An Ehrenfeucht-Fraïssé Game for $L_{\omega_1\omega}$

MSc Thesis *(Afstudeerscriptie)*

written by

Tong Wang
(born September 5th, 1988 in Jinan, China)

under the supervision of **Prof. Dr. Jouko Väänänen** and **Prof. Dr. Benedikt Löwe**, and submitted to the Board of Examiners in partial fulfillment of the requirements for the degree of

MSc in Logic

at the *Universiteit van Amsterdam.*

Date of the public defense:	**Members of the Thesis Committee:**
September 19th, 2012	Prof. Dr. Dick de Jongh
	Prof. Dr. Benedikt Löwe
	Prof. Dr. Jouko Väänänen
	Prof. Dr. Yde Venema

INSTITUTE FOR LOGIC, LANGUAGE AND COMPUTATION

Abstract

The Ehrenfeucht-Fraïssé Game is very useful in studying separation and equivalence results in logic. The usual finite Ehrenfeucht-Fraïssé Game EF_n characterizes separation in first order logic $L_{\omega\omega}$. The infinite Ehrenfeucht-Fraïssé Game EF_ω and the Dynamic Ehrenfeucht-Fraïssé Game EFD_α characterize separation in $L_{\infty\omega}$, the logic with arbitrary conjunctions and disjunctions of formulas. The logic $L_{\omega_1\omega}$ is the extension of first order logic with countable conjunctions and disjunctions of formulas. It is the most immediate, and perhaps the most important infinitary logic. However, there is no Ehrenfeucht-Fraïssé Game in the literature that characterizes separation in $L_{\omega_1\omega}$.

In this thesis we introduce an Ehrenfeucht-Fraïssé Game for the logic $L_{\omega_1\omega}$. This game is based on a game for propositional and first order logic introduced by Hella and Väänänen. Unlike the usual Ehrenfeucht-Fraïssé Games which are modeled solely after the behavior of quantifiers, this new game also takes into account the behavior of boolean connectives in logic. We prove the adequacy theorem for this game. In the final part of the thesis we apply this game to prove complexity results about infinite binary strings.

Contents

1 Why Do We Want an Ehrenfeucht-Fraïssé Game for $L_{\omega_1\omega}$?

There is a close connection between logic and games. Three central concepts in logic: *truth*, *consistency* and *separation* are characterized respectively by three types of games. Let L be a vocabulary, $\mathcal{M}$ be an L-structure[1] and T be an L-theory . The Semantic Game SG$(\mathcal{M}, T)$ deals with the concept of *truth*. It aims to determine whether a set of L-sentences T is true in a model $\mathcal{M}$. The Model Existence Game MEG(T, L) deals with the concept of *consistency*. It tackles the question whether a given set of L-sentences T is consistent, in the sense that it has a model, or equivalently in the sense that it is impossible to derive a contradiction syntactically from this set of sentences. The Ehrenfeucht-Fraïssé Game EF$(\mathcal{A}, \mathcal{B})$ deals with the concept of *separation*. We say that we can separate two models $\mathcal{A}$ and $\mathcal{B}$ if there exists a sentence that is true in one model and false in another. Separation can also be considered in terms of the more familiar concept of *equivalence*: to say that two models can be separated by a sentence in a certain language, is just another way of saying that these two model are not equivalent in this language. Each of these three games involves two players: one trying to assert that the concept at issue (truth, consistency or separation) is true of the situation, and the other trying to challenge this assertion. These three games characterize the corresponding concepts in the sense that in each case, the affirmation of the concept is equivalent to the existence of a winning strategy of one player.

