

EXPRESSING CARDINALITY QUANTIFIERS IN MONADIC SECOND-ORDER LOGIC OVER CHAINS

VINCE BÁRÁNY, LUKASZ KAISER, AND ALEXANDER RABINOVICH

Abstract. We investigate the extension of monadic second-order logic of order with cardinality quantifiers “there exists uncountably many sets such that ...” and “there exists continuum many sets such that ...”. We prove that over the class of countable linear orders the two quantifiers are equivalent and can be effectively and uniformly eliminated. Weaker or partial elimination results are obtained for certain wider classes of chains. In particular, we show that over the class of ordinals the uncountability quantifier can be effectively and uniformly eliminated. Our argument makes use of Shelah’s composition method and Ramsey-like theorem for dense linear orders.

§1. Introduction. The study of extensions of first-order logic with cardinality quantifiers goes back to at least Mostowski [14]. For a cardinal κ the quantifier $\exists^\kappa x$ asserts the existence of at least κ many *elements* with a given property. For this reason these cardinality quantifiers, along with their many generalisations to be found in the literature, e.g. the Magidor-Malitz quantifiers, can rightfully be called first-order cardinality quantifiers. Model theoretic properties and axiomatisability of first-order logic extended with the first-order uncountability quantifier $\exists^{\aleph_1} x$ as well as the possibility of eliminating $\exists^{\aleph_1} x$ have been widely investigated [12]. The book [2] presents results on decidability and other properties of first-order logic extended with cardinality quantifiers over various natural classes of structures.

In this paper we investigate over various classes of chains the expressive power of the extension of monadic second-order logic of order by the cardinality quantifiers $\exists^{\aleph_1} X$ and $\exists^{2^{\aleph_0}} X$ meaning “there exist uncountably many sets X ” and “there exist continuum many sets X ”, respectively.

Monadic second-order logic of order (henceforth 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, 6, 17, 18] when the logic was proved to be decidable over the class of finite linear orders. Büchi proved the decidability of MLO on ω [4] and later, together with Siefkes, extended this result to the monadic theory of every ordinal less than ω_2 [5]. Shelah [16] improved model-theoretic techniques that allow him to reprove almost all known decidability results about MLO theories as well as new decidability results for the case of linear orderings, and in particular dense orderings. He proved in particular that the MLO theory of the real line is undecidable. The frontier between decidable and undecidable cases was explored

in later papers by Gurevich and Shelah [8, 11, 10]; we refer the reader to the survey [9].

For a cardinal κ and a formula ψ let $\exists^\kappa X \psi$ express that “there are at least κ many sets X such that ψ holds”. We will denote by $\text{MLO}(\exists^\kappa)$ the extension of MLO by \exists^κ . Furthermore we will briefly consider extensions of MLO by the predicates $\mathbf{Inf}(X) \equiv$ “the set X is infinite” and $\mathbf{Unc}(X) \equiv$ “the set X is uncountable”, denoted $\text{MLO}(\mathbf{Inf})$ and $\text{MLO}(\mathbf{Unc})$, respectively. This is of course equivalent to allowing the use of the first-order cardinality quantifiers $\exists^{\aleph_0} x$ and $\exists^{\aleph_1} x$, respectively, counting *elements* with a certain property. We prefer the predicate notation to avoid confusion with the second-order cardinality quantifiers.

Call a linear order *almost complete* if its completion contains only countably many new points. This class includes all ordinals, all countable scattered linear orders and of course all complete linear orders. As a kind of starting point we will prove the following.

THEOREM 1. *For every MLO formula $\varphi(X, \bar{Y})$ there exists an $\text{MLO}(\mathbf{Unc})$ formula $\psi(\bar{Y})$ that is equivalent to $\exists^{\aleph_1} X \varphi(X, \bar{Y})$ over the class of almost complete linear orders.*

The predicate \mathbf{Unc} is easily expressible in MLO over chains of certain order types including, e.g., all ordinals. In the case of the real line little can be said in the absence of further set theoretic assumptions. Recall that in [7] Gurevich has shown, assuming the continuum hypothesis, that $\mathbf{Unc}(X)$ is expressible in MLO over a class of linear orders including the real line and relying, in fact, solely on topological properties of these orderings.

Theorem 1 immediately yields complete elimination of the uncountability quantifier over countable scattered chains. Next we will prove this for chains of order type of the rationals, which ultimately enables the extension to all countable chains. Our main results are summarised in the next two theorems.

THEOREM 2 (Elimination of the uncountability quantifier).

- (1) *For every $\text{MLO}(\exists^{\aleph_1})$ formula $\varphi(\bar{Y})$ there exists an MLO formula $\psi(\bar{Y})$ that is equivalent to $\varphi(\bar{Y})$ over the class of all ordinals.*
- (2) *For every $\text{MLO}(\exists^{\aleph_1})$ formula $\varphi(\bar{Y})$ there exists an MLO formula $\psi(\bar{Y})$ that is equivalent to $\varphi(\bar{Y})$ over the class of all countable linear orders.*

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

In addition to the above, the reduction will show that over countable linear orders 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.

THEOREM 3. *Let $\varphi(X, \bar{Y})$ be an arbitrary MLO formula. Then $\exists^{\aleph_1} X \varphi(X, \bar{Y})$ is equivalent to $\exists^{2^{\aleph_0}} X \varphi(X, \bar{Y})$ over all countable chains.*

All of the above trivially extend to cardinality quantifiers $\exists^{\aleph_0} \bar{X}$, $\exists^{\aleph_1} \bar{X}$ and $\exists^{2^{\aleph_0}} \bar{X}$ counting finite tuples of sets given that for any cardinal $\kappa \geq \aleph_0$

$$\exists^\kappa (X_0, X_1) \varphi \equiv \exists^\kappa X_0 \exists X_1 \varphi \vee \exists^\kappa X_1 \exists X_0 \varphi .$$

Our results generalise similar results of Kuske and Lohrey [13] obtained for ω using automata techniques.

Organisation. After recapitulating basic notions including elements of the composition method, the main vehicle of our arguments, over linear orders and fixing some notation in Section 2 and summarily dealing with the infinity quantifier in Section 3, we divide the discussion into several subsections gradually working our way towards establishing the above stated theorems. In Section 4 we introduce the notions of unique intervals and doubling intervals and of finite covers underlying our analysis and pin down their fundamental properties. Our first result, presented in Section 5, is the partial reduction of cardinality quantifiers over almost-complete linear orders as stated in Theorem 1. As corollaries we obtain complete elimination over the class of ordinals as well as over countable scattered linear orders. Section 6 sees another partial elimination result effective over all linear orders showing that the application of cardinality quantifiers can be restricted to cuts, i.e. to points of the completion of any given linear order. In Section 7 we prove that \exists^{\aleph_1} and $\exists^{2^{\aleph_0}}$ are equivalent and MLO-expressible over the ordering of the rationals. In Section 8 we show how to eliminate the cardinality quantifiers over linear orders uniformly decomposable into sums of linear orders when uniform elimination over the summands and elimination over the index structure are at hand. Finally, in Section 9 the results of the previous sections are combined to obtain uniform and effective elimination of cardinality quantifiers over all countable linear orders.

§2. Preliminaries. For a number $l \in \mathbb{N}$, $l > 0$, an l -labelled linear order (or simply l -chain) is a structure $\mathcal{L} = (L, <, P_1, \dots, P_l)$, where the P_i 's are unary predicates and $(L, <)$ is a linear order.

We denote the standard ordering of natural numbers by ω or $(\mathbb{N}, <)$, the orderings of integers and rational numbers are denoted $(\mathbb{Z}, <)$ and $(\mathbb{Q}, <)$, respectively. Recall that $(\mathbb{Q}, <)$ is dense, i.e. between any two elements $x < y$ there is another element z such that $x < z < y$. A linear order $(L, <)$ is *scattered* if $(\mathbb{Q}, <)$ cannot be embedded into it.

