sets of branches of the binary tree are Borel. This opens the way to formalizing Condition C in plain MLO first over the binary tree and finally, in Section 8, over arbitrary simple trees.

We summarize the proof in Section 9 and mention further results for linear orders that can be obtained using this method in Section 10.

2 Infinity quantifier

Before we proceed to the uncountability quantifier, let us consider the the second-order infinity quantifier $\exists^{\aleph_0}X$. This quantifier can be eliminated from MLO uniformly over all structures with the aid of the "X is infinite" predicate, or equivalently, using the first-order infinity quantifier $\exists^{\aleph_0}x$.

Proposition 3. *For every* $\mathrm{MLO}(\exists^{\aleph_0})$ *formula* $\varphi(\overline{Y})$ *there exists an* $\mathrm{MLO}(\mathrm{Inf})$ *formula* $\psi(\overline{Y})$ *equivalent to* $\varphi(\overline{Y})$ *over all structures.*

Proof. Observe that the following are equivalent:

(1) There are only finitely many X which satisfy $\varphi(X,\overline{Y})$.

(2) There is a finite set Z such that any two different sets X_1, X_2 which both satisfy $\varphi(X_i,\overline{Y})$ differ on Z, i.e.

$$\exists Z\Big(\neg\mathrm{Inf}(Z) \;\wedge\; \forall X_1X_2\; \big((\varphi(X_1,\overline{Y})\wedge\varphi(X_2,\overline{Y})\wedge X_1\neq X_2) \;\rightarrow\; \exists z\in Z\,(z\in X_1 \leftrightarrow z\notin X_2)\,\big)\Big).$$

Item (2) implies (1) as a collection of sets pairwise differing only on a finite set Z has cardinality at most $2^{|Z|}$. Conversely, if $X_1,\ldots,X_k$ are all the sets that satisfy $\varphi(X_i,\overline{Y})$, then choose for every pair of distinct sets X_i, X_j an element $z_{i,j}$ in the symmetric difference of X_i and X_j and define Z as the set of these chosen elements. ☐

As the predicate $\mathrm{Inf}(X)$ is uniformly MLO-definable over all simple trees (cf. Lemma 11) we have the following corollary.

Corollary 4. $\mathrm{MLO}(\exists^{\aleph_0})$ *collapses effectively to* MLO *over the class of simple trees.*

Observe that the converse of Proposition 3 holds as well. In fact, the predicate $\mathrm{Inf}(X)$ can be defined over all structures by the formula $\exists^{\kappa} Y\,(Y \subseteq X)$ for any $\aleph_0 \leq \kappa \leq 2^{\aleph_0}$. Therefore, by Proposition 3, any of the quantifiers $\exists^{\kappa}$ with $\aleph_0 < \kappa \leq 2^{\aleph_0}$ can be used to define $\exists^{\aleph_0}$.

3 Preliminaries

For a given set A we denote by A^* the set of all finite sequences of elements of A, by A^{ω} the set of all infinite sequences of elements of A (i.e. functions $\omega \to A$), and $A^{\leq\omega} = A^* \cup A^{\omega}$. For any sequence $s = s_0 s_1 s_2 \ldots \in A^{\leq\omega}$ we denote by $|s|$ the length of s (either a natural number or ω) and by $s|_n = s_0 \ldots s_{n-1}$ the finite sequence composed of the first n elements of s, with $s|_0 = \varepsilon$, the empty sequence. We write $s[n]$ for the $(n+1)$st element of s (we start counting from 0), so $s[n] = s_n$ for $n \in \mathbb{N}$. Given a finite sequence s and a sequence $t \in A^{\leq\omega}$ we denote by $s \cdot t$ (or just st) the concatenation of s and t. Moreover, we write $s \preceq t$ if s is a prefix of t, i.e. if there exists a sequence r such that $t = sr$. A subset B of $A^{\leq\omega}$ is said to be prefix-closed if for every $t \in B$ and $s \preceq t$ it holds that $s \in B$.

3.1 Trees

For a number $l \in \mathbb{N}$, $l > 0$, an l-tree is a structure $\mathfrak{T} = (T, <, P_1, \ldots, P_l)$, where the P_i's are unary predicates and $<$ is the irreflexive and transitive binary *ancestor* relation with a least element called the *root of $\mathcal{T}$* and such that for every $v \in T$ the set $\{u \in T \mid u < v\}$ of ancestors of v is linearly ordered by $<$. Elements of a tree are referred to as *nodes*, maximal linearly ordered sets of nodes are called *branches*, ancestor-closed and linearly ordered sets of nodes are called *paths*, whereas *chains* are arbitrary linearly ordered subsets. An *antichain* is a set of pairwise incomparable nodes. Given a node v, the subtree of $\mathfrak{T}$ rooted in v is obtained by restricting the structure to the domain $T_v = \{u \in T \mid u \geq v\}$ and is denoted $\mathfrak{T}_v$.

Given a finite set A, we denote by $\mathfrak{T}(A)$ the full tree over A, which is a structure with the universe A^*, $<$ interpreted as the prefix ordering and unary predicates $P_a = A^* a$ for each $a \in A$. For finite A with $|A| = n$, this structure is axiomatizable in MLO and its MLO theory is the same as SnS, the monadic second-order theory of n successors (modulo trivial MLO-interpretations in $\mathfrak{T}(n)$).

We identify a path B of $\mathfrak{T}(A)$ with the sequence $\beta = a_0 a_1 a_2 \ldots \in A^{\leq\omega}$ such that $B = \{a_0 \ldots a_s \mid s \leq |\beta|\}$. Conversely, given a sequence $\beta \in A^{\leq\omega}$ we write $\mathrm{Pref}(\beta)$ for the corresponding path B.

Ordered sums of trees are defined as follows.

Definition 5. Let $l > 0$, $\mathfrak{I} = (I, <^{\mathcal{I}})$ be an unlabeled tree and let $\mathfrak{T}_i = (T_i, <^i, P_1^i, \ldots, P_l^i)$ be an l-tree for each $i \in I$. The *tree sum* of $(\mathfrak{T}_i)_{i \in \mathfrak{I}}$, denoted $\sum_{i \in \mathfrak{I}} \mathfrak{T}_i$, is the l-tree

$$\mathfrak{T} = (\bigcup_{i \in I} \{i\} \times T_i, <^{\mathfrak{T}}, \bigcup_{i \in I} \{i\} \times P_1{}^i, \ldots, \bigcup_{i \in I} \{i\} \times P_l{}^i),$$

such that $(i, a) <^{\mathfrak{T}} (j, b)$ for $i, j \in I$, $a \in T_i$, $b \in T_j$ iff:

$i <^{\mathfrak{I}} j$ and a is the root of $\mathfrak{T}_i$, or

$i = j$ and $a <^i b$.

Unless explicitly noted, we will not make a distinction between $\mathfrak{T}_i$ and the isomorphic subtree $\{i\} \times \mathfrak{T}_i$ of $\mathfrak{T}$.

A particular special case of the sum we will be using is when the index structure $\mathfrak{I}$ consists of a single branch, i.e. is a linear ordering. For every linear order $(I, <)$ and chain $\langle \mathfrak{T}_i \mid i \in I \rangle$ of trees, the sum $\mathfrak{T} = \sum_{i \in I} \mathfrak{T}_i$ is well defined, and $(I, <)$ forms a path (not necessarily maximal) of $\mathfrak{T}$.

We remark that not every tree can be decomposed as a sum along an arbitrarily chosen path. Such discrepancies can be ruled out by requiring that every two nodes possess a greatest common ancestor, i.e. an infimum. In this paper we work only with *simple trees*, which trivially fulfill this requirement.

Definition 6. A *simple tree* is a finitely branching tree every branch of which is either finite or of order type ω.

3.2 MLO and the composition method

We will work with labeled trees in the relational signature $\{<, P_1, \ldots, P_l\}$ where $<$ is a binary relation symbol denoting the ancestor relation of the tree, and the P_i's are unary predicates representing a labeling.

Monadic second-order logic of order, MLO for short, extends first-order logic by allowing quantification over *subsets* of the domain. MLO uses first-order variables $x, y, \ldots$ interpreted as elements, and set variables $X, Y, \ldots$ interpreted as subsets of the domain. Set variables will always be capitalized to distinguish them from first-order variables. The atomic formulas are $x < y$, $x \in P_i$ and $x \in X$, all other formulas are built from the atomic ones by applying boolean connectives and the universal and existential quantifiers for both kinds of variables. Concrete formulas will be given in this syntax, taking the usual liberties and shorthands, such as $X \cup Y, X \cap Y, X \subseteq Y$, guarded quantifiers and relativizations of formulas to a set.