The Semantic Game, the Model Existence Game, and the Ehrenfeucht-Fraïssé Game are closely linked to each other. There is a sense in saying that they are essentially three variants of just one basic game. This basic game is modeled after our understanding of the nature of quantifiers in logic. The Model Existence Game MEG(T, L) is like SG$(\mathcal{M}, \phi)$ with the model $\mathcal{M}$ missing. The Ehrenfeucht-Fraïssé Game EF$(\mathcal{A}, \mathcal{B})$ is like SG$(\mathcal{A}, T)$ and SG$(\mathcal{B}, T)$ with the theory T missing. The three games form an organic unity which we may call the *Strategic Balance of Logic* (see Figure 1). For a detailed survey of this topic, see Väänänen's book [23].

The prime example for the Strategic Balance is first order logic. For the precise definition of the Semantic Game, the Model Existence Game and the Ehrenfeucht-Fraïssé Game for first order logic, the reader is again referred to [23]. In first order logic the three games characterize the three corresponding concepts neatly. Moreover, the diagram in Figure 1 is commutative in the sense that a winning strategy in one game can be transferred into a winning strategy in another game.

The Semantic Game characterizes the notion of truth in first order logic. The idea of interpreting the quantifiers in terms of moves in a game, as in the Semantic Game, is due to Henkin [10].

Theorem 1 (Semantic Game)**.** *Suppose L is a vocabulary, T a set of first order*

[1] In this paper we use the terms 'model' and 'structure' interchangeably.

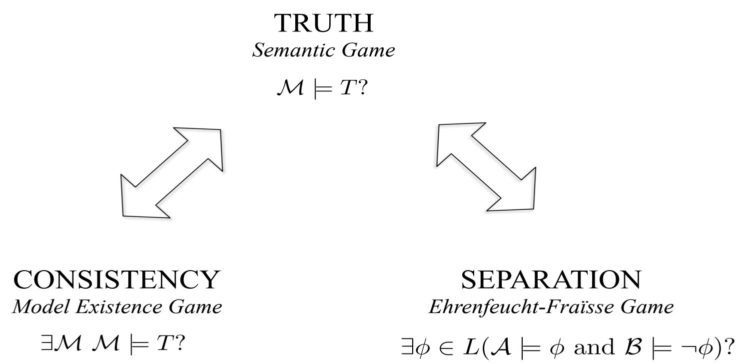

Figure 1: Strategic Balance for First Order Logic

L-sentences and $\mathcal{M}$ an L-structure. Then the following are equivalent:

1. *$\mathcal{M} \models T$.*
2. *Player **II** has a winning strategy in* $\mathrm{SG}(\mathcal{M}, T)$.

The Model Existence Game characterizes the notion of consistency in first order logic. The Model Existence Game is a game-theoretic rendering of the method of semantic tableaux of Beth [2, 3]. The following theorem proves the 'left leg' of the balance, namely the marriage of truth and consistency.

Theorem 2 (Model Existence Game)**.** *Suppose L is a countable vocabulary and T a set of first order L-sentences. Then the following are equivalent:*

1. *There is an L-structure $\mathcal{M}$ such that $\mathcal{M} \models T$.*
2. *Player **II** has a winning strategy in* $\mathrm{MEG}(T, L)$.

The Ehrenfeucht-Fraïssé Game characterizes the notion of separation in first order logic. The Ehrenfeucht-Fraïssé Game is first formulated by Ehrenfeucht in [4] and [5], whose idea is based on the work of Fraïssé [6]. The next theorem proves the 'right leg' of the balance, namely the marriage of truth and separation. For L-structures $\mathcal{A}$, $\mathcal{B}$ and a natural number n, let $\mathcal{A} \equiv^n_{L_{\omega\omega}} \mathcal{B}$ denote that $\mathcal{A}$ and $\mathcal{B}$ satisfy the same first order L-sentences up to quantifier rank n.

Theorem 3 (Ehrenfeucht-Fraïssé Game)**.** *Suppose L is a relational vocabulary. Suppose $\mathcal{A}$ and $\mathcal{B}$ are L-structures and $n \in N$. The following are equivalent:*

1. $\mathcal{A} \equiv^n_{L_{\omega\omega}} \mathcal{B}$.