A subset I of a linear order $(L, <)$ is convex, if for all $x < y < z$ with $x, z \in I$ also $y \in I$. We use the word *intervals* referring to all convex subsets, not just when they are bounded, or have endpoints in L . For intervals with endpoints $a, b \in L$, whether open or closed on either end, we will use the standard notation, such as $[a, b) = \{x \in L \mid a \leq x < b\}$, etc. We write $L|_{[a,b]}$ for the order $L \cap [a, b)$, and for $X \subseteq L$ we use analogous notation, i.e. $X|_{[a,b]}$ for $X \cap [a, b)$. This notation is extended to k -tuples in the natural way, e.g. $\bar{X}|_I = (X_1|_I \dots X_k|_I)$.

A linear order is *complete* if every one of its subsets has a least upper bound. In this paper a *cut* of a linearly ordered set $(L, <)$ is a downward closed set $C \subseteq L$ such that if C has a least upper bound in L then it is contained in C . A *proper cut* is a cut that has no least upper bound in L , i.e. one that has no maximal element. The *completion* of the linear order L , denoted \bar{L} , is defined as the linear order $(\mathcal{C}(L), \subseteq)$ of cuts of L with the mapping $L \ni a \mapsto \{b \in L \mid b \leq a\} \in \mathcal{C}(L)$ as the canonical embedding.

2.1. Monadic logic of order. We work with l -chains in the relational signature $\{<, P_1, \dots, P_l\}$ where $<$ is a binary relation symbol interpreted as a total ordering and the P_i 's are unary predicates representing a labelling of the chain.

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, \dots interpreted as elements, and set variables X, Y, \dots interpreted as subsets of the domain. Set variables will always be capitalised 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 short-hands, such as $X \cup Y$, $X \cap Y$, $X \subseteq Y$, and relativisation of formulas to a set.

The quantifier rank of a formula φ , denoted $\text{qr}(\varphi)$, is the maximum depth of nesting of quantifiers in φ . For fixed n and l we denote by $\text{Form}_{n,l}$ the set of formulas of quantifier depth $\leq n$ and with free variables among X_1, \dots, X_l . Let $n, l \in \mathbb{N}$ and $\mathcal{L}_1, \mathcal{L}_2$ be l -chains. We say that \mathcal{L}_1 and \mathcal{L}_2 are n -equivalent, denoted $\mathcal{L}_1 \equiv^n \mathcal{L}_2$, if for every $\varphi \in \text{Form}_{n,l}$, $\mathcal{L}_1 \models \varphi$ iff $\mathcal{L}_2 \models \varphi$.

Clearly, \equiv^n is an equivalence relation. For any $n \in \mathbb{N}$ and $l > 0$, the set $\text{Form}_{n,l}$ 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.

LEMMA 4 (Hintikka Lemma). *For each $n, l \in \mathbb{N}$, we can compute a finite set $H_{n,l} \subseteq \text{Form}_{n,l}$ such that:*

- For every l -chain \mathfrak{A} there is a unique $\tau \in H_{n,l}$ such that $\mathfrak{A} \models \tau$.
- If $\tau \in H_{n,l}$ and $\varphi \in \text{Form}_{n,l}$, 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 -chain \mathcal{L} we denote by $\text{Tp}^n(\mathcal{L})$ the unique element of $H_{n,l}$ satisfied in \mathcal{L} and call it the n -type of \mathcal{L} . Thus, $\text{Tp}^n(\mathcal{L})$ determines (effectively) which formulas of quantifier-depth $\leq n$ are satisfied in \mathcal{L} .

We sometimes speak of the n -type of a tuple of subsets $\bar{V} = V_1, \dots, V_m$ of a given l -chain \mathcal{L} . This is synonymous with the n -type of the $(l+m)$ -chain (\mathcal{L}, \bar{V}) obtained as an expansion of \mathcal{L} with the predicates P_{l+1}, \dots, P_{l+m} interpreted by the sets V_1, \dots, V_m . This type will be denoted by $\text{Tp}^n(\mathcal{L}, \bar{V})$ and often referred to as an n -type in m variables, whereby the n -type of the $(l+m)$ -structure (\mathcal{L}, \bar{V}) is understood. Moreover, when considering substructures, e.g. $\mathcal{L}' \subseteq \mathcal{L}$, and given sets $\bar{X} \subseteq \mathcal{L}$, we write $\text{Tp}^n(\mathcal{L}', \bar{X})$ to denote $\text{Tp}^n(\mathcal{L}', \bar{X} \cap \mathcal{L}')$.

2.2. The composition method. The essence of the composition method is that certain operations on structures, such as disjoint union and ordered sums of linear orders, can be projected to n -theories, i.e. there are corresponding operations mapping n -theories of constituent structures to the n -theory of the resulting structure. In other words, n -theories can be composed. The method

was introduced by Shelah as an adaptation of the Feferman-Vaught Theorem to MLO [16].

DEFINITION 5 (Sums of l -chains). Given a linear order $(I, <)$ and a family of l -chains $\mathcal{L}_i = (L^i, <^i, \overline{P}^i)$ for every $i \in I$ the sum $\sum_I \mathcal{L}_i$ is defined as the l -chain over the set $\bigcup_{i \in I} L^i \times \{i\}$ such that

$$(a, i) < (a', i') \iff i < i' \text{ or } i = i' \text{ and } a <^i a',$$

and with monadic predicates $P_1, \dots, P_k, \dots, P_l$ defined as $(a, i) \in P_k$ iff $a \in P_k^i$.

We write $\mathcal{L}_0 + \mathcal{L}_1$ for the sum over $(\{0, 1\}, <)$. For example $(\mathbb{Z}, <)$ is isomorphic to $\omega^* + \omega$, where ω^* is the standard ordering of negative integers.

In its weakest form the composition theorem states that the n -type of a sum of chains, $\text{Tp}^n(\sum_{i \in I} \mathcal{L}_i)$, is uniquely determined by the chain of n -types of the summands $\langle \text{Tp}^n(\mathcal{L}_i) \mid i \in I \rangle$. A stronger statement is cited below as Theorem 28.

THEOREM 6 (Composition on linear orders I).

Let $(I, <)$ be a linear ordering. If $\langle \mathcal{L}_i \mid i \in I \rangle$ and $\langle \mathcal{L}'_i \mid i \in I \rangle$ are I -indexed sequences of chains such that $\text{Tp}^n(\mathcal{L}_i) = \text{Tp}^n(\mathcal{L}'_i)$ for all $i \in I$, then $\text{Tp}^n(\sum_{i \in I} \mathcal{L}_i) = \text{Tp}^n(\sum_{i \in I} \mathcal{L}'_i)$.

§3. Infinity quantifier. We begin with the simple observation that the second-order infinity quantifier $\exists^{\aleph_0} X$ can be eliminated uniformly over all structures with the aid of the predicate $\mathbf{Inf}(X)$ expressing that X is infinite or, equivalently, using the first-order infinity quantifier $\exists^{\aleph_0} x$.

PROPOSITION 7. $\text{MLO}(\exists^{\aleph_0})$ collapses effectively to $\text{MLO}(\mathbf{Inf})$.

PROOF. All occurrences of the quantifier \exists^{\aleph_0} can be successively eliminated observing that for any formula $\varphi(X, \overline{Y})$ 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.

$$\exists Z \left(\neg \mathbf{Inf}(Z) \wedge \forall X_1 X_2 \left((\varphi(X_1, \overline{Y}) \wedge \varphi(X_2, \overline{Y}) \wedge X_1 \neq X_2) \rightarrow \exists z \in Z (z \in X_1 \leftrightarrow z \notin X_2) \right) \right).$$

Indeed, (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, \dots, 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 the chosen elements. \dashv

Note that in an any linear order $\mathbf{Inf}(X)$ holds iff X contains an increasing or decreasing ω -sequence, i.e., $\exists Y (Y \subseteq X) (\text{decr}_\omega(Y) \vee \text{incr}_\omega(Y))$, where $\text{decr}_\omega(Y) = Y \neq \emptyset \wedge \forall t \in Y \exists t' (t' < t \wedge t' \in Y)$ and $\text{incr}_\omega(Y) = Y \neq \emptyset \wedge \forall t \in Y \exists t' (t' > t \wedge t' \in Y)$. Hence \mathbf{Inf} is uniformly MLO-definable over all linear orders.