The quantifier rank of a formula φ, denoted $\mathrm{qr}(\varphi)$, is the maximum depth of nesting of quantifiers in φ. For fixed n

and l we denote by $\mathrm{Form}_{n,l}$ the set of formulas of quantifier depth $\leq n$ and with free variables among $X_1, \ldots, X_l$. Let $n, l \in \mathbb{N}$ and $\mathfrak{T}_1, \mathfrak{T}_2$ be l-trees. We say that $\mathfrak{T}_1$ and $\mathfrak{T}_2$ are *n-equivalent*, denoted $\mathfrak{T}_1 \equiv^n \mathfrak{T}_2$, if for every $\varphi \in \mathrm{Form}_{n,l}$, $\mathfrak{T}_1 \models \varphi$ iff $\mathfrak{T}_2 \models \varphi$.

Clearly, $\equiv^n$ is an equivalence relation. For any $n \in \mathbb{N}$ and $l > 0$, the set $\mathrm{Form}_{n,l}$ is infinite. However, it contains only finitely many semantically distinct formulas, so there are only finitely many $\equiv^n$-classes of l-structures. In fact, we can compute representatives for these classes as follows.

Lemma 7 (Hintikka Lemma). *For $n, l \in \mathbb{N}$, we can compute a* finite *set $H_{n,l} \subseteq \mathrm{Form}_{n,l}$ such that:*

- *For every l-tree $\mathfrak{T}$ there is a* unique *$\tau \in H_{n,l}$ such that $\mathfrak{T} \models \tau$.*
- *If $\tau \in H_{n,l}$ and $\varphi \in \mathrm{Form}_{n,l}$, then either $\tau \models \varphi$ or $\tau \models \neg\varphi$. Furthermore, there is an algorithm that, given such τ and φ, decides which of these two possibilities holds.*

Elements of $H_{n,l}$ are called (n,l)-Hintikka formulas.

Given an l-tree $\mathfrak{T}$ we denote by $\mathrm{Tp}^n(\mathfrak{T})$ the unique element of $H_{n,l}$ satisfied in $\mathfrak{T}$ and call it the *n-type* of $\mathfrak{T}$. Thus, $\mathrm{Tp}^n(\mathfrak{T})$ determines (effectively) which formulas of quantifier-depth $\leq n$ are satisfied in $\mathfrak{T}$.

We sometimes speak of the n-type of a tuple of subsets $\overline{V} = V_1, \ldots, V_m$ of a given l-tree $\mathfrak{T}$. This is synonymous with the n-type of the $(l+m)$-tree $(\mathfrak{T}, \overline{V})$ obtained by expansion of $\mathfrak{T}$ with the predicates $P_{l+1}, \ldots, P_{l+m}$ interpreted as the sets $V_1, \ldots, V_m$. This type will be denoted by $\mathrm{Tp}^n(\mathfrak{T}, \overline{V})$ and often referred to as an n-type in m variables, whereby the n-type of the $(l+m)$-tree $(\mathfrak{T}, \overline{V})$ is understood. Moreover, when considering substructures, e.g. $\mathfrak{T}' \subseteq \mathfrak{T}$, and given sets $\overline{X} \subseteq \mathfrak{T}$, we write $\mathrm{Tp}^n(\mathfrak{T}', \overline{X})$ to denote $\mathrm{Tp}^n(\mathfrak{T}', \overline{X} \cap \mathfrak{T}')$.

The essence of the composition method is that certain operations on structures, such as disjoint union and certain ordered sums, can be projected to n-types. A general composition theorem for MLO from which most other follow was proved by Shelah in [13]. We only cite the composition theorem that we use [8], a more detailed presentation of the method can be found in [14, 5].

Theorem 8 (Composition Theorem for Trees). *For every* MLO*-formula $\varphi(\overline{X})$ in the signature of l-trees having m free variables and quantifier rank n, and given the enumeration $\tau_1(\overline{X}), \ldots, \tau_k(\overline{X})$ of $H_{n,l+m}$, there exists an* MLO*-formula $\theta(Q_1, \ldots, Q_k)$ such that for every tree $\mathfrak{I} = (I, <^I)$ and family $\{\mathfrak{T}_i \mid i \in I\}$ of l-trees and subsets $V_1, \ldots, V_m$ of $\sum_{i \in I} \mathfrak{T}_i$,*

$$\sum_{i \in I} \mathfrak{T}_i \models \varphi(\overline{V}) \iff \mathfrak{I} \models \theta(Q_1, \ldots, Q_k)$$

where $Q_r = Q_r^{I;\overline{V}} = \{i \in I \mid \mathrm{Tp}^n(\mathfrak{T}_i, \overline{V}) = \tau_r\}$ for each $1 \leq r \leq k$. Moreover, θ is computable from φ, and does not depend on the decomposition of $\mathfrak{T}$.

4 U-D colorings and the three conditions

To eliminate the uncountability quantifier from $\exists^{\aleph_1} X\ \varphi(X, \overline{Y})$ over an l-tree $\mathfrak{T}$, we will consider certain colorings of segments of $\mathfrak{T}$. Let us first fix m sets $\overline{Y}$, n as the quantifier rank of φ, and k as the number of n-types in $l+m+1$ variables.

An *interval* of a tree is a connected and convex set I of nodes, i.e. such that for every $u, w \in I$ if u and w are incomparable, then their greatest common ancestor is in I, and if $u < w$ then for every $u < v < w$ also $v \in I$. We denote by $\mathfrak{T}|_I$ the restriction of an l-tree $\mathfrak{T}$ to the interval I.

An interval having a minimal element is called a *tree segment*. Observe that every interval of a simple tree is a tree segment and that the summands $\mathfrak{T}_i$ of a tree sum $\mathfrak{T} = \sum_{i \in I} \mathfrak{T}_i$ are tree segments of $\mathfrak{T}$. In fact any subtree $\mathfrak{T}_z$ of a tree $\mathfrak{T}$ is a tree segment.

Let Z be a subset of a tree $\mathfrak{T}$ and z be an element of $\mathfrak{T}$. We use the notation $\mathfrak{T}_{z \setminus Z}$ for the restriction of $\mathfrak{T}$ to the set $\mathfrak{T}_z \setminus (\bigcup_{w \in Z \setminus \{z\}} \mathfrak{T}_w)$. Any tree segment $\mathfrak{T}'$ with a minimal element z can be written in the form $\mathfrak{T}_{z \setminus Z}$, where Z is the set $\{u \mid u \geq z \wedge u \notin \mathfrak{T}'\}$.

Definition 9. Let $\mathfrak{T} = (T, <, \overline{P}, X, \overline{Y})$ be an $l+m+1$-tree such that $\mathfrak{T} \models \varphi(X, \overline{Y})$ and let I be an interval of $\mathfrak{T}$.

(1) I is a *U-interval* for $\varphi, X, \overline{Y}$ iff

$$\mathfrak{T}|_I \models \forall Z\, \tau(Z, \overline{Y}) \rightarrow Z = X.$$

where $\tau(X, \overline{Y})$ is the n-type of $\mathfrak{T}|_I$ in $m+1$ variables.[1]

(2) I is a *D-interval* for $\varphi, X, \overline{Y}$ iff it is not a U-interval.

(3) In the special case of $I = \{u \mid u \geq z\}$ we say that the subtree $\mathfrak{T}_z$ is a *U-tree* or *D-tree*, respectively, and further say that z is a *U-node* or *D-node* for $\varphi, X, \overline{Y}$.

(4) The set of D-nodes for $\varphi, X, \overline{Y}$ is denoted $D(X)$.

(5) An infinite path P is called a *D-path* for $\varphi, X, \overline{Y}$ if every $v \in P$ is a D-node for $\varphi, X, \overline{Y}$, i.e. if $P \subseteq D(X)$.

Whenever $\varphi, X, \overline{Y}$ are clear from the context, we will write "D-interval for X" instead of "D-interval for $\varphi, X, \overline{Y}$", and similarly for the other notions above.

Observe that $D(X)$ is prefix-closed since if $u < v$ and $\mathfrak{T}_v$ is a D-tree then, by composition, $\mathfrak{T}_u$ is a D-tree as well. Therefore $D(X)$ can be thought of as a tree whose infinite paths are precisely the infinite D-paths for X.