2. *Player **II** has a winning strategy in* $\mathrm{EF}_n(\mathcal{A}|_{L'}, \mathcal{B}|_{L'})$ *for all finite* $L' \subseteq L$.

The presence of function symbols will make the Ehrenfeucht-Fraïssé Game more involved. As a way out we may consider a variant of the game tailored for *unnested formulas*. Every formula can be converted to a logically equivalent unnested formula at the price of raising its quantifier rank. An analogous equivalence result holds for this variant of the game and unnested formulas. For details, see Chapter 3 in [11]. For the sake of clarity, in this paper we always restrict our attention to relational structures when we consider separation and the Ehrenfeucht-Fraïssé Game.

Given its success in first order logic, a natural motive is to extend the Strategic Balance to a wider range of logics. In this paper we focus on the logic $L_{\omega_1\omega}$, the extension of first order logic with countable conjunctions and disjunctions of formulas. The logic $L_{\omega_1\omega}$ is, perhaps, the most important infinitary logic. We ask the question: does the Strategic Balance hold for the logic $L_{\omega_1\omega}$?

With respect to the 'left leg' of the balance—the connection between the Semantic Game and the Model Existence Game, that is—the answer is a clear 'yes'. The Semantic Game for first order logic extends to infinitary logic naturally (see Chapter 7 in [23]).

Theorem 4. *Suppose* L *is a vocabulary,* T *a set of* L*-sentences in* $L_{\omega_1\omega}$ *and* $\mathcal{M}$ *an* L*-structure. Then the following are equivalent:*

1. $\mathcal{M} \models T$.

2. *Player **II** has a winning strategy in* $\mathrm{SG}(\mathcal{M}, T)$.

The Model Existence Game for first order logic can also be modified to a Model Existence Game for $L_{\omega_1\omega}$ (see Chapter 8 in [23]). Theorem 1 finds the following analogue in the context of $L_{\omega_1\omega}$.

Theorem 5 (Model Existence Theorem for $L_{\omega_1\omega}$)**.** *Suppose* L *is a countable vocabulary and* ϕ *is an* L*-sentence of* $L_{\omega_1\omega}$*. The following are equivalent:*

1. *There is an* L*-structure* $\mathcal{M}$ *such that* $\mathcal{M} \models \phi$.

2. *Player **II** has a winning strategy in* $\mathrm{MEG}(\phi, L)$.

The Model Existence Theorem is essential to the model theory of $L_{\omega_1\omega}$. We may say that it more or less takes the role of the Compactness theorem, which is the corner stone of first order model theory but unfortunately fails for infinitary logic. In particular, the Model Existence Theorem can be used to prove the Interpolation Theorem for $L_{\omega_1\omega}$. For details, we refer again to [23].

Let us now consider the other leg of the balance, namely the connection between truth and separation. There is no straightforward extension of the Ehrenfeucht-Fraïssé Game to $L_{\omega_1\omega}$ like in the case of the Semantic Game and the Model Existence Game. The problem of finding an Ehrenfeucht-Fraïssé

Game that characterizes separation in $L_{\omega_1\omega}$, rather than in $L_{\infty\omega}$ up to a fixed quantifier rank, has been open since the 70s. This has been open even for propositional $L_{\omega_1\omega}$. Figure 2 shows the Strategic Balance for $L_{\omega_1\omega}$ with the 'right leg' missing.