COROLLARY 8. $\text{MLO}(\exists^{\aleph_0})$ collapses effectively to MLO over the class of linear orders.

The converse of Proposition 7 holds as well. In fact, the predicate $\mathbf{Inf}(X)$ 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 7, any of the quantifiers $\exists^\kappa Y$ with $\aleph_0 < \kappa \leq 2^{\aleph_0}$ can be used to define \exists^{\aleph_0} .

§4. Unique and doubling intervals. To eliminate a single occurrence of the uncountability quantifier from a formula $\exists^{\aleph_1} X \varphi(X, \bar{Y})$ we will make extensive use of the following notions for intervals.

DEFINITION 9. Let \mathcal{L} be a labelled chain and X, \bar{Y} subsets such that $\mathcal{L} \models \varphi(X, \bar{Y})$ and let I be an interval of \mathcal{L} .

- (1) We say that I is a *U-interval* for φ, X, \bar{Y} whenever $X \cap I$ is the *unique* subset of its type on $\mathcal{L}|_I$. More precisely, if $\mathcal{L}|_I \models \forall Z \tau(Z, \bar{Y}) \rightarrow Z = X$, where $\tau = \text{Tp}^n(\mathcal{L}|_I, X, \bar{Y})$ with n the quantifier rank of φ .
- (2) I is a *D-interval* for φ, X, \bar{Y} iff it is not a U-interval.
- (3) I is an *unsplittable D-interval* for φ, X, \bar{Y} iff it cannot be split into disjoint D-intervals.

The name ‘‘U-interval’’ attests to the fact that the set X in question is *uniquely* determined by its type on a given interval, as opposed to ‘‘D-intervals’’ offering two (or more) distinct choices for X with the same type on the interval, thus (at least) *doubling* the total number of choices for X over the entire domain. Whenever φ, X, \bar{Y} are clear from the context we will take the liberty of saying ‘‘ I is an U-interval’’ instead of ‘‘ I is U-interval for φ, X, \bar{Y} ’’. Similarly, for D-intervals and unsplittable D-intervals.

Note that all of these notions can be formalised in MLO. For example, there is an MLO formula $DINT_\varphi(X, \bar{Y}, I)$ expressing that I is a D-interval for φ, X, \bar{Y} and a formula $UNSP_\varphi(X, \bar{Y}, I)$ such that $L \models UNSP_\varphi(X, \bar{Y}, I)$ iff I is an unsplittable D-interval for φ, X, \bar{Y} .

LEMMA 10. *If there is an infinite set of pairwise disjoint D-intervals for some X satisfying $\varphi(X, \bar{Y})$ then there are at least continuum many such X .*

PROOF. If there are infinitely many disjoint D-intervals for some X satisfying $\varphi(X, \bar{Y})$ then there is an increasing or decreasing ω -sequence of disjoint D-intervals among them. The two cases being symmetric assume we have an increasing ω -sequence of D-intervals. The chain $(\mathcal{L}, X, \bar{Y})$ can then be written as a sum $\sum_{i \leq \omega} (\mathcal{L}|_{L_i}, X, \bar{Y})$ such that for all $i < \omega$ the interval L_i contains a D-interval, hence L_i is itself a D-interval for X , and allowing L_ω to be empty. This means that for all $i \in \omega$, there is a subset $X'_i \subseteq L_i$ such that $X'_i \neq X_i \cap L_i$ and $\text{Tp}^n(\mathcal{L}|_{L_i}, X, \bar{Y}) = \text{Tp}^n(\mathcal{L}|_{L_i}, X'_i, \bar{Y})$, where n is the quantifier rank of φ . We associate to every $H \subseteq \omega$ the distinct set

$$X_H = \bigcup \{X'_i \mid i \in H\} \cup \bigcup \{X \cap L_i \mid i \notin H\}.$$

From Theorem 6 it follows that $\text{Tp}^n(\mathcal{L}, X_H, \bar{Y}) = \text{Tp}^n(\mathcal{L}, X, \bar{Y})$ for every $H \subseteq \omega$. In particular, each of the continuum many X_H satisfies $\varphi(X_H, \bar{Y})$ in \mathcal{L} . \dashv

This observation motivates the introduction of the following key notion.

DEFINITION 11 (Finite U-U cover). Let $(L, <, X, \bar{Y})$ be a labelled chain such that $L \models \varphi(X, \bar{Y})$. Let I be an interval. Intervals $I_1 \dots I_k$ constitute a *finite U-U cover* of I for φ, X, \bar{Y} if $I = \bigcup_j I_j$ and each I_j is either a U-interval or an unsplittable D-interval for φ, X, \bar{Y} .

Again, we will most often take no mention of either φ, X , or \bar{Y} when these are understood, or can be arbitrary, as in the following observation.

EXAMPLE 12. Consider the formula $\varphi(X) = \forall xy (x < y \wedge y \in X \rightarrow x \in X)$ defining downwards-closed sets, i.e. cuts, on $L = (\mathbb{Q}, <)$. Observe that \mathbb{Q} is itself an unsplittable D-interval constituting a trivial finite U-U cover for every cut X . For every proper cut X there is a unique finite U-U cover of \mathbb{Q} consisting of two U-intervals, namely $(-\infty, \sup X)$ and $(\sup X, \infty)$. Let X_π be the set of all rationals smaller than π . Note that $[3, 4]$ is a D-interval for X_π because for any $r \in [3, 4] \setminus \mathbb{Q}$, the set $\{x \in \mathbb{Q} \mid x < r\}$ has the same type as X_π on $[3, 4]$. In $[3, 4]$ we can find a left and a right point, say $l = 3.1$ and $r = 3.5$, such that both $[l, 4]$ and $[3, r]$ are D-intervals for X_π – for this reason we will later say that the interval $[3, 4]$ is unbalanced. In fact, every interval containing π in its interior is a D-interval for X_π – a property, which we later formalise by defining essential points for X_π .

LEMMA 13. *If I has no finite U-U cover, then I can be split into two D-intervals such that one of them has no finite U-U cover.*

PROOF. Because I has no finite U-U cover, it is necessarily a D-interval, but not an unsplittable D-interval. It can thus be split into D-intervals I_1, I_2 with $I_1 \cap I_2 = \emptyset$ and $I_1 \cup I_2 = I$. Now if both I_1 and I_2 had a finite U-U cover, then this would yield a finite U-U cover of I . Therefore I_1 or I_2 has no finite U-U cover. \dashv

As a conclusion we obtain the following lemma.

LEMMA 14. *The following dichotomy holds (for each X):*

- (1) *either I contains infinitely many disjoint D-intervals (for X)*
- (2) *or I has a finite U-U cover (for X).*

PROOF. Assume that I has a finite U-U cover. Then it has a finite U-U cover I_1, \dots, I_k such that $I_i \cap I_j = \emptyset$ for all $1 \leq i < j \leq k$. As each I_j is either a U-interval or an unsplittable D-interval it cannot contain two disjoint D-intervals. This gives an upper bound $k + (k - 1)$ on the size of any collection of pairwise disjoint D-intervals inside I : at most k D-intervals each contained properly in separate I_j 's and at most $k - 1$ D-intervals intersecting two or more of the I_j 's.

Conversely, if $I_0 = I$ has no finite U-U cover then, by Lemma 13, it can be split into disjoint D-intervals I_1 and J_0 , with I_1 having no finite U-U cover. By induction we obtain D-intervals I_n and J_n for X such that $I_{n+1} = I_n \setminus J_n$ has no finite U-U cover for X . Then, for $m < n$, we have $J_m \cap J_n \subseteq J_m \cap I_n = \emptyset$. Hence we have found infinitely many pairwise disjoint D-intervals J_n for X . \dashv

Next we refine the notion of finite U-U covers as follows.

DEFINITION 15 (Balanced cover).