[1] As set before, n is the quantifier rank of φ and m is the length of $\overline{Y}$.

We note that each of the notions introduced in Definition 9 is formalizable in MLO. Let us start by constructing the formula $\mathrm{DINT}_\varphi(I, X, \overline{Y})$, expressing that I is a D-interval for φ, X and $\overline{Y}$. By the Hintikka Lemma (L.7), the set of n-types $H_{n,l+m+1}$ can be computed and is finite. Thus, we can write the formula

$$\psi_{\mathrm{eqtp}}(X, X', \overline{Y}) = \bigwedge_{\tau \in H_{n,l+m+1}} \tau(X, \overline{Y}) \leftrightarrow \tau(X', \overline{Y}),$$

expressing that X and X' have the same n-type on the tree $\mathfrak{T}$. Let $\psi^{\mathrm{rel}}_{\mathrm{eqtp}}(X, Z, \overline{Y}, I)$ be the relativization of $\psi_{\mathrm{eqtp}}(X, Z, \overline{Y})$ to an interval I, which expresses that X and Z have the same n-type on I. $\mathrm{DINT}_\varphi(I, X, \overline{Y})$ can now be written as

$$\varphi(X, \overline{Y}) \wedge \exists Z(\psi^{\mathrm{rel}}_{\mathrm{eqtp}}(X, Z, \overline{Y}, I) \wedge X \cap I \neq Z \cap I).$$

Using this formula we can also write the formulas $\mathrm{DPATH}_\varphi(P, X, \overline{Y})$ and $\mathrm{DNODE}_\varphi(v, X, \overline{Y})$, expressing, respectively, that P is a D-path and that v is a D-node for $\varphi, X, \overline{Y}$, and the formula $\mathrm{DSET}_\varphi(D, X, \overline{Y})$ which holds iff $D = D(X)$.

The following lemma is the first step in eliminating the $\exists^{\aleph_1}$ quantifier from MLO over simple trees.

Lemma 10. *Let $\mathfrak{T}$ be a simple l-tree and $\varphi(X, \overline{Y})$ an* MLO*-formula in the signature of l-trees. Then for every tuple of subsets $\overline{V}$ of $\mathfrak{T}$*

$$\mathfrak{T} \models \exists^{\aleph_1} X\, \varphi(X, \overline{V})$$

if and only if one of the following conditions is satisfied.

- A. *There is a set U satisfying $\mathfrak{T} \models \varphi(U, \overline{V})$ and there is an infinite antichain A of D-nodes for $\varphi, U, \overline{V}$.*
- B. *There is an infinite branch B which is a D-path for uncountably many U satisfying $\mathfrak{T} \models \varphi(U, \overline{V})$.*
- C. *The set of branches*

  $$\{B \mid \textit{ exists a } U \textit{ such that } B \textit{ is a D-path for } \varphi, U, \overline{V}\}$$

  is uncountable.

Proof. Condition B explicitly requires the existence of uncountably many sets satisfying $\varphi(X, \overline{V})$, so it is clearly sufficient for $\exists^{\aleph_1} X\, \varphi(X, \overline{V})$ to hold. We first show that Condition A is in itself sufficient, and then that if Condition A does not hold, then Condition C is sufficient as well. Last we prove that the disjunction of the three is also necessary.

Sufficiency of Condition A.
To see that Condition A is sufficient, let U and A be the sets guaranteed to exist in Condition A, and let v_0 denote the root of $\mathcal{T} = (\mathfrak{T}, U, \overline{V})$. Then $\mathcal{T}$ can be decomposed as

$$\mathcal{T} = \mathcal{T}_{v_0 \setminus A} + \sum_{w \in A} \mathcal{T}_w.$$

Applying the Composition Theorem (Th.8) to this decomposition, we get that $\mathfrak{T} \models \varphi(U', \overline{V})$ for every U' such that $U' \cap \mathfrak{T}_{v_0 \setminus A} = U \cap \mathfrak{T}_{v_0 \setminus A}$ and $\mathrm{Tp}^n(\mathfrak{T}_w, U', \overline{V}) = \mathrm{Tp}^n(\mathfrak{T}_w, U, \overline{V})$ for all $w \in A$. By the choice of A, U can be modified independently on each subtree $\mathcal{T}_w$ without changing its type $\mathrm{Tp}^n(\mathcal{T}_w)$. Hence there are continuum many different sets U' as above.

Sufficiency of Condition C *when Condition* A *fails.*
If Condition A does not hold, then for each U satisfying $\varphi(U, \overline{V})$, the set $D(U)$ does not contain an infinite antichain. Thus, since $D(U)$ is a simple tree and König's Lemma applies, it is comprised of only finitely many branches. In particular, there are only finitely many infinite D-paths for each such U. Thus, if Condition C holds and there are uncountably many D-paths altogether, then there are uncountably many sets U satisfying $\varphi(U, \overline{V})$ as well.

Necessity of the three conditions.
As already observed above, if Condition A fails, then for each U satisfying $\varphi(U, \overline{V})$, $D(U)$ is a tree comprised of only finitely many branches. In particular, there are only finitely many infinite D-paths for each such U. In case Condition C fails too, there are at most countably many sets $D(U)$ altogether for all the sets U satisfying $\varphi(U, \overline{V})$.

It remains to show that if Condition B is not fulfilled either, then for every set D there can be at most countably many sets U satisfying $\varphi(U, \overline{V})$ and having $D(U) = D$.

For all those D containing an infinite path, this is explicitly guaranteed by the failure of Condition B. Let us consider the other case, so let D be a finite prefix-closed set and F be the set of maximal points in D, i.e. its frontier nodes. If $D = D(U)$, then U is fully determined by $U \cap D$ and the n-types of all successor nodes of the frontier nodes. Over finitely branching trees this only allows for a finite number of choices of U. The simultaneous failure of all three conditions therefore implies that $\exists^{\leq \aleph_0} X\, \varphi(X, \overline{V})$. □

5 Formalization of Condition A

The definition of Condition A can be directly cast in $\mathrm{MLO}(\mathrm{Inf})$. It suffices therefore to note that the predicate $\mathrm{Inf}(X)$ can be formalized in pure MLO over simple trees, as proved in Appendix A.

Lemma 11. *There exists an* MLO *formula $\psi_{\mathrm{Inf}}(X)$ that holds on a simple tree $\mathfrak{T}$ if and only if X is infinite.*

Condition A can now be formalized in MLO as

$$\psi_{\mathsf{A}}(\overline{Y}) = \exists U \exists A\big(\varphi(U, \overline{Y}) \wedge \psi_{\mathrm{Inf}}(A) \wedge \mathsf{antich}(A) \wedge (\forall w \in A\, \mathrm{DNODE}_\varphi(w, U, \overline{Y}))\big),$$

where $\mathsf{antich}(A) = \forall x, y \in A\, \neg(x < y \vee y < x)$.

Already in the proof of Lemma 10 we have pointed out that if condition A is satisfied, then there are continuum many sets X satisfying the formula $\varphi(X, \overline{Y})$.

6 Condition B

In this section, we show that a branch B is a witness for Condition B if and only if this branch satisfies a disjunction of three sub-conditions: Ba, Bb and Bc. Moreover, if both Condition A and Condition C fail, then already the sub-conditions Ba and Bc are sufficient. Finally, we express both Ba and Bc in MLO and show, that in fact both these sub-conditions guarantee the existence of continuum many sets X satisfying the formula $\varphi(X, \overline{Y})$ in consideration.

As in the previous section, we assume that the formula $\varphi(X, \overline{Y})$ of quantifier rank n is fixed together with a simple l-tree $\mathfrak{T}$ and m parameters $\overline{Y}$, and let k be the number of n-types in $l+m+1$ variables. Additionally, we fix a branch B and introduce the formula $\psi(X, \overline{Y}, P)$ stating that P is an infinite D-path for X and that $\varphi(X, \overline{Y})$ holds:

$$\psi(X, \overline{Y}, P) = \mathrm{DPATH}_{\varphi}(P, X, \overline{Y}) \wedge \mathrm{Inf}(P) \wedge \varphi(X, \overline{Y}).$$

Note that the branch B witnesses Condition B if and only if $\exists^{\aleph_1} U\, \psi(U, \overline{Y}, B)$.