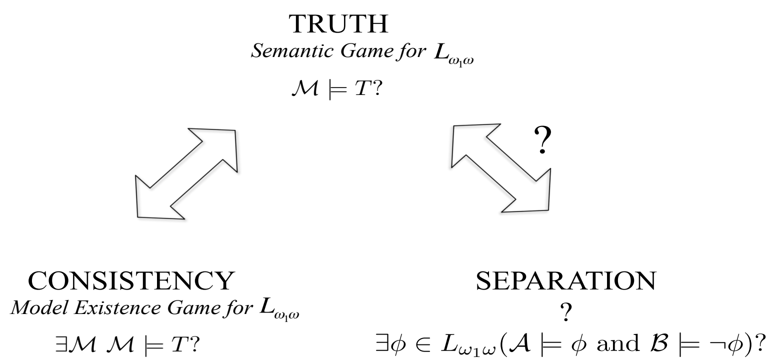

Figure 2: Strategic Balance for $L_{\omega_1\omega}$

In this thesis we fill in the missing connection between truth and separation in the Strategic Balance of $L_{\omega_1\omega}$. The text is divided as follows. We start our quest with a quick survey of the mathematical concept of game and the infinitary logic $L_{\omega_1\omega}$ in Section 2. In Section 3 we look at the 'usual' Ehrenfeucht-Fraïssé Games. The usual Ehrenfeucht-Fraïssé Games characterize separation in first order logic $L_{\omega\omega}$ and in the infinitary logic $L_{\infty\omega}$. The logic $L_{\omega_1\omega}$ falls in the gap between $L_{\omega\omega}$ and $L_{\infty\omega}$. The reason that the usual Ehrenfeucht-Fraïssé Games fail to characterize separation in $L_{\omega_1\omega}$ is, as we have indicated, that these games reflect our understanding of the nature of quantifiers in logic. But in order to characterize separation in the logic $L_{\omega_1\omega}$, where the 'boolean size' of a formula is also at issue, the game needs to take into account not only the behaviour of quantifiers, but also of boolean connectives. Based on this idea and a game introduced by Hella and Väänänen in [9], we introduce in Section 4 an Ehrenfeucht-Fraïssé Game for $L_{\omega_1\omega}$. We prove the adequacy theorem for this new game (Theorem 23), which is the main result of this thesis. In Section 5 we apply this new game to prove a complexity result about infinite binary strings (Theorem 36). This is the first propositional complexity result for $L_{\omega_1\omega}$.

2 Preliminaries

2.1 The Mathematical Concept of Game

Everyone has an intuitive notion of game. When we think of a game, the first idea that comes to mind is that of playing the game as an *act*. As an example, consider the game of chess. The game of chess involves two *players*. They take turns to make their moves. The possible moves of the players are given by a set of explicitly stated *rules*. A player may choose, among the possible moves, the next move that she makes using her intelligence and the information she has about the history and the current situation of the game. The winner is determined by a set of explicit *winning conditions*. If a player has a plan that guarantees her to win no matter how the other participant of the game plays, we say that she has a *winning strategy*. The games that we encounter in logic are very much like chess, the difference being that now the two players do not play with pieces on a chessboard but rather with models and formulas, and the goal is not to checkmate the opponent but to assert or to refute a logical property. In game-theoretic terms, the games that we deal with in this paper, together with the game of chess, fall under the category *zero-sum two-person games of perfect information*.

The game-approach in logic owes much of its intuitive appeal to the picture of playing the game as an interactive act. However, an 'act' is not a mathematical concept. In order to reason rigorously about games, it is important to formulate the relevant notions about games in a mathematically precise way. In this section we give an overview of the mathematical details of zero-sum two-person games of perfect information. This will give us a uniform framework to deal with the various games in this paper.

Let us fix two players **I** and **II**. As a convention we always refer to player **I** as 'he' and player **II** as 'she'. Let A be an arbitrary set. We may compare it to the set of all possible configurations on a chess board. We also introduce a measure for the length of a game. For the moment, let a natural number $n \in \mathbb{N}$ be such a measure. A *play* of one of the players is any sequence $x = (x_0, \ldots, x_{n-1})$ of elements of A. A sequence

$$(\bar{x}, \bar{y}) = (x_0, y_0, \ldots, x_{n-1}, y_{n-1})$$

of elements of A is called a *play* of the game.