- (1) An unsplittable D-interval I is *left balanced* iff $I_{<v} = \{w \in I \mid w < v\}$ is a U-interval for every $v \in I$.
- (2) Similarly, an unsplittable D-interval I is *right balanced* iff for every $v \in I$ the interval $I_{>v} = \{w \in I \mid w > v\}$ is a U-interval.
- (3) An unsplittable D-interval I is *balanced* iff it is either left-balanced or right-balanced.
- (4) A finite U-U cover I_1, \dots, I_k is *balanced* iff for each $1 \leq j \leq k$, I_j is either a U-interval or a balanced unsplittable interval.

LEMMA 16. *An interval I has a finite U-U cover for X iff I has a balanced U-U cover for X .*

PROOF. We show that every non-balanced unsplittable D-interval I for X can be split¹ into two intervals L and R constituting a balanced U-U cover of I for X . This immediately implies the conclusion of the lemma.

Let I be a non-balanced unsplittable D-interval for X . Because I is not right-balanced, there is a point $v \in I$ such that $I_{>v}$ is a D-interval, consequently, $I_{<v}$ must be a U-interval, since I cannot be split into two D-intervals. The following set is therefore not empty.

$$L = \{v \in I \mid I_{<v} \text{ is a U-interval for } X\}$$

By definition, L is a downward-closed subinterval of I , and it is either a U-interval or a left-balanced unsplittable D-interval.

Let $R = I \setminus L$. Because I is not left-balanced R cannot be empty. Then one of R or L is a U-interval. We have seen that if L is a D-interval then it is left-balanced. Similarly, we need to show that if R is a D-interval then it is right-balanced. Notice that $I_{<v}$ is a D-interval for X for every $v \in R$, otherwise v would be in L . Therefore, since I cannot be split into disjoint D-intervals for X , $I_{>v}$ is a U-interval for X for every $v \in R$. Which means precisely that R is right-balanced. \dashv

LEMMA 17. *There is a function $N(k, l, \varphi)$ such that for every l -chain $\mathcal{L} = (L, <, \bar{Y})$ if $\{I_1, \dots, I_k\}$ is a balanced U-U cover of an interval I of \mathcal{L} for each of X_1, \dots, X_n then $n \leq N(k, l, \varphi)$ or $X_i \cap I = X_j \cap I$ for some $i \neq j$.*

PROOF. Let K be the number of $\text{qr}(\varphi)$ -types in $l + 1$ variables. Then, if J is a U-interval for $K + 1$ sets then two of these must realise the same type on J and hence have to coincide on J . Assume now that J is left-balanced for $2K + 1$ sets X_1, \dots, X_{2K+1} , so for each $v \in J$ the interval $J_{<v}$ is a U-interval for each of these sets X_i . If for each pair X_i, X_j with $i \neq j$ there is a point $p_{i,j} \in J$ on which these two sets differ, then all the $2K + 1$ sets differ on the interval $J_{\leq p}$ with $p = \max\{p_{i,j} \mid i, j \leq 2K + 1\}$. Therefore there are at least $K + 1$ among them which are pairwise different on $J_{\leq p}$, which contradicts the fact that $J_{<p}$ is a U-interval for all of these. The case of right-balanced intervals is treated symmetrically.

Classify the sets X_i into 2^k classes according to which of the I_1, \dots, I_k are left- or right-balanced for each X_i (considering U-intervals, say, as left-balanced). By

¹It is important to point out that the split depends on X .

the above, no class can contain more than $(2K)^k$ sets pairwise different on I . Therefore $N(k, l, \varphi) = (4K)^k$ satisfies the claim. \dashv

Combining Lemmas 10, 14, 16 and 17 we obtain the following criterion.

PROPOSITION 18. *Let $\mathcal{L} = (L, <, \bar{Y})$ be an l -chain and $\varphi(X, \bar{Y})$ an MLO-formula and $\aleph_1 \leq \kappa \leq 2^{\aleph_0}$. Then*

$$\mathcal{L} \models \neg \exists^\kappa X \varphi(X, \bar{Y})$$

if and only if there exists a subset U of the completion of L such that $|U| < \kappa$, and for every X satisfying $\varphi(X, \bar{Y})$ there is a finite balanced U - U cover of \mathcal{L} the end-points of which lie in U .

Every point of the completion of L can be represented by a cut – a subset of L . Hence, a direct formalisation of this criterion referring to the cardinality of a set of cuts would require a third-order predicate. So what have we gained? For one, from Proposition 18 it is immediate that over countable scattered linear orders, where the restriction on U becomes vacuous, \exists^{\aleph_1} and $\exists^{2^{\aleph_0}}$ are equivalent and can be effectively eliminated, cf. Corollary 19. In [13] Kuske and Lohrey obtained similar results for ω using more intricate automata techniques.

More generally, we observe that over almost complete linear orders the condition stated in Proposition 18 for $\kappa = \aleph_1$ can be formulated in $\text{MLO}(\mathbf{Unc})$. This yields further elimination results for \exists^{\aleph_1} , in particular, over ordinal chains.

Later in Section 6 we will show that in general U can be replaced by a subset of L and a *definable set of cuts*. This will allow us to reduce cardinality quantifiers over arbitrary sets to the weaker first-order cardinality quantifiers and the ostensibly simpler use of cardinality quantifiers applied solely to definable sets of cuts. Of course, the latter subsumes also the first-order cardinality quantifiers.

§5. Almost complete linear orders. After the preparations of the previous section we are ready to prove the first item of Theorem 2 concerning the collapse of $\text{MLO}(\exists^{\aleph_1})$ to MLO over the class of all ordinals. This will be a corollary of the elimination step embodied in Theorem 1 and valid uniformly over all *almost complete* linear orders. Recall that these have at most countably many proper cuts, i.e. cuts without a maximal element.

THEOREM 1. To every MLO-formula $\varphi(X, \bar{Y})$ one can effectively associate an $\text{MLO}(\mathbf{Unc})$ -formula $\psi(\bar{Y})$ that is equivalent to $\exists^{\aleph_1} X \varphi(X, \bar{Y})$ over the class of almost complete linear orders.

PROOF. We are going to show that over almost complete linear orders the condition stated in Proposition 18 can be formulated in $\text{MLO}(\mathbf{Unc})$. First, let U be as in Proposition 18 and let $V = U \cap L$. Note that $V \cup (\bar{L} \setminus L)$ is an overapproximation of U , which also fulfils the condition stated in Proposition 18.

Let M be a subset of L . Define an equivalence relation \sim_M as follows: $x \sim_M y$ if $[x, y] \subseteq M$ or $[x, y]$ is disjoint from M . Note that for every M , the equivalence classes of \sim_M are intervals of L and the following can be formalised in MLO:

- (i) The number of \sim_M classes is finite.
- (ii) Each \sim_M class is a U-interval or a balanced unsplittable D-interval.
- (iii) Whenever an end-point of a \sim_M -class lies in L then it is contained in V .

Note also, that if I_0, \dots, I_k are disjoint intervals that partition L , then there is an M such that the I_i are \sim_M -equivalence classes. Indeed if for $i < j$ the interval I_i precedes I_j then we can take as M the set $I_0 \cup I_2 \cup \dots \cup I_{2\lfloor k/2 \rfloor}$.

Over almost complete linear orders the conditions of Proposition 18 can thus be formalised by an MLO(**Unc**) formula expressing that there is a countable set V such that for all X satisfying $\varphi(X, \bar{Y})$ there is a set M such that the \sim_M classes constitute a finite balanced U-U cover for X and all end-points of \sim_M classes fall in $V \cup (\bar{L} \setminus L)$. \dashv

In cases where the uncountability predicate, equivalently, the first-order uncountability quantifier is MLO-definable the above technique can be used inductively to completely reduce MLO(\exists^{\aleph_1}) to MLO.

COROLLARY 19.