To break up Condition B, we decompose $\mathcal{T} = (\mathfrak{T}, X, \overline{Y})$ along the branch B, $\mathcal{T} = \sum_{w \in B} \mathcal{T}_{w \setminus B}$, and apply the Composition Theorem (Th.8) to this decomposition and the formula ψ. This yields a formula θ such that

$$\mathcal{T} \models \psi(X, \overline{Y}, B) \iff (B, <) \models \theta(P_1, \ldots, P_r),$$

where r is the number of $\mathrm{qr}(\psi)$-types in $l+m+2$ variables, which we enumerate as $\tau_1, \ldots, \tau_r$, and

$$P_i = \{w \in B \mid (\mathcal{T}_{w \setminus B}, \{w\}) \models \tau_i\}.$$

Note that we use the expansion of $\mathcal{T}_{w \setminus B}$ by $\{w\}$ as w is the only element of $\mathcal{T}_{w \setminus B}$ that belongs to B. The above application of the Composition Theorem allows us to formulate and prove the following lemma.

Lemma 12. *There are uncountably many $X \subseteq \mathfrak{T}$ satisfying the formula $\psi(X, \overline{Y}, B)$ in $\mathfrak{T}$ iff one of the following sub-conditions holds.*

(Ba) *There exists a set X such that $\mathfrak{T}_{w \setminus B}$ is a D-interval for $\varphi, X, \overline{Y}$ for infinitely many $w \in B$.*

(Bb) *There exists a set X satisfying ψ and a $w \in B$ so that*

$$\mathfrak{T}_{w \setminus B} \models \exists^{\aleph_1} X'\, \tau_i(X', \overline{Y} \cap \mathfrak{T}_{w \setminus B}, \{w\}),$$

where $\tau_i = \mathrm{Tp}^{\mathrm{qr}(\psi)}(\mathfrak{T}_{w \setminus B}, X, \overline{Y}, \{w\})$.

(Bc) *It holds that*

$$(B, <) \models \exists^{\aleph_1} \overline{P} \left(\theta(\overline{P}) \wedge \bigwedge_{i=1}^{r} P_i \subseteq Q_i \wedge \forall x \left(\bigvee_{i=1}^{r} (x \in P_i \wedge \bigwedge_{j \neq i} x \notin P_j) \right) \right),$$

where Q_i is the set of nodes on the branch B in which the type τ_i is satisfied by some set X, i.e.

$$Q_i = \{w \in B \mid \mathfrak{T}_{w \setminus B} \models \exists X\, \tau_i(X, \overline{Y} \cap \mathfrak{T}_{w \setminus B}, \{w\})\}$$

for each $1 \leq i \leq r$.

Proof. By the application of the Composition Theorem done above, $\mathcal{T} \models \psi(X, \overline{Y}, B)$ iff $(B, <) \models \theta(P_1, \ldots, P_r)$. Let us consider the following cases.

Case 1: There exists a tuple $\overline{P}$ such that $(B, <) \models \theta(\overline{P})$ and there are uncountably many sets X for which $P_i = \{w \in B \mid (\mathfrak{T}_{w \setminus B}, X, \overline{Y}, \{w\}) \models \tau_i\}$ for each $1 \leq i \leq r$.
In this case the branch B witnesses Condition B, so we only need to show that one of the sub-conditions holds. By contradiction, assume that sub-condition (Ba) does not hold. Then, for every set X satisfying $\psi(X, \overline{Y}, B)$, the segment $\mathfrak{T}_{w \setminus B}$ is a D-interval only for finitely many $w \in B$. Consider one of the uncountably many sets X which have $\mathrm{qr}(\psi)$-types on $\mathfrak{T}_{w \setminus B}$ described by $\overline{P}$. Since $\mathrm{qr}(\psi) \geq \mathrm{qr}(\varphi)$ and $\mathfrak{T}_{w \setminus B}$ is a U-interval for X for all but finitely many w's, all of the continuum many sets that share $\overline{P}$ must be equal to X on all but finitely many $\mathfrak{T}_{w \setminus B}$. Thus, there is as well a single w for which there are continuum many different sets sharing the types with X on $\mathfrak{T}_{w \setminus B}$, and thus Condition (Bb) is satisfied.

Case 2: For each tuple $\overline{P}$ such that $(B, <) \models \theta(\overline{P})$ there are only countably many sets X for which $P_i = \{w \in B \mid (\mathfrak{T}_{w \setminus B}, X, \overline{Y}, \{w\}) \models \tau_i\}$.
In this case, we show that Condition (Bc) is both necessary and sufficient for the existence of uncountably many sets X satisfying ψ.

Necessity of Condition (Bc).
As a direct consequence of the application of the Composition Theorem above and the condition of this case, if there are uncountably many sets X satisfying ψ then there are uncountably many corresponding tuples $\overline{P}$ for which $(B, <) \models \theta(\overline{P})$. By definition, P_i is the set of w's for which $(\mathfrak{T}_{w \setminus B}, X, \overline{Y}, \{w\}) \models \tau_i$. Taking the X above we get $\mathfrak{T}_{w \setminus B} \models \exists X \tau_i(X, \overline{Y} \cap \mathfrak{T}_{w \setminus B}, \{w\})$, and therefore $P_i \subseteq Q_i$ holds. Since Hintikka formulas are mutually exclusive, each two sets P_i, P_j for $i \neq j$ are disjoint. This guarantees that the remaining conjunct $\forall x \big(\bigvee_{i=1}^{r} (x \in P_i \wedge \bigwedge_{s \neq r} x \notin P_s \big)$ of Condition (Bc) is satisfied, and thus Condition (Bc) holds.

Sufficiency of Condition (Bc).
By definition of the sets Q_i, for each $w \in Q_i$ there is a set $X_{w,i}$ which makes the type τ_i satisfied on the extension of $\mathfrak{T}_{w \setminus B}$. Assuming that Condition (Bc) holds, let $\mathcal{P}$ be the uncountable set of tuples $\overline{P}$ that witness this condition. For each such tuple $\overline{P}$ and each $w \in B$ the last conjunct of Condition (Bc) guarantees that there is a unique $i = i(w)$ for which $w \in P_i$. Construct $X_{\overline{P}}$ as the sum of $X_{w,i(w)}$ over all $w \in B$. Since $P_i \subseteq Q_i$, the tuple $\overline{P}$

indeed describes the types of the set $X_{\overline{P}}$. Therefore for different tuples $\overline{P}_1, \overline{P}_2$ the sets $X_{\overline{P_1}}, X_{\overline{P_2}}$ are different as well. Moreover, since $\theta(\overline{P})$ holds, the above application of the Composition Theorem guarantees that $\psi(X_{\overline{P}}, \overline{Y}, B)$ holds. Thus $\{X_{\overline{P}} \mid \overline{P} \in \mathcal{P}\}$ constitutes an uncountable family of sets satisfying ψ. $\square$

While Condition (Bb) in itself is just another instance of the problem we started with, we claim that when conditions A and C fail, it can simply be ignored.

Lemma 13. *If over a finitely branching tree $\mathfrak{T}$ both Condition* A *and Condition* C *fail, then Condition* B *holds if and only if there exists a branch that satisfies Condition* (Ba) *or Condition* (Bc).

Proof. If conditions A and C fail, then, as we have already seen, the set $\mathcal{D} = \{D(X) \mid \mathfrak{T} \models \varphi(X, \overline{Y})\}$ is countable. Moreover, each $D \in \mathcal{D}$ is a union of finitely many paths.

If Condition B holds then there are uncountably many sets X satisfying $\varphi(X, \overline{Y})$ and thus, as $\mathcal{D}$ is countable, there is a set D such that $D = D(X)$ for uncountably many X satisfying φ. Fix such a set D and consider all its labelings by the types of X on the partial trees $\mathfrak{T}_{w\setminus D}$, i.e. the set $\mathcal{L} = \{\overline{L}^X \mid D(X) = D\}$ where $\overline{L}^X = \langle L_1^X \dots L_k^X \rangle$ and

$$L_j^X = \{w \in D \mid \mathrm{Tp}^n(\mathfrak{T}_{w\setminus D}, X, \overline{Y}, \{w\}) = \tau_j\}.$$

We are going to show that the failure of Condition (Bc) guarantees that the set $\mathcal{L}$ is countable.