Let W be a subset of A^{2n}. The game $\mathcal{G}_n(A, W)$ is defined as follows:

Player **II** wins the play $(\bar{x}, \bar{y})$ if $(x_0, y_0, \ldots, x_{n-1}, y_{n-1}) \in W$.
Player **I** wins the play $(\bar{x}, \bar{y})$ if $(x_0, y_0, \ldots, x_{n-1}, y_{n-1}) \notin W$.

Note that in the set W we incorporate both the 'rules' and the 'winning conditions' of the game in the ordinary sense. If a player does not, or could not play according to the rules, then he or she loses immediately.

A *strategy* of player **I** is a sequence

$$\sigma = (\sigma_0, \ldots, \sigma_{n-1})$$

of functions $\sigma_i : A^i \to A$. We say that player **I** has *used the strategy* σ *in the play* $(\bar{x}, \bar{y})$ if

$$x_0 = \sigma_0$$

and for all $0 < i < n$:

$$x_i = \sigma_i(y_0, \ldots, y_{n-1}).$$

The strategy σ of player **I** is a *winning strategy* if every play where **I** has used σ is a win for him. The concept of strategy and winning strategy can be defined similarly for player **II**.

A *position* of the game $\mathcal{G}_n(A, W)$ is any initial segment

$$p = (x_0, y_0, \ldots, x_{i-1}, y_{i-1})$$

of a play $(\bar{x}, \bar{y})$, where $i \leq n$. A *strategy for player **I** in position* $p = (x_0, y_0, \ldots, x_{i-1}, y_{i-1})$ is a sequence

$$\sigma = (\sigma_0, \ldots, \sigma_{n-i-1})$$

of functions $\sigma_j : A^j \to A$. The concept of *using the strategy* σ *after position* p and *winning strategy in position* p are defined analogously. We say that p is a *winning position* for player **I** if he has a winning strategy at this position.

One characteristic feature of the game $\mathcal{G}_n(A, W)$ is that at the end of the play there is always exactly one winner. There is no room for a draw. The next theorem tells us that, moreover, the winner is already determined at the beginning of the game.

Definition 1. A game is called *determined* if one of the players has a winning strategy. Otherwise the game is *non-determined.*

Theorem 6 (Zermelo)**.** *If A is any set, n a natural number, and $W \subseteq A^{2n}$, then the game $\mathcal{G}_n(A, W)$ is determined.*

To prove this theorem we need the following lemma.

Lemma 7 (Survival Lemma)**.** *Consider the game $\mathcal{G}_n(A, W)$. If player **I** does not have a winning strategy in position $p = (x_0, y_0, \ldots, x_{i-1}, y_{i-1})$, then for every $x_i \in A$ there is $y_i \in A$ such that player **I** does not have a winning strategy in position $p' = (x_0, y_0, \ldots, x_i, y_i)$.*

Proof. The main idea of the proof is backward induction. Suppose to the contrary that in position p, there exists $x_i \in A$ such that for any $y \in A$, player **I** has a winning strategy σ_y in the remaining game. We claim that player **I** has a winning strategy in p: he first plays x_i, and then depending on which $y \in A$ player **II** plays, he chooses to follow the strategy σ_y in the rest of the game. Clearly this strategy will win the game for him. This is a contradiction with the assumption that player **I** does not have a winning strategy in position p. □

Proof of the theorem. Suppose player **I** does not have a winning strategy at the beginning. Then player **II** makes sure she stays in the game by repeated use of the Survival Lemma. After n rounds the game terminates and player **I** still does not have a winning strategy. This means that player **II** has won the game and player **I** has lost. What we have described above is a winning strategy for **II**. □

Our interest is not limited to games with finitely many rounds. Consider the variant of the game $\mathcal{G}_n(A, W)$ where the measure for the length of game is not a natural number but the first infinite ordinal ω. This game goes on for infinitely many rounds. The infinite game $\mathcal{G}_\omega(A, W)$ consists of a set A and a set $W \subseteq A^{\mathbb{N}}$ of infinite sequences of elements of A.