- (1) For every MLO(\exists^{\aleph_1}) formula $\varphi(\bar{Y})$ there exists an MLO formula $\psi(\bar{Y})$ that is equivalent to $\varphi(\bar{Y})$ over the class of countable scattered linear orders.
- (2) For every MLO(\exists^{\aleph_1}) formula $\varphi(\bar{Y})$ there exists an MLO formula $\psi'(\bar{Y})$ that is equivalent to $\varphi(\bar{Y})$ over the class of all ordinals.
- (3) (Under CH) For every MLO(\exists^{\aleph_1}) formula $\varphi(\bar{Y})$ there exists an MLO formula $\psi''(\bar{Y})$ that is equivalent to $\varphi(\bar{Y})$ over the reals.

PROOF. In all three cases one eliminates successively all uncountability quantifiers from ψ from the inside out by an application of Theorem 1 followed by the elimination of the predicate **Unc**(X).

Over countable linear orders the predicate **Unc** is vacuously always false, hence the first claim.

Gurevich [7] proved (assuming the continuum hypothesis) that the predicate “the set X is uncountable” is expressible in MLO over the reals. This proves the third claim.

It is well-known that “the set X is uncountable” is expressible in MLO over the class of all ordinals. Recall that a subset X of an ordinal α is uncountable iff the order type of $(X, <)$ is greater than or equal to ω_1 iff there is a subset $Y \subseteq X$ such that the cofinality of the order-type of $(Y, <)$ is strictly greater than ω . This is the case precisely if every subset $Z \subseteq Y$ such that the order-type of $(Z, <)$ is ω is bounded in Y . Formally, “the set X is uncountable” is equivalent over ordinals to

$$\exists Y \subseteq X \forall Z \subseteq Y \omega(Z) \rightarrow \exists y \in Y \forall z \in Z z < y$$

where $\omega(Z)$ expresses that the order type of $(Z, <)$ is ω , for instance by saying that Z is infinite and $[0, z) \cap Z$ is finite for every $z \in Z$. Finiteness can be expressed e.g. as shown in Proposition 7. \dashv

Let us stress again that, by Proposition 18, \exists^{\aleph_1} and $\exists^{2^{\aleph_0}}$ are equivalent over all countable scattered linear orders.

§6. Reduction to counting cuts. In this section we show that the existence of κ many sets satisfying an MLO formula can be reduced to the existence of κ many cuts (downward closed sets) satisfying some MLO formula effectively obtainable from the prior one. Cuts are of course just representations of points

of the completion of the underlying linear order. Hence we show that a single use of the second-order cardinality quantifier $\exists^\kappa X$ over a linear order L reduces to a single use of the corresponding first-order cardinality quantifier $\exists^\kappa x$ over its completion \bar{L} .

We say that a point of \bar{L} is a *splitting point* of a given finite U-U cover if it is an end-point of an interval in this cover. When convenient we may blur the distinction between a cut C of a linear order L and the corresponding point $\text{sup}C$ of the completion \bar{L} .

DEFINITION 20. A cut C is an *essential cut for X* , equivalently, $\text{sup}C \in \bar{L}$ is an *essential point for X* , if every interval I such that I intersects both C and its complement is a D-interval for X .

LEMMA 21.

- (1) If $\text{sup}C$ is an essential point for X then it is a splitting point of every finite balanced U-U cover for X .
- (2) If there is a finite balanced U-U cover for some X then there is also one whose non-essential splitting points belong to L .

PROOF.

- (1) Assume indirectly that there is an interval I of some balanced U-U cover for X and points $v, w \in I$ such that $v \in C$ and $w \notin C$. Then (v, w) must be a U-interval for X because either $\{x \in I \mid x < w\}$ or $\{x \in I \mid v < x\}$ is a U-interval and (v, w) is contained in both.
- (2) Consider wlog. a finite balanced U-U cover consisting of disjoint intervals bounded by consecutive elements $\sigma_1 < \sigma_2 < \dots < \sigma_t$ of the completion \bar{L} . If some $\sigma_j \in \bar{L} \setminus L$ is not an essential cut for X then, by definition, there is a U-interval I such that $\text{inf}I < \sigma_j < \text{sup}I$. Hence there are points $v, w \in L$ such that $[v, w] \subset I$ is a U-interval for X , and wlog. $\sigma_{j-1} < v < \sigma_j < w < \sigma_{j+1}$. The sequence $\sigma_1 < \dots < \sigma_{j-1} < v < w < \sigma_{j+1} < \dots < \sigma_t$ of splitting points thus gives rise to a new balanced U-U cover for X with fewer non-essential cut-points in $\bar{L} \setminus L$. Continuing this way in a finite number of refinement steps we arrive at a finite balanced U-U cover for X all of whose non-essential splitting points belong to L .

–

Now we can refine Proposition 18 as follows.

THEOREM 22. Let $\mathcal{L} = (L, <, \bar{Y})$ be an l -chain and $\varphi(X, \bar{Y})$ an MLO formula and $\aleph_1 \leq \kappa \leq 2^{\aleph_0}$. Then

$$\mathcal{L} \models \neg \exists^\kappa X \varphi(X, \bar{Y})$$

holds if and only if

- (i) there exists a set $V \subseteq L$ such that $|V| < \kappa$, and every X satisfying $\varphi(X, \bar{Y})$ has a finite balanced U-U cover with non-essential splitting points in V ;
- (ii) and the total number of essential cuts for all X satisfying $\varphi(X, \bar{Y})$ is strictly less than κ .

Note the slight advantage of the above over Proposition 18. Condition (i) can be formalised using the first-order cardinality quantifier and (ii), unlike the

condition of Proposition 18, refers to the cardinality of a definable set of cuts, i.e., a definable subset of the completion of \mathcal{L} .

PROPOSITION 23. *To every MLO-formula $\varphi(X, \bar{Y})$ one can effectively associate MLO-formulas $\alpha(V, \bar{Y})$ and $\beta(C, V, \bar{Y})$ such that for all $\aleph_0 \leq \kappa \leq 2^{\aleph_0}$ over all linear orders $\exists^\kappa X$ φ is equivalent to $\forall V (\alpha \rightarrow \exists^\kappa C(\text{cut}(C) \wedge \beta))$.*

Note that, in particular, each of the above formulas makes use of only a single occurrence of the respective \exists^κ , and only restricted to cuts.

PROOF. It is straightforward to give an MLO formula $ECUT_\varphi(C, X, \bar{Y})$ expressing that C is an essential cut for X . Thus, much as in the proof of Theorem 1, we can construct the formula $\alpha(V, \bar{Y})$ to express that “every X satisfying $\varphi(X, \bar{Y})$ has a finite balanced U-U cover with non-essential splitting points in V ”. This amounts to condition (i) of Theorem 22 without the cardinality constraint. Let further $\beta(C, V, \bar{Y}) = \text{“max}C \in V\text{”} \vee \exists X \varphi(X, \bar{Y}) \wedge ECUT_\varphi(C, X, \bar{Y})$, where $\text{max}C \in V$ is a shorthand for $\exists x \in V \forall y (y \in C \leftrightarrow y \leq x)$. Then there are at least κ many cuts C satisfying $\beta(C, V, \bar{Y})$ precisely if $|V| \geq \kappa$ or if condition (ii) of Theorem 22 fails. \dashv

§7. Rationals (counting Dedekind cuts). In this section we show how the uncountability quantifier can be eliminated in monadic second-order logic of order over the rationals. It will also be apparent that over the rationals $\exists^{\aleph_1} X\varphi$ and $\exists^{2^{\aleph_0}} X\varphi$ are equivalent for any MLO formula φ . Our argument makes use of a Ramsey-like theorem for additive colourings of dense chains due to Shelah.

DEFINITION 24.

- (1) A colouring of a chain C is a function $col : [C]^2 \rightarrow T$ where $[C]^2$ is the set of unordered pairs of distinct elements of C and T is a finite set (the set of colours).
- (2) The colouring f is additive if for every $x_1 < y_1 < z_1$ and $x_2 < y_2 < z_2$ in C , it hold that $col(x_1, y_1) = col(x_2, y_2)$ and $col(y_1, z_1) = col(y_2, z_2)$ implies $col(x_1, z_1) = col(x_2, z_2)$. In this case a partial operation $+$ is well defined on T : $t_1 + t_2 = t$ iff there are $x < y < z$ such that $col(x, y) = t_1$, $col(y, z) = t_2$ and $col(x, z) = t$.
- (3) A sub-chain $D \subseteq C$ is homogeneous (for col) if there exists some $t_0 \in T$ such that for every $x, y \in D$, $col(x, y) = t_0$.