First, D is the union of a finite set of branches, therefore there is a finite set $E = \{e_1, \dots, e_s\}$ of maximal branching points of D. For $i = 1 \dots s$, let $\mathrm{Path}_i = \{v \in D \mid v > e_i\}$, let B_i be the unique branch of D that contains Path_i and let $T_{\text{fin}} = D \setminus \cup_i \mathrm{Path}_i$. Note that T_{fin} is a finite subtree of D and hence there are only finitely many possible labelings of T_{fin}. Note also that B_i are infinite branches.

If $\mathcal{L}$ was uncountable then there would exist an i with uncountably many different labelings of Path_i, i.e. the set $\mathcal{H} = \{\overline{H}^X \mid D(X) = D\}$ where $\overline{H}^X = \langle H_1^X \dots H_k^X \rangle$,

$$H_j^X = \{w \in Path_i \mid \mathrm{Tp}^n(\mathfrak{T}_{w\setminus D}, X, \overline{Y}, \{w\}) = \tau_j\},$$

would be uncountable. However, for $w \in \mathrm{Path}_i$, $\mathfrak{T}_{w\setminus D} = \mathfrak{T}_{w\setminus \mathrm{Path}_i} = \mathfrak{T}_{w\setminus B_i}$. Therefore, $\mathcal{Q} = \{\overline{Q}^X \mid D(X) = D\}$ where $\overline{Q}^X = \langle Q_1^X, \dots Q_k^X \rangle$ and

$$Q_j^X = \{w \in B_i \mid \mathrm{Tp}^n(\mathfrak{T}_{w\setminus B_i}, X, \overline{Y}, \{w\}) = \tau_j\}$$

would be uncountable. Since $\mathrm{qr}(\psi) \geq n$, different n-types induce different $\mathrm{qr}(\psi)$-types, so the set $\mathcal{P} = \{\overline{P}^X \mid D(X) = D\}$, with $\overline{P}^X = \langle P_1^X, \dots P_r^X \rangle$ and

$$P_j^X = \{w \in B_i \mid \mathrm{Tp}^{\mathrm{qr}(\psi)}(\mathfrak{T}_{w\setminus B_i}, X, \overline{Y}, \{w\}) = \tau_j\},$$

is uncountable as well. (Note that here τ_j is an $\mathrm{qr}(\psi)$-type.) As shown in the part on necessity of Condition (Bc) in the proof of Lemma 12, each such $\overline{P}^X$ satisfies the formula in Condition (Bc), so this condition holds for B_i.

As shown above, $\mathcal{L}$ is countable. Since there are uncountably many X with $D(X) = D$, there exists a single type labeling $\overline{L}$ such that $\overline{L} = \overline{L}^X$ for uncountably many of these sets X. Thus each of these uncountably many sets X has the same type $\mathrm{Tp}^n(\mathfrak{T}_{w\setminus D}, X, \overline{Y}, \{w\})$ for each $w \in B$, which we denote $\tau_{(w)}$.

If Condition (Ba) is not satisfied either, all but finitely many of these $\tau_{(w)}$ uniquely define X on the respective tree segments $\mathfrak{T}_{w\setminus D}$.

Thus, there exists a $w \in D$ such that there are uncountably many X as above pairwise differing on the tree segment $\mathfrak{T}_{w\setminus D}$. However, by definition, every subtree of $\mathfrak{T}_{w\setminus D}$ is a U-tree relative to every of these X, because $D(X) = D$. Hence if $\mathfrak{T}$ is finitely branching, i.e. if $\mathfrak{T}_{w\setminus D} \setminus \{w\}$ is a finite union of such U-trees, then there can be only finitely many X as above pairwise differing on $\mathfrak{T}_{w\setminus D}$, which is a contradiction. $\square$

In the next subsections we construct MLO formulas $\psi_{\mathsf{Ba}}(B, \overline{Y})$ and $\psi_{\mathsf{Bc}}(B, \overline{Y})$ that formalize the subconditions (Ba) and (Bc). By the above lemma, we can then use the formula $\psi_{\mathsf{B}}(\overline{Y}) = \exists B(\psi_{\mathsf{Ba}}(B, \overline{Y}) \vee \psi_{\mathsf{Bc}}(B, \overline{Y}))$ for Condition B of Lemma 10.

6.1 Formalization of Condition Ba

Condition (Ba) is clearly expressible in $\mathrm{MLO(Inf)}$ and thus, over simple trees, in pure MLO as well, by the formula

$$\psi_{\mathsf{Ba}}(B, \overline{Y}) = \exists X\, \exists^{\aleph_0} w\, \mathrm{DINT}(T_{w\setminus B}, X, \overline{Y}),$$

where $T_{w\setminus B}$ is just a notation for the set defined by

$$x \in T_{w\setminus B} \iff w \leq x \wedge \neg \exists b \in B\, (b > w \wedge b \leq x).$$

The fact that Condition (Ba) is sufficient for the existence of continuum many sets U satisfying $\varphi(U, \overline{V})$ can be arrived at by appealing to the Composition Theorem in the same manner as for Condition A in the proof of Lemma 10, because the set X can be left intact or changed to another one with the same type on any of the infinitely many trees $\mathfrak{T}_{w\setminus B}$ which are D-intervals for X.

6.2 Formalization of Condition Bc

In order to eliminate the explicit use of the uncountability quantifier from Condition (Bc) over $(B, <) \cong (\omega, <)$, we use Proposition 2.5 from [7] reformulated using the standard equivalence of automata and MLO on $(\omega, <)$, as stated in the following proposition.

Proposition 14. *For every* MLO *formula* $\varphi(\overline{X},\overline{Y})$ *there exists an effectively constructable formula* $\psi(\overline{Y})$ *such that over* $(\omega,<)$

$$\psi(\overline{Y}) \equiv \exists^{\aleph_1}\overline{X}\,\varphi(\overline{X},\overline{Y}) \equiv \exists^{2^{\aleph_0}}\overline{X}\,\varphi(\overline{X},\overline{Y}).$$

Applying this result to the formula on the right hand side of Condition (Bc), with $\overline{Q}$ as parameters, we obtain a formula $\vartheta(\overline{Q})$ such that Condition (Bc) holds iff $(B,<) \models \vartheta(\overline{Q})$, with $\overline{Q}$ as specified there.

By Proposition 14, if $\vartheta(\overline{Q})$ holds, then there are even continuum many sets $\overline{P}$ satisfying Condition (Bc). As shown in the proof of Lemma 12 above, this ensures the existence of continuum many sets X satisfying $\psi(X,\overline{Y},B)$, because for each $\overline{P}$ a corresponding X exists. Thus, in this case there are continuum many sets X satisfying $\varphi(X,\overline{Y})$.

To formalize Condition (Bc) in MLO over the tree $\mathfrak{T}$, we first define the sets Q_i on $\mathfrak{T}$. As the set of types is computable, we can compute each τ_i and thus effectively construct the formula $\alpha_i(w,B,\overline{Y})$ expressing that w is a node on the branch B such that $\mathfrak{T}_{w\setminus B} \models \exists X\,\tau_i(X,\overline{Y}\cap\mathfrak{T}_{w\setminus B},\{w\})$, i.e. $w \in Q_i$. Using this formula we can express Condition (Bc) as $\psi_{\mathsf{Bc}}(B,\overline{Y}) =$

$$\exists\overline{Q}\left(\bigwedge_{i=1}^{r}\left(w\in Q_i \leftrightarrow \alpha_i(w,B,\overline{Y})\right) \wedge \vartheta^B(\overline{Q})\right),$$

where ϑ^B is a relativization of ϑ to the branch B.

7 The full binary tree and the Cantor space

In order to formalize Condition C in MLO over simple trees, we first analyze the problem only on the full binary tree and identify and prove the following key topological property that distinguishes counting branches from counting arbitrary sets.

On the full binary tree $\mathfrak{T}(2) = (\{0,1\}^*,\prec,S_0,S_1)$ where $\prec$ is the prefix-order and $S_i = \{0,1\}^*i$, we show that the set of branches satisfying any given MLO formula is a Borel set in the Cantor topology and hence it has the *perfect set property*: it is uncountable iff it contains a perfect subset iff it has the cardinality of the continuum. A *perfect set* is a closed set without isolated points.

The argument we present is based on basic results of descriptive set theory and the theory of finite automata on infinite words in connection with monadic second-order logic and the Borel hierarchy of the Cantor space. Let us recall a few basic notions from descriptive set theory and prove one theorem about the topological complexity of definable sets of paths. A thorough introduction to descriptive set theory can be found in [9], we only mention a few basic facts.