A *play* of the game $\mathcal{G}_\omega(A, W)$ is an infinite sequence:

$$(\bar{x}, \bar{y}) = (x_0, y_0, x_1, y_1, \ldots).$$

A *play* of one of the players is likewise any infinite sequence $\bar{x} = (x_0, x_1, \ldots)$ of elements of A.

The play $(\bar{x}, \bar{y})$ is a *win* for player **II** if

$$(x_0, y_0, x_1, y_1, \ldots) \in W$$

and otherwise a win for player **I**.

A *position* in the game $\mathcal{G}_\omega(A, W)$ is any initial segment

$$p = (x_0, y_0, \ldots, x_{i-1}, y_{i-1})$$

of a play $(\bar{x}, \bar{y})$. The notions of strategy, winning strategy, using a strategy etc. are likewise defined for this game.

Given an infinite game $\mathcal{G}_\omega(A, W)$, we can restrict it to an n-round game $\mathcal{G}_n(A, W)$ by declaring that a play in $\mathcal{G}_n(A, W)$ is a win for player **I** if any infinite sequence extending it is a win for him in the game $\mathcal{G}_\omega(A, W)$. In a sense, we can think of the game $\mathcal{G}_\omega(A, W)$ as the limit of the finite games $\mathcal{G}_n(A, W)$ as n goes to infinity. In general, the outcome of the finite games does not tell us much about the outcome of the infinite game—it can happen that player **I** has a winning strategy in none of the games $\mathcal{G}_n(A, W)$, yet he does have a winning strategy in $\mathcal{G}_\omega(A, W)$. However there is an important class of games where the outcome of the finite games tell us all we need about the infinite game. They are what we call *closed* games and *open* games.

A subset W of $A^{\mathbb{N}}$ is *open*[2] if

$$(x_0, y_0, x_1, y_1, \ldots) \in W$$

implies the existence of $n \in \mathbb{N}$ such that

$$(x_0, y_0, \ldots, x_{n-1}, y_{n-1}, x'_n, y'_n, x'_{n+1}, y'_{n+1}, \ldots) \in W$$

for all $x'_n, y'_n, x'_{n+1}, y'_{n+1}, \ldots \in A$. A subset W is *closed* if its complement is open. We say that the game $\mathcal{G}_\omega(A, W)$ is *closed* (or *open*) if the set W is. If player **I** wins a play $(\bar{x}, \bar{y})$ in a closed game $\mathcal{G}_\omega(A, W)$ he must have essentially

[2]The collection of open sets is a topology on $A^{\mathbb{N}}$, hence the name.

already won it at a finite stage, in the sense that in such a play if I manages to keep playing smartly up to a certain point, then whatever he does afterwards will not spoil his victory.

In this paper we are particularly interested in closed games. An important reason is that determinacy results for finite games carry over to closed (and open) infinite games.

Lemma 8 (Infinite Survival Lemma)**.** *Consider the game $\mathcal{G}_\omega(A, W)$. If player **I** does not have a winning strategy in position $p = (x_0, y_0, \ldots, x_{i-1}, y_{i-1})$, then for every $x_i \in A$ there is $y_i \in A$ such that player **I** does not have a winning strategy in position $p' = (x_0, y_0, \ldots, x_i, y_i)$.*

Proof. The proof is virtually the same as that of the finite Survival Lemma. □

With the Infinite Survival Lemma we can prove the following theorem of Gale and Stewart [7].