Shelah [16, Theorem 1.3] proved the following remarkable theorem.

THEOREM 25 (Ramsey theorem for additive colourings). *Let $col : [C]^2 \rightarrow T$ be an additive colouring of a dense chain C using a finite set of colours T . Then there exists a homogeneous sub-chain $D \subseteq C$ for col that is everywhere dense in some open interval I of C .*

Recall that a *proper cut* is a cut having no supremum in the underlying linear order. A cut is non-trivial if neither it nor its complement is empty.

LEMMA 26. *Let $\psi(C, Y_1, \dots, Y_M)$ be an MLO formula and $V_1, \dots, V_M \subseteq \mathbb{Q}$. Then there are uncountably many — and in fact continuum many — Dedekind cuts C of \mathbb{Q} satisfying $(\mathbb{Q}, <) \models \psi(C, \bar{V})$ if and only if there is a subset $D \subseteq \mathbb{Q}$*

such that $(D, <)$ is dense and for every non-trivial proper cut C of $(D, <)$ the Dedekind cut $C' = \{q \in \mathbb{Q} \mid \exists p \in C : q < p\}$ of rationals satisfies $\psi(C', \bar{V})$.

PROOF. Only the necessity of the above condition requires consideration. To that end assume that there are uncountably many cuts satisfying ψ and say that two rationals q and q' are *close*, $q \asymp q'$, if $[\min(q, q'), \max(q, q')]$ contains only countably many cuts satisfying ψ ; and *far* otherwise. This defines an equivalence relation each equivalence class of which constitutes an interval of the rationals. These intervals are naturally linearly ordered and form a dense ordering. Indeed, by assumption there are at least two classes and by definition no two classes can form adjacent intervals, for otherwise their union would have to be part of a single class. In other words between any two points far apart there must be a third, which is far from both of these.

Assign to every pair $[q]_{\asymp} < [q']_{\asymp}$ of \asymp -classes as its colour the n -theory of the interval $L_{[q, q']} = \bigcup \{[p]_{\asymp} \mid [q]_{\asymp} \leq [p]_{\asymp} < [q']_{\asymp}\}$, where n is the quantifier rank of ψ .

$$\nu([q]_{\asymp}, [q']_{\asymp}) = \text{Tp}^n(L_{[q, q']}, <, Y_1 \cap L_{[q, q']}, \dots, Y_M \cap L_{[q, q']})$$

By composition, ν defines an additive binary colouring on \mathbb{Q}/\asymp . Now Theorem 25 asserts that there is an open interval \mathcal{I} of \mathbb{Q}/\asymp and a subset $\mathcal{O} \subset \mathcal{I}$, which is dense in \mathcal{I} and is ν -homogeneous. In other words there exists an n -theory τ such that $\nu([q]_{\asymp}, [q']_{\asymp}) = \tau$ for all $[q]_{\asymp} < [q']_{\asymp}$ in \mathcal{O} . Let D be an arbitrary complete set of representatives of the \asymp -classes in \mathcal{I} . In particular $(D, <)$ is dense, countable and without endpoints. Let $D_0 = \{q \in D \mid [q]_{\asymp} \in \mathcal{O}\}$ and let $I = \bigcup \mathcal{I}$.

Consider now a non-trivial proper cut C of D and let $C' = \{q \in \mathbb{Q} \mid \exists p \in C : q < p\}$ as in the statement of this lemma. Because D_0 is dense in D there exist \mathbb{Z} -chains $\dots < p_{-2} < p_{-1} < p_0 < p_1 < \dots$ in $D_0 \cap C$ and $\dots < q_{-2} < q_{-1} < q_0 < q_1 < \dots$ in $D_0 \setminus C$ such that $C = \{d \in D \mid \exists z p_z \leq d < p_{z+1}\}$ and, similarly, $D \setminus C = \{d \in D \mid \exists z q_z \leq d < q_{z+1}\}$. In particular, there is no $d \in D$ such that $p_z < d < q_z$ for all $z \in \mathbb{Z}$, which also means that there is in fact no $q \in \mathbb{Q}$ such that $p_z < q < q_z$ for all $z \in \mathbb{Z}$. Therefore we have $I \cap C' = \{q \in \mathbb{Q} \mid \exists z p_z \leq q < p_{z+1}\}$ and $I \setminus C' = \{q \in \mathbb{Q} \mid \exists z q_z \leq q < q_{z+1}\}$.

By composition and homogeneity of \mathcal{O} , the n -theories $\text{Tp}^n(I \cap C', <, Y_1 \cap I \cap C', \dots, Y_M \cap I \cap C')$ and $\text{Tp}^n(I \setminus C', <, Y_1 \cap I \setminus C', \dots, Y_M \cap I \setminus C')$ are obtained as the \mathbb{Z} -fold product of τ with itself and as such are independent of the choice of C . By the composition theorem again it follows that either every C' as above satisfies $\psi(C', Y_1, \dots, Y_M)$ or none does. The latter possibility can be immediately ruled out on the grounds that any two points of D are by definition far apart meaning that there must be uncountably many Dedekind cuts between them satisfying ψ of which at most countably many do not induce, equivalently, are not induced by a proper cut of D . \dashv

The “only if” condition of Lemma 26 is clearly MLO expressible. Combined with Proposition 23 this yields full and effective elimination of \exists^{\aleph_1} over $(\mathbb{Q}, <)$.

PROPOSITION 27 (Elimination of \exists^{\aleph_1} over the rationals). *For every MLO-formula $\varphi(X, \bar{Y})$ one can compute an MLO-formula $\psi(\bar{Y})$, which is equivalent over the standard ordering of the rationals to both $\exists^{\aleph_1} X \varphi(X, \bar{Y})$ and $\exists^{2^{\aleph_0}} X \varphi(X, \bar{Y})$.*

PROOF. The proof is by induction on the structure of the formula. To eliminate an inner-most occurrence of the uncountability quantifier one applies first Proposition 23 followed by an application of Lemma 26. \dashv

§8. Sums of linear orderings. In the following we will make use of a more informative statement on composition of types on sums of linear orderings as formulated by Shelah. Recall that we denote by $H_{n,k}$ the set of Hintikka formulas of quantifier depth n having k free variables (see Lemma 4).

THEOREM 28 (Composition on linear orders II [16]). *For every MLO-formula $\varphi(\bar{X})$ in the signature of l -chains having m free variables and quantifier rank n , and given the enumeration $\tau_1(\bar{X}), \dots, \tau_k(\bar{X})$ of $H_{n,l+m}$, there exists an MLO-formula $\theta(Q_1, \dots, Q_k)$ computable from the above and such that for every linear ordering $\mathfrak{J} = (I, <^I)$ and family $\{\mathcal{L}_i \mid i \in I\}$ of l -chains and subsets V_1, \dots, V_m of $\sum_{i \in I} \mathcal{L}_i$,*

$$\sum_{i \in I} \mathcal{L}_i \models \varphi(\bar{V}) \iff \mathfrak{J} \models \theta(Q_1, \dots, Q_k)$$

where the predicates \bar{Q} form a partition of I induced by \bar{V} as follows: for each $1 \leq r \leq k$,

$$Q_r = Q_r^{I; \bar{V}} = \{i \in I \mid \text{Tp}^n(\mathcal{L}_i, \bar{V}) = \tau_r\}.$$

Using this theorem we can formulate some general conditions allowing to reduce the problem of eliminating the uncountability quantifier \exists^{\aleph_1} over sums of linear orderings to eliminating \exists^{\aleph_1} over the index structure as well as eliminating it uniformly over the summands. This, of course, assuming that the sum is already given.