The Cantor space is the topological space with the product topology on $\{0,1\}^\omega$. It is a Polish space with the topology generated by basic neighborhoods $w\{0,1\}^\omega$ with the prefix $w \in \{0,1\}^*$. Alternatively, it can be defined by the metric $d(\alpha,\beta) = 2^{-\min\{n\,:\,\alpha[n]\neq\beta[n]\}}$.

The hierarchy of Borel sets is generated starting from open sets, i.e. unions of basic neighborhoods, denoted $\mathbf{\Sigma}^0_1$, and closed sets, which are complements of open sets and denoted $\mathbf{\Pi}^0_1$. Further on by transfinite induction for any countable ordinal α, $\mathbf{\Sigma}^0_\alpha$ is defined as $\{\bigcup_{i\in\omega} A_i \mid \forall i\,\exists\beta_i < \alpha\ A_i \in \mathbf{\Pi}^0_{\beta_i}\}$ and the $\mathbf{\Pi}^0_\alpha$-sets are the complements of $\mathbf{\Sigma}^0_\alpha$-sets. The projective hierarchy is built on top of the Borel hierarchy, starting with $\mathbf{\Sigma}^1_0 = \mathbf{\Pi}^1_0$ as the class of Borel sets. On the first level one has the class $\mathbf{\Sigma}^1_1$ of *analytic sets*, which are projections of Borel sets, and the class $\mathbf{\Pi}^1_1$ of *co-analytic sets*, whose complements of analytic. The hierarchy is built in this manner with sets in $\mathbf{\Sigma}^1_{\alpha+1}$ being projections of $\mathbf{\Pi}^1_\alpha$-sets, and $\mathbf{\Pi}^1_{\alpha+1}$ sets being complements of $\mathbf{\Sigma}^1_\alpha$ sets.

The connection between the topological complexity of MLO-definable tree languages and the complexity of tree-automata recognizing them is well understood [15, 11]. By Rabin's complementation theorem, all MLO-definable tree languages are in $\mathbf{\Sigma}^1_2 \cap \mathbf{\Pi}^1_2$. There are $\mathbf{\Sigma}^1_1$-complete as well as $\mathbf{\Pi}^1_1$-complete regular tree languages. For instance, the set of $\{a,b\}$-labeled binary trees, which have on every path only finitely many a's, is $\mathbf{\Pi}^1_1$-complete [1, 11]. There also exist regular tree languages not contained in $\mathbf{\Sigma}^1_1 \cup \mathbf{\Pi}^1_1$, however languages accepted by deterministic tree automata are contained in $\mathbf{\Pi}^1_1$. In contrast, by McNaughton's theorem, ω-regular languages, i.e. MLO-definable sets of ω-words, are boolean combinations of $\mathbf{\Pi}^0_2$ sets [15].

The Cantor-Bendixson Theorem states that closed subsets of a Polish space have the *perfect set property*: they are either countable or contain a perfect subset and thus have cardinality continuum. A set P is *perfect* if it is closed and if every point $p \in P$ is a condensation point of P, i.e. if every neighborhood of p contains another point from P. We shall rely on the following fundamental result due to Souslin.

Theorem 15 (cf. e.g. in [9])**.** *A subset of a Polish space is Borel if and only if it is both analytic and co-analytic. Moreover, every uncountable analytic set contains a perfect subset.*

Note that whether co-analytic sets, or all sets on higher levels of the projective hierarchy, satisfy the continuum hypothesis is independent of ZFC [9].

A key observation that our formalization will exploit is that even though there are non-Borel sets of trees definable in MLO, sets of definable paths are Borel. Recall that for a sequence $\pi \in \{0,1\}^*$ we denote by $\mathrm{Pref}(\pi)$ the path through $\mathfrak{T}(2)$ that corresponds to this sequence, which formally is the set of prefixes of π.

Theorem 16 (MLO definable sets of branches are Borel)**.** *Let* $U_1,\ldots,U_m$ *be subsets of* $\mathfrak{T}(2)$ *and let* $\psi(X,\overline{Y})$ *be an* MLO *formula over* $\mathfrak{T}(2)$*. Then the set*

$$\mathcal{X} = \{\,\pi\in\{0,1\}^\omega \mid \mathfrak{T}(2) \models \psi(\mathrm{Pref}(\pi),\overline{U})\,\}$$

of branches of the binary tree satisfying $\psi(X, \overline{U})$ is Borel and therefore has the perfect set property.

Proof. Note that the complement of $\mathcal{X}$ is also definable by $\neg\psi(X, \overline{U})$. We will show that every definable set of branches is analytic. Therefore, by Souslin's Theorem, it is Borel. To prove this, we will use the following variation of the Composition Theorem (c.f [8]), proved in Appendix B.

Lemma 17. *Let $\psi(X, Y_1, \ldots, Y_m)$ be an* MLO *formula with quantifier rank $n \geq 2$, and let k be the number of $(n+2)$-types in $m+1$ variables. Then there exists an* MLO *formula $\theta(I, Z_1, \ldots, Z_k)$ such that*

$$\mathfrak{T}(2) \models \psi(\mathrm{Pref}(\pi), \overline{U}) \iff$$
$$\iff (\omega, <) \models \theta(\{n \mid \pi[n] = 1\}, \overline{Q}),$$

where for each $1 \leq i \leq k$ we define $Q_i = Q_i^{\pi, \overline{U}}$ as

$$Q_i = \{j \in \omega \mid \mathrm{Tp}^{n+2}(\mathfrak{T}(2)_{\pi|_j}, \overline{U}) = \tau_i\}.$$

Let θ be the formula obtained by applying the above lemma to ψ. Then, by the well-known correspondence of MLO and finite automata on ω-words, there is an ω-regular language $\mathcal{L}_\theta \subseteq (\{0,1\}^{k+1})^\omega \cong \{0,1\}^\omega \times (\{0,1\}^k)^\omega$, such that $\mathcal{L}_\theta$ consists of those pairs of sequences (π, ρ) for which $(\omega, <) \models \theta(P, \overline{Q})$, where P and $\overline{Q}$ are subsets of ω with characteristic sequences $\pi \in \{0,1\}^\omega$ and $\rho \in (\{0,1\}^k)^\omega$. By McNaughton's theorem, cf. [15], $\mathcal{L}_\theta \in \mathbf{\Sigma}^0_3$.

Let $\mathcal{T}$ be the extension of $\mathfrak{T}(2)$ with each node w labeled by $(\sigma, \overline{q})$ such that w is the σ-th successor of its parent (i.e. $w \in S_\sigma$) and $\overline{q} = (0, \ldots, 0, 1, 0, \ldots, 0)$ with the 1 in position i if $\mathrm{Tp}^{n+2}(\mathfrak{T}(2)_w, \overline{U}) = \tau_i$. The set $[\mathcal{T}]$ of labeled infinite branches of $\mathcal{T}$ is closed in the Cantor topology.

By construction, $\mathcal{X}$ is the projection of $\mathcal{L}_\theta \cap [\mathcal{T}]$ to its first component, and is analytic as $\mathcal{L}_\theta \in \mathbf{\Sigma}^0_3$ and $[\mathcal{T}] \in \mathbf{\Pi}^0_1$. $\square$

8 Formalizing Condition C

In this section, we show how Theorem 16 above can be transferred to all simple trees by interpretation. This gives a characterization of Condition C that is expressible in MLO.

A perfect tree is a tree without isolated branches, where a branch is isolated iff it contains a node not contained in any other branch of the tree. Equivalently, a tree is perfect, if for every node u of this tree there are $v, w > u$ such that v and w are incomparable. Perfectness is thus first-order definable. The set $\mathcal{P}$ of branches of a perfect simple tree $\mathfrak{T}$ is topologically perfect. Conversely, every perfect set $\mathcal{P}$ of branches of a tree $\mathfrak{T}$ constitutes a perfect tree $(\bigcup \mathcal{P}, <^{\mathfrak{T}})$.

Proposition 18 (Eliminating the uncountably–many-branches quantifier)**.** *For every* MLO *formula $\varphi(X, \overline{Y})$ there is an* MLO *formula $\psi(\overline{Y})$ such that over all simple trees "$\exists^{\aleph_1} B$ branch$(B) \wedge \varphi(B, \overline{Y})$" is equivalent to $\psi(\overline{Y})$. Furthermore, ψ is computable from φ and implies the existence of continuum many branches B satisfying $\varphi(B, \overline{Y})$.*

Proof. Let $\psi(\overline{Y})$ be the MLO formula expressing that there is a prefix-closed set of nodes Λ, such that $(\Lambda, <)$ is a perfect tree and every infinite branch $B \subset \Lambda$ satisfies $\varphi(B, \overline{Y})$.