Theorem 9 (Gale-Stewart)**.** *If A is any set and $W \subseteq A^{\mathbb{N}}$ is open or closed, then the game $\mathcal{G}_\omega(A, W)$ is determined.*

Proof. Suppose W is closed and player **I** does not have a winning strategy at the beginning. We define the following strategy for player **II**. At each step she makes sure that player **I** does not have a winning strategy in the remaining game by repeated use of the Infinite Survival Lemma. In other words she plays by 'hanging on' in the game. Suppose

$$(\bar{x}, \bar{y}) = (x_0, y_0, x_1, y_1, \ldots)$$

is a play where player **II** has used this strategy. We claim that $(\bar{x}, \bar{y})$ is a win for **II**. Suppose otherwise. Since the game is closed, player **I** must have a winning strategy at some finite stage. More precisely, there exists a natural number $n \in \mathbb{N}$ such that

$$(x_0, y_0, \ldots, x_{n-1}, y_{n-1}, x'_n, y'_n, x'_{n+1}, y'_{n+1}, \ldots) \in W$$

for all $x'_n, y'_n, x'_{n+1}, y'_{n+1}, \ldots \in A$. This means that *any* strategy is a winning strategy for **I** in the position $(x_0, y_0, \ldots, x_{n-1}, y_{n-1})$. This is in contradiction with our assumption.

The proof for the case where W is open is similar. □

The games that we encounter in the rest of this paper are either $\mathcal{G}_n(A, W)$ or $\mathcal{G}_\omega(A, W)$ for some A and W. Sometimes we may present a game in a more intuitive way by describing its rules and winning conditions, without mentioning explicitly the sets A and W. However, we claim that such a game can always be formulated in this mathematically precise form if needed. The games in this paper are all determined.

2.2 The Infinitary Logic $L_{\omega_1\omega}$

Traditionally, symbolic logic has been associated with the study of finite formal expressions which are, at least in principle, capable of actually being written out in primitive notation on a piece of paper. The classical first order predicate logic is one eminent example. However, the fact that first order formulas can be identified with natural numbers (via 'Gödel numbering') and hence with finite *sets*, suggests the possibility of fashioning abstract languages some of whose formulas would be naturally identified as *infinite sets*. On the other hand, certain limitations of finitary logic call for the extension of logic with infinitary expressions. One main defect of first order logic is that it is usually inadequate for expressing mathematical concepts that involve explicitly or implicitly the notion of the infinite, which abound in mainstream mathematics.

Infinitary languages were first introduced in propositional calculus in Scott and Tarski [20] and in predicate logic in Tarski [22], both in the year 1958. Karp's 1964 book [13] is an early book on infinitary languages. In this book Karp introduced the following notation for infinitary languages. Let κ and λ be cardinals numbers. Let $L_{\kappa\lambda}$ denote the logic with conjunctions and disjunctions over index sets of cardinality less than κ, and quantification of sets of variables of cardinality less than λ. All of the logics that we need to consider for our purpose admit only formulas with finite strings of quantifiers, therefore $\lambda = \omega$ throughout this paper. In this notation $L_{\omega\omega}$ is just first order logic. We also denote by $L_{\infty\omega}$ the logic with conjunctions and disjunctions over arbitrary index sets.

In this paper we are mainly concerned with the logic $L_{\omega_1\omega}$, which adds to first order logic countable conjunctions and disjunctions of formulas. The first wave of results about the logic $L_{\omega_1\omega}$ came around the year 1963. The Completeness Theorem for $L_{\omega_1\omega}$, as well as for other infinitary languages (but in a weaker form), was proved by Karp [13] in 1964. The Interpolation Theorem for $L_{\omega_1\omega}$ was proved by Lopez-Escobar [17] in 1965 and Scott's Isomorphism Theorem for $L_{\omega_1\omega}$ by Scott [19] in 1965. In this section we give a quick survey of the results from these early pioneers of infinitary logic.

Let L be a vocabulary, and let $x_1, x_2, \ldots$ denote variables. The terms of $L_{\omega_1\omega}$ are defined in the usual way. The formulas of $L_{\omega_1\omega}$ are defined as follows.

Definition 2. We define the class of all formulas of $L_{\omega_1\omega}$ inductively.