LEMMA 29. *Let $\varphi(X, \bar{Y})$ be an MLO formula of quantifier rank N , and let $\mathcal{L} = \sum_{i \in I} \mathcal{L}_i$ be an ordered sum of chains, and \bar{V} subsets of \mathcal{L} . Let the enumeration of the N -types be given by $\tau_1(X, \bar{Y}), \dots, \tau_K(X, \bar{Y})$ and let $\theta(\bar{T})$ be as in Theorem 28. Then there are uncountably many $U \subseteq \mathcal{L}$ satisfying $\mathcal{L} \models \varphi(U, \bar{V})$ if and only if one of the following conditions holds:*

- (a) *there is one such U having infinitely many disjoint D -intervals; or*
- (b) *there is one such U and an index $i \in I$ so that $\mathcal{L}_i \models \exists^{\aleph_1} Z \tau_r(Z, \bar{V}|_{\mathcal{L}_i})$ where τ_r is the N -type of $(U|_{\mathcal{L}_i}, \bar{V}|_{\mathcal{L}_i})$ on \mathcal{L}_i ; or*
- (c) *the set of those partitions \bar{P} of I that are induced by \bar{V} and some U satisfying $\varphi(U, \bar{V})$ is uncountable. This can be expressed by an MLO(\exists^{\aleph_1})-formula over the index structure:*

$$(I, <) \models \exists^{\aleph_1} \bar{P} : \text{Part}(\bar{P}) \wedge \bigwedge_{r=1}^K P_r \subseteq Q_r \wedge \theta(\bar{P})$$

where $\text{Part}(\bar{P})$ states that \bar{P} partition I and for each $r = 1 \dots K$ the set $Q_r = \{i \in I \mid \mathcal{L}_i \models \exists X' \tau_r(X', \bar{V}|_{\mathcal{L}_i})\}$.

If moreover \exists^{\aleph_1} is equivalent to $\exists^{2^{\aleph_0}}$ both over the index structure and on each of the summands then these two quantifiers are also equivalent over the sum.

PROOF. Each of the three conditions is sufficient to yield uncountably many sets U satisfying φ . For (a) this was proved in Lemma 10 by the weaker form of the composition theorem. Similarly, for (b) this also follows directly already from the weaker composition theorem. Finally, for condition (c) this follows from the fact that, for every one of the uncountably many tuples \bar{P} accounted for, there is a distinct set U inducing the type-predicates \bar{P} and fulfilling $\varphi(U, \bar{V})$.

Conversely, if condition (c) fails then there are only countably many colourings of I with type predicates \bar{P} induced by some U satisfying $\varphi(U, \bar{V})$. By failure of (a) for each of these type predicates we have for all but finitely many indices i that $i \in P_r$ implies that τ_r uniquely defines $U \cap L_i$ from $\bar{V}|_{L_i}$. Finally, if condition (b) fails too, then on each of the finitely many remaining intervals L_i there are also only countably many choices for $U \cap L_i$.

To see that the formalisation of condition (c) provided above is sound note that by Theorem 28 every U satisfying $\varphi(U, \bar{V})$ induces (together with \bar{V}) a partition \bar{P} satisfying the given formula. Conversely, each tuple \bar{P} satisfying it fulfils all the following: It forms a partition, it is induced by some set U together with \bar{V} as ensured by $\bigwedge_{r=1}^K P_r \subseteq Q_r$, and every U inducing it must satisfy $\varphi(U, \bar{V})$ thanks to $\theta(\bar{P})$. \dashv

A crucial point as to the applicability of the above claim is that it assumes a given factorisation of a linear ordering as a sum. We introduce the notion of definable splitting to facilitate the use of the above technique on classes of linear orderings over which an appropriate factorisation is uniformly definable.

DEFINITION 30 (Splitting). Let $\mathcal{L} = (L, <, \dots)$ be a chain and $\theta(x, y)$ a formula with x and y first-order variables.

- (1) We call θ a *splitting* of \mathcal{L} if $\{(a, b) \in L^2 \mid \mathcal{L} \models \theta(a, b)\}$ is an equivalence relation every class of which is an interval.
- (2) For a splitting θ of \mathcal{L} let $\sim_\theta^\mathcal{L}$ denote the equivalence relation defined by θ in \mathcal{L} , let $I_{\mathcal{L}/\theta}$ be the set of $\sim_\theta^\mathcal{L}$ -classes and $Ind_{\mathcal{L}/\theta} := (I_{\mathcal{L}/\theta}, <)$ the natural ordering of $I_{\mathcal{L}/\theta}$ according to representatives. Call $Ind_{\mathcal{L}/\theta}$ the *indexing order* of \mathcal{L} and $\mathcal{S}_{\mathcal{L}/\theta} = \{\mathcal{L}|_I \mid I \in I_{\mathcal{L}/\theta}\}$ the *summand structures* of \mathcal{L} w.r.t. θ .
- (3) Let \mathcal{C} be a class of labelled chains. We call θ a *splitting* of \mathcal{C} iff θ splits every $\mathcal{L} \in \mathcal{C}$. Let $Ind_{\mathcal{C}/\theta} := \{Ind_{\mathcal{L}/\theta} \mid \mathcal{L} \in \mathcal{C}\}$ and $\mathcal{S}_{\mathcal{C}/\theta} := \bigcup_{\mathcal{L} \in \mathcal{C}} \mathcal{S}_{\mathcal{L}/\theta}$. Call $Ind_{\mathcal{C}/\theta}$ and $\mathcal{S}_{\mathcal{C}/\theta}$ the *class of indexing chains* of \mathcal{C} and the *class of summand structures* of \mathcal{C} w.r.t. θ , respectively.

THEOREM 31. Let \mathcal{C} be a class of labelled chains and θ a splitting of \mathcal{C} . If

- (1) $MLO(\exists^{\aleph_1})$ collapses effectively to MLO over the class of indexing chains of \mathcal{C} w.r.t. θ , and
- (2) $MLO(\exists^{\aleph_1})$ collapses effectively to MLO over the class of summand structures of \mathcal{C} w.r.t. θ ,

then $MLO(\exists^{\aleph_1})$ collapses effectively to MLO over \mathcal{C} .

PROOF. Consider a formula $\varphi(X, \bar{Y})$ of MLO. We give a formula $\alpha \vee \beta \vee \gamma$ expressing in MLO the disjunction of the three conditions of Lemma 29 equivalent to $\exists^{\aleph_1} X \varphi(X, \bar{Y})$ uniformly over each $\mathcal{L} \in \mathcal{C}$ with the factorisation as defined by θ . Let τ_1, \dots, τ_K be an enumeration of $\text{Tp}(N, 1 + M)$ where N is the quantifier rank of φ and $M = |\bar{Y}|$.

Condition (a) can be expressed in MLO uniformly over all chains of a given signature. For instance by requiring the existence of an X satisfying $\varphi(X, \bar{Y})$ and an infinite set D such that every interval containing at least two points of D is a D-interval for X :

$$\alpha = \exists X \exists D \mathbf{Inf}(D) \wedge \forall \text{ interval } I (\exists d \neq d' \in D \cap I) \rightarrow DINT_\varphi(X, \bar{Y}, I).$$

The use of $\mathbf{Inf}(D)$ above is, of course, just a shorthand, it can be eliminated as in Proposition 7.

Condition (b) can be easily expressed in MLO relying on the elimination procedure for $\mathcal{S}_{C/\theta}$. By the latter, one obtains for each N -type $\tau_r(X, \bar{Y})$ an MLO formula $\nu_r(\bar{Y})$ equivalent to $\exists^{\aleph_1} Z \tau_r(Z, \bar{Y})$ over $\mathcal{S}_{C/\theta}$. Using these, condition (b) can be written as

$$\beta = \exists X \varphi(X, \bar{Y}) \wedge \exists I \exists x CLASS_\theta(x, I) \wedge \bigvee_r (\tau_r^I(X, \bar{Y}) \wedge \nu_r^I(\bar{Y}))$$

where $CLASS_\theta(x, I) = \forall y (y \in I \leftrightarrow \theta(x, y))$ defines I as the equivalence class of x with respect to θ ; and where for a formula ψ , we denote by ψ^I the relativisation of ψ to I .