By definition of perfecness, ψ implies that there are uncountably many branches B satisfying $\varphi(B, \overline{Y})$ over any tree. As we have shown in Theorem 16, over the full binary tree with arbitrary additional unary predicates, ψ is equivalent to this condition. We transfer the result to all simple trees using an encoding of any simple tree in $\mathfrak{T}(2)$ with appropriate predicates, as follows.

Every simple l-tree $\mathfrak{T}$ is isomorphic to some $(T, \prec, P_1, \ldots, P_l)$ where $T \subseteq \mathbb{N}^*$ is a prefix-closed subset of finite sequences of natural numbers and $\prec$ is the prefix relation. Consider the following encoding $\mu : \mathbb{N}^* \to \{0,1\}^*$

$$(n_0, n_1, \ldots, n_s) \mapsto 0^{n_0} 1 0^{n_1} 1 \ldots 0^{n_s} 1,$$

and set $S = \mu(T)$ and $Q_i = \mu(P_i)$ for each $i = 1 \ldots l$.

Given that $v \prec w$ in $\mathfrak{T}$ iff $\mu(v) \prec \mu(w)$ in $\mathfrak{T}(2)$, this defines an interpretation of $\mathfrak{T}$ inside $(\mathfrak{T}(2), S, Q_1, \ldots, Q_l)$. In particular, for every MLO-formula $\vartheta(\overline{X})$ of l-trees

$$\mathfrak{T} \models \vartheta(\overline{U}) \iff (\mathfrak{T}(2), S, Q_1, \ldots, Q_l) \models \vartheta^*(\mu(\overline{U})),$$

where ϑ^* is obtained from ϑ by interpreting each P_i with Q_i and relativizing all quantifiers to subsets/elements of S.

Observe that μ induces a function μ^* mapping each infinite branch B of $\mathfrak{T}$ to the unique infinite branch $\mu^*(B)$ of $\mathfrak{T}(2)$ containing $\mu(w)$ for all $w \in B$. Conversely, every infinite branch of $\mathfrak{T}(2)$ containing the μ-image of infinitely many nodes of $\mathfrak{T}$ is the μ^* image of the unique infinite branch of $\mathfrak{T}$ containing all of these nodes. Hence μ^* is injective (but not surjective).

Consider the formula $\varphi(B, \overline{Y})$ defining, with parameters $\overline{V}$ over $\mathfrak{T}$, an uncountable set of branches. Thus, over simple trees, it defines an uncountable set of infinite branches $\mathcal{D} = \{B \mid \mathfrak{T} \models \varphi(B, \overline{V})$ and B is an infinite branch$\}$.

Then, according to earlier remarks, $\mathcal{D}^* = \{\mu^*(B) \mid B \in \mathcal{D}\}$ is an uncountable set of branches of $\mathfrak{T}(2)$ and it is defined by "branch$(B) \wedge \exists$ infinite $P \subseteq B\ \varphi^*(P, \mu(\overline{V}))$" over $(\mathfrak{T}(2), S, Q_1, \ldots, Q_l)$.

Thus, by Theorem 16, there is a $\Lambda^* \subseteq \mathfrak{T}(2)$ inducing a perfect tree $(\Lambda^*, <)$, every infinite branch of which is in $\mathcal{D}^*$.

We claim that $\Lambda = \mu^{-1}(\Lambda^*)$ induces a perfect tree in $\mathfrak{T}$, every infinite branch of which is then in $\mathcal{D}$.

For one, because Λ^* is prefix-closed, so is Λ, therefore it induces a tree in $\mathfrak{T}$. We know moreover, as Λ^* is perfect, that the image $\mu^*(B)$ of every infinite branch $B \subset \Lambda$ is not isolated in Λ^*. Hence for every $w \in B$ there is an infinite branch $C^* \subset \Lambda^*$ different from $\mu^*(B)$ and such that

$\mu(w) \in C^*$. Therefore $w \in \mu^{-1}(C^*)$, which is a branch through Λ and is different from B. This shows that B is not isolated in Λ, and so Λ is perfect. $\square$

Applying the above proposition to the formula $\exists X\, \mathrm{DPATH}_\varphi(B, X, \overline{Y})$ from Condition C, we get the formula $\psi_{\mathsf{C}}(\overline{Y})$ which satisfies the following.

Corollary 19. *On any simple tree $\mathfrak{T}$, the formula $\psi_{\mathsf{C}}(\overline{Y})$ holds if and only if Condition C of Lemma 10 is satisfied. Moreover, if $\psi_{\mathsf{C}}(\overline{Y})$ holds, then there are continuum many D-paths altogether for all sets U satisfying $\varphi(U, \overline{Y})$.*

9 Summary of the proofs

As we have shown above, the conditions of Lemma 10 can be formalized in MLO over simple trees, thus we can again state the conclusion of this Lemma: $\mathfrak{T} \models \exists^{\aleph_1} X\ \varphi(X, \overline{Y})$ holds if and only if

$$\mathfrak{T} \models \varphi_{\mathsf{A}}(\overline{Y}) \vee \exists B(\varphi_{\mathsf{Ba}}(B, \overline{Y}) \vee \varphi_{\mathsf{Bc}}(B, \overline{Y})) \vee \varphi_{\mathsf{C}}(\overline{Y}).$$

Using this lemma, we can reduce any formula of $\mathrm{MLO}(\exists^{\aleph_1})$ to an MLO formula equivalent over the class of simple trees by inductively eliminating the inner-most occurrence of a cardinality quantifier. Theorem 1 follows.

Moreover, as we have shown in the corresponding sections, each of the conditions of Lemma 10 implies the existence of continuum many sets X satisfying $\varphi(X, \overline{Y})$. Therefore Theorem 2 follows as well.

10 Further results

The technique we used here can be applied to linear orders and leads to a proof of the following generalization of the theorem of Kuske and Lohrey (c.f. Proposition 14).

Theorem 20 (Elimination of the uncountability quantifier)**.**

(1) For every $\mathrm{MLO}(\exists^{\aleph_1})$ formula $\varphi(\overline{Y})$ there exists an MLO formula $\psi(\overline{Y})$ that is equivalent to $\varphi(\overline{Y})$ over the class of all ordinals.

(2) For every $\mathrm{MLO}(\exists^{\aleph_1})$ formula $\varphi(\overline{Y})$ there exists an MLO formula $\psi(\overline{Y})$ that is equivalent to $\varphi(\overline{Y})$ over the class of all countable linear orders. Moreover, $\exists^{\aleph_1} X\ \varphi(X, \overline{Y})$ is equivalent to $\exists^{2^{\aleph_0}} X\ \varphi(X, \overline{Y})$ over the class of countable linear orders.

Furthermore, in all these cases ψ is computable from φ.

The proof will be provided in an extension of this paper. Let us remark that (2) cannot be obtained by interpretations of countable linear orders in the full binary tree and the expressive equivalence of $\mathrm{MLO}(\exists^{\aleph_1})$ to MLO and to $\mathrm{MLO}(\exists^{2^{\aleph_0}})$ over the full binary tree.

Acknowledgment. We are very grateful to Sasha Rubin for insightful discussions at an earlier stage of this work.

References

[1] A. Arnold, J. Duparc, F. Murlak, and D. Niwiński. On the topological complexity of tree languages. In *Logic and Automata: History and Perspectives*, volume 2 of *Texts in Logic and Games*, pages 9–29. Amsterdam Uni. Press, 2007.

[2] A. Baudisch, D. Seese, H.-P. Tuschik, and M. Weese. *Decidability and Generalized Quantifiers*. Akademie-Verlag Berlin, 1980.

[3] J. R. Büchi. Weak second-order arithmetic and finite automata. *Zeitschrift für mathematische Logik und Grundlagen der Mathematik*, 6:66–92, 1960.

[4] C. C. Elgot. Decision problems of finite automata design and related arithmetics. *Transactions of the American Mathematical Society*, 98:21–51, 1961.

[5] Y. Gurevich. Monadic second-order theories. In J. Barwise and S. Feferman, editors, *Model-Theoretical Logics*, pages 479–506. Springer, 1985.

[6] D. Kuske and M. Lohrey. First-order and counting theories of ω-automatic structures. In *Proceedings of FOSSACS'06*, pages 322–336, 2006.