Finally, to express condition (c) of Lemma 29 one needs to

- choose a set I of representatives of all the equivalence classes defined by θ ,
- relativise to I the MLO formula $\rho(\bar{Q})$ equivalent to $\exists^{\aleph_1} \bar{P} \dots$ of condition (c) over $Ind_{C/\theta}$ as delivered by the elimination procedure for this class,
- and substitute into ρ^I the sets \bar{Q} defined as

$$\Omega(\bar{Q}) = \bigwedge_{r=1}^K \forall x (x \in Q_r \leftrightarrow \exists X', L : CLASS_\theta(x, L) \wedge X' \subseteq L \wedge \tau_r^L(X', \bar{Y})).$$

With the customary shorthand “ $\exists! y$ ” meaning “there is a unique y such that” this formula takes the form

$$\gamma = \exists I (\forall x \exists! y \in I \theta(x, y)) \wedge \exists Q_1, \dots, Q_K \Omega(\bar{Q}) \wedge \rho^I(\bar{Q}).$$

–

§9. All countable linear orders. At last we are in a position to conclude that the quantifiers \exists^{\aleph_1} and $\exists^{2^{\aleph_0}}$ are equivalent and can be effectively eliminated from $MLO(\exists^{\aleph_1}, \exists^{2^{\aleph_0}})$ uniformly over all countable chains. Indeed, it is well known that every countable linear order arises as a dense sum of scattered linear orders, i.e. in the form $\sum_{q \in D} L_q$ where each L_q is a countable scattered linear order and D is either a singleton or is isomorphic to the standard ordering of the rationals with or without an additional minimal and/or maximal element.

Let $\theta(x, y)$ be the MLO-formula expressing that for no subset A of $L_{[x, y]}$ is $(A, <)$ a dense ordering, i.e. that $L_{[x, y]}$ is a scattered linear order. Over any countable chain θ defines an equivalence relation partitioning it into intervals coinciding with the summands L_q as above. Thus, θ is a splitting of the class of all countable linear orders.

Taking advantage of Theorem 31 and using the previously proven collapse results over the class of countable scattered linear orders (Corollary 19) and over the rationals (Proposition 27), — which trivially extends to the rationals with

either one or both endpoints added — we obtain uniform effective elimination of \exists^{\aleph_1} over the class of all countable chains. This completes the proof of Theorem 2(2) and Theorem 3 follows similarly.

§10. Further results. Combining Theorems 2 and 31 further elimination results can be obtained for a hierarchy of classes of linear orders. Starting from the classes of ordinals (and their reverses) and the class of countable linear orders, effective elimination of \exists^{\aleph_1} can be derived for e.g. ordinal sums of countable linear orders, or for countable sums of ordinals and reverse ordinals, and so on for any finite number of iterations of summation. However, this transformation of formulas is not uniform over the union of these classes.

Furthermore, the same technique employed here can be adapted to obtain similar results over trees. The more complex structure of trees does pose, however, some interesting additional challenges. In [1] we have considered the extension of MLO with cardinality quantifiers over finitely branching trees and obtained the following theorem generalising an earlier result of Niwinski [15].

THEOREM 32 ([1]). *Over finitely branching trees every $\text{MLO}(\exists^{\aleph_0}, \exists^{\aleph_1}, \exists^{2^{\aleph_0}})$ formula is effectively equivalent to an MLO formula. Moreover, $\exists^{\aleph_1} X \varphi(X, \bar{Y})$ and $\exists^{2^{\aleph_0}} X \varphi(X, \bar{Y})$ are equivalent for every MLO formula φ with parameters \bar{Y} over every finitely branching tree.*

Acknowledgement. We thank the anonymous referee for valuable suggestions that greatly helped to improve the paper. This research was facilitated by the ESF project AutoMathA. The first author was supported in part by ANR-06-MDCA-05 (2007-2009) DocFlow, and by the EPSRC grant EP/E010865/1. The third author was partially supported by ESF project Games and EPSRC grant EP/H018581/1.

REFERENCES

[1] V. BÁRÁNY, L. KAISER, and A. RABINOVICH, *Expressing Cardinality Quantifiers in Monadic Second-Order Logic over Trees*, *Fundamenta Informaticae*, vol. 100 (2010).
 [2] A. BAUDISCH, D. SEESE, H.-P. TUSCHIK, and M. WEESE, *Decidability and generalized quantifiers*, Akademie-Verlag Berlin, 1980.
 [3] JULIUS R. BÜCHI, *Weak second-order arithmetic and finite automata*, *Zeitschrift für mathematische Logik und Grundlagen der Mathematik*, vol. 6 (1960), pp. 66–92.
 [4] ———, *On decision method in restricted second order arithmetic*, *Proceedings of the international congress on logic, methodology and philosophy of science*, 1962, pp. 1–11.
 [5] JULIUS R. BÜCHI and DIRK SIEFKES, *The monadic second order theory of all countable ordinals*, Lecture Notes in Mathematics, vol. 328, Springer, 1973.
 [6] CALVIN C. ELGOT, *Decision problems of finite automata design and related arithmetics*, *Transactions of the American Mathematical Society*, vol. 98 (1961), pp. 21–51.
 [7] YURI GUREVICH, *Monadic theory of order and topology, I*, *Israel Journal of Mathematics*, vol. 27 (1977), pp. 299–319.
 [8] ———, *Modest theory of short chains I*, *The Journal of Symbolic Logic*, vol. 44 (1979), pp. 481–490.
 [9] ———, *Monadic second-order theories*, *Model-theoretical logics* (Jon Barwise and Solomon Feferman, editors), Springer, 1985, pp. 479–506.

- [10] YURI GUREVICH, MENACHEM MAGIDOR, and SAHARON SHELAH, *The monadic theory of ω_2* , *Journal of Symbolic Logic*, vol. 48 (1983), pp. 387–398.
- [11] YURI GUREVICH and SAHARON SHELAH, *Modest theory of short chains II*, *The Journal of Symbolic Logic*, vol. 44 (1979), pp. 491–502.
- [12] MATT KAUFMANN, *The Quantifier “There Exists Uncountably Many” and Some of Its Relatives*, *Model Theoretic Logics* (J. Barwise and S. Feferman, editors), Springer-Verlag, 1985, pp. 123–176.
- [13] DIETRICH KUSKE and MARKUS LOHREY, *First-order and counting theories of omega-automatic structures*, *The Journal of Symbolic Logic*, vol. 73 (2008), pp. 129–150.
- [14] ANDRZEJ MOSTOWSKI, *On a generalization of quantifiers*, *Fundamenta Mathematicae*, vol. 44 (1957), pp. 12–36.
- [15] DAMIAN NIWIŃSKI, *On the cardinality of sets of infinite trees recognizable by finite automata*, *Mathematical Foundations of Computer Science, MFCS’91, Proceedings*, LNCS, vol. 520, Springer, 1991, pp. 367–376.
- [16] SAHARON SHELAH, *The monadic theory of order*, *Annals of Mathematics*, vol. 102 (1975), pp. 379–419.
- [17] BORIS A. TRAKHTENBROT, *Finite automata and the logic of monadic predicates*, *Dokl. Akad. Nauk SSSR*, vol. 140 (1961), pp. 326–329.
- [18] ———, *Finite automata and the logic of one-place predicates*, *Siberian Mathematical Journal*, vol. 13 (1962), pp. 103–131, (in Russian).

VINCE BÁRÁNY

OXFORD UNIVERSITY COMPUTING LABORATORY
WOLFSON BUILDING, PARKS ROAD
OX1 3QD, OXFORD, UNITED KINGDOM

E-mail: vbarany@comlab.ox.ac.uk

LUKASZ KAISER

MATHEMATICAL FOUNDATIONS OF COMPUTER SCIENCE
RWTH AACHEN D-52056 AACHEN, GERMANY

E-mail: kaiser@logic.rwth-aachen.de

ALEXANDER RABINOVICH

TEL AVIV UNIVERSITY, THE BLAVATNIK SCHOOL OF COMPUTER SCIENCE
RAMAT AVIV, TEL AVIV 69978, ISRAEL

E-mail: rabinoa@post.tau.ac.il