[7] D. Kuske and M. Lohrey. First-order and counting theories of omega-automatic structures. *Journal of Symbolic Logic*, 73:129–150, 2008.

[8] S. Lifsches and S. Shelah. Uniformization, choice functions and well orders in the class of trees. *Journal of Symbolic Logic*, 61:1206–1227, 1996.

[9] Y. N. Moschovakis. *Descriptive Set Theory*, volume 100 of *Studies in Logic and the Foundations of Mathematics*. North-Holland Publishing Company, 1980.

[10] D. Niwiński. On the cardinality of sets of infinite trees recognizable by finite automata. In *Mathematical Foundations of Computer Science, MFCS'91, Proceedings*, volume 520 of *LNCS*, pages 367–376. Springer, 1991.

[11] D. Niwiński and I. Walukiewicz. A gap property of deterministic tree languages. *Theoretical Computer Science*, 1(303):215–231, 2003.

[12] M. O. Rabin. Decidability of second-order theories and automata on infinite trees. *Transactions of the American Mathematical Society*, 141:1–35, 1969.

[13] S. Shelah. The monadic theory of order. *Annals of Mathematics*, 102:379–419, 1975.

[14] W. Thomas. Ehrenfeucht games, the composition method, and the monadic theory of ordinal words. In *Structures in Logic and Computer Science*, volume 1261 of *Lecture Notes in Computer Science*, pages 118–143. Springer, 1997.

[15] W. Thomas and H. Lescow. Logical specifications of infinite computations. In J. W. de Bakker, W. P. de Roever, and G. Rozenberg, editors, *REX School/Symposium*, volume 803 of *LNCS*, pages 583–621. Springer, 1993.

[16] B. A. Trakhtenbrot. Finite automata and the logic of monadic predicates. *Dokl. Akad. Nauk SSSR*, 140:326–329, 1961.

A Infinity predicate on simple trees

Lemma 11. *There exists an* MLO *formula* $\psi_{\mathrm{Inf}}(X)$ *that holds on a simple tree* $\mathfrak{T}$ *if and only if* X *is infinite.*

Proof. By König's Lemma, a set $X \subseteq \mathfrak{T}$ is infinite if and only if there exists a path in $\mathfrak{T}$ containing infinitely many elements of X. The condition that a path P contains infinitely many elements of X can in turn be expressed in MLO by the formula

$$\psi_{\mathrm{infpath}}(P, X) = \forall x \in P\, \exists y \in P\, (x < y \wedge y \in X).$$

Thus, the correct formula $\psi_{\mathrm{Inf}}(X)$ is given by

$$\psi_{\mathrm{Inf}}(X) = \exists P\, (\mathrm{path}(P) \;\wedge\; \psi_{\mathrm{infpath}}(P, X)),$$

where $\mathrm{path}(P)$ expresses that P is a path. □

B Proof of Lemma 17

Lemma 17 is weaker than the full Composition Theorem for trees (Th. 8) of Lifsches and Shelah [8], as the index structure on which the tree is decomposed is a single branch and we consider a specific labeling. However, even if it is not very likely to be useful for other applications, we need this particular version for our proof.

Lemma 17. *Let* $\psi(X, Y_1, \ldots, Y_m)$ *be an* MLO *formula with quantifier rank* $n \geq 2$*, and let* k *be the number of* $(n+2)$*-types in* $m+1$ *variables. Then there exists an* MLO *formula* $\theta(I, Z_1, \ldots, Z_k)$ *such that*

$$\begin{aligned}\mathfrak{T}(2) \;\models\; \psi(\mathrm{Pref}(\pi), \overline{U}) \quad &\iff \\ \iff \quad (\omega, <) \models \theta(\{n \mid \pi[n] = 1\}, \overline{Q}),&\end{aligned}$$

where for each $1 \leq i \leq k$ *we define* $Q_i = Q_i^{\pi, \overline{U}}$ *as*

$$Q_j = \{j \in \omega \mid \mathrm{Tp}^{n+2}(\mathfrak{T}(2)_{\pi|_j}, \overline{U}) = \tau_i\}.$$

Proof. To construct θ, we first apply the Composition Theorem (Th.8) to $\psi(X, \overline{Y})$ on the full binary tree $\mathfrak{T}(2)$ decomposed along any branch B. This yields an MLO formula $\theta_0(\overline{T})$ such that, for every branch B of $\mathfrak{T}(2)$,

$$\mathfrak{T}(2) \models \psi(\mathrm{Pref}(\pi), \overline{U}) \iff (B, \prec) \models \theta_0(\overline{P}).$$

Here, by definition of $P_r = P_r^{B;\mathrm{Pref}(\pi), \overline{U}}$, holds for each n-type τ_r, each $\iota \in \{0, 1\}$ and $v \in P_r$ that $v\iota \in B$ if and only if τ_r is the n-type of $(\mathrm{Pref}(\pi), \overline{U})$ on the tree segment $\mathfrak{T}(2)_v \setminus \mathfrak{T}(2)_{v\iota}$.

As a first step we refine $\theta_0(\overline{P})$ to a formula $\theta_1(I, \overline{P})$ such that $(B, \prec) \models \theta_1\left(I, \overline{P}^{B;\mathrm{Pref}(\pi), \overline{U}}\right)$ if and only if all of the following three conditions hold:

- $(B, \prec) \models \theta_0\left(\overline{P}^{B, \mathrm{Pref}(\pi), \overline{U}}\right)$,
- $B = \mathrm{Pref}(\pi)$, and
- $I = B \cap S_1$.

Observe that a node $v \in B$ lies on the path π or is a 1-successor precisely if the n-type $\tau_r(X, \overline{Y})$ such that $v \in P_r$ stipulates that X is not empty, or that the root belongs to S_1, respectively. As we assumed that $n \geq 2$, let H and G be the sets of those n-types $\tau_r(X, \overline{Y})$ from which $\exists x\, (x \in X)$, respectively, $\exists x\, \forall y (x \leq y) \wedge x \in S_1$, can be inferred. Then we set $\theta_1(I, \overline{P})$ to be

$$\theta_0(\overline{P}) \wedge \forall v \left(\bigvee_{\tau_r \in H} v \in P_r \wedge (v \in I \leftrightarrow \bigvee_{\tau_r \in G} v \in P_r) \right),$$

and it indeed has the above property, i.e.

$$\begin{aligned}\mathfrak{T}(2) \;\models\; \psi(\mathrm{Pref}(\pi), \overline{U}) \quad &\iff \\ \iff \quad (\omega, <) \models \theta_1(\{n \mid \pi[n] = 1\}, \overline{T}^{(\pi, \overline{U})}),&\end{aligned}$$

with $T_r = \{i \in \omega \mid \tau_r = \mathrm{Tp}^n(\mathfrak{T}(2)_{\pi|_i \setminus \pi|_{i+1}}, \{\pi|_i\}, \overline{U})\}$ for each n-type τ_r.

Finally, for each $i \in \{0, 1\}$ and $(n+2)$-type $\sigma_s(\overline{Y})$ and n-type $\tau_r(X, \overline{Y})$ we define the relationship $\sigma_s \vdash_i \tau_r$, meaning that σ_s ensures that τ_r is the n-type of the tree segment obtained by removing the subtree of the i-th successor of the root. This condition is expressible with a formula of quantifier rank $n+2$ as follows: (This explains the need for $(n+2)$-types.)

$$\begin{aligned}\sigma_s(\overline{Y}) \;\models\; \exists z\, \exists Z \Big(& \forall x (z \leq x) \wedge \\ & \exists y \Big(\; y \in S_i \wedge \\ & \qquad \forall x (x < y \rightarrow x = z) \wedge \\ & \qquad \forall x\, (x \in Z \leftrightarrow y \not\leq x) \Big) \wedge \\ & \tau_r^Z(\{z\}, \overline{Y}|_Z) \; \Big),\end{aligned}$$

where the superscript Z denotes relativization to Z. Finally, θ can be defined as promised by $\theta(I, \overline{Q}) =$

$$\exists \overline{P}\, \forall n \bigwedge_{\sigma_s \vdash_i \tau_r} (n \in Q_s \wedge \mathrm{s}(n) \in S_i \rightarrow n \in P_r) \wedge \theta_1(I, \overline{P}).$$

where $\mathrm{s}(n)$ refers to the immediate successor of n, which is of course definable, but used here in functional notation for brevity. □