1. If t_1 and t_2 are terms, then $t_1 = t_2$ is a formula.
2. If R is an n-place relational symbol and $t_1, \ldots, t_n$ are terms, then $Rt_1 \ldots t_n$ is a formula.
3. If ϕ is a formula, then so is $\neg\phi$.
4. If ϕ and ψ are formulas, then so is $\phi \vee \psi$.
5. If ϕ and ψ are formulas, then so is $\phi \wedge \psi$.

6. If I is a countable set and for every $i \in I$, ϕ_i is a formula, then so is $\bigvee_{i \in I} \phi_i$.
7. If I is a countable set and for every $i \in I$, ϕ_i is a formula, then so is $\bigwedge_{i \in I} \phi_i$.
8. If ϕ is a formula and x_n a variable, then $\exists x_n \phi$ is a formula.
9. If ϕ is a formula and x_n a variable, then $\forall x_n \phi$ is a formula.

We use $\rightarrow$ and $\leftrightarrow$ as abbreviations for implication and equivalence in the usual way. Substitution for L-formulas in $L_{\omega_1 \omega}$ is also defined as usual. The truth definition for $L_{\omega_1 \omega}$ is standard.

The logic $L_{\omega_1 \omega}$ has a complete proof calculus by adding to first order logic an introduction rule for infinite conjunction and disjunction. The following axiomatization is given essentially in Keisler's book [15], with slight modification. We first introduce some convenient notation. Given a formula ϕ of $L_{\omega_1 \omega}$, the formula $\phi\neg$, obtained by 'moving a negation inside', is defined inductively as follows:

1. If ϕ is atomic, then $\phi\neg$ is $\neg\phi$.
2. If ϕ is $(\neg\psi)$, then $\phi\neg$ is ψ.
3. If ϕ is $\psi \vee \theta$, then $\phi\neg$ is $\neg\psi \wedge \neg\theta$.
4. If ϕ is $\psi \wedge \theta$, then $\phi\neg$ is $\neg\psi \vee \neg\theta$.
5. If ϕ is $\bigvee_{i \in \omega} \phi_i$, then $\phi\neg$ is $\bigwedge_{i \in \omega} \neg\phi_i$.
6. If ϕ is $\bigwedge_{i \in \omega} \phi_i$, then $\phi\neg$ is $\bigvee_{i \in \omega} \neg\phi_i$.
7. If ϕ is $\forall x_n \psi$, then $\phi\neg$ is $\exists x_n \neg\psi$.
8. If ϕ is $\exists x_n \psi$, then $\phi\neg$ is $\forall x_n \neg\psi$.

LOGICAL AXIOMS FOR $L_{\omega_1 \omega}$:

1. Every instance of a tautology of finitary propositional logic is an axiom.
2. $(\neg\phi) \leftrightarrow (\phi\neg)$.
3. $\phi_1 \wedge \phi_2 \rightarrow \phi_i$, $i = 1, 2$.
4. $\bigwedge_{i \in \omega} \phi_i \rightarrow \phi_n$, $n \in \omega$.
5. $\forall x \phi(x, \bar{y}) \rightarrow \phi(t, \bar{y})$, where t is a term.
6. $x = x$.
7. $x = y \rightarrow y = x$.
8. $\phi(x, \bar{y}) \wedge t = x \rightarrow \phi(t, \bar{y})$.

INFERENCE RULES FOR $L_{\omega_1\omega}$:

1. From $\psi, \psi \to \phi$, infer ϕ.
2. From $\psi \to \phi(x, \bar{y})$, infer $\psi \to \forall x \phi(x, \bar{y})$, where x does not occur free in ϕ.
3. From the sequence $\psi \to \phi_i, i \in \omega$, infer $\psi \to \bigwedge_{i\in\omega} \phi_i$.

As a consequence of the last rule, the notion of a *proof* in $L_{\omega_1\omega}$ has to be generalized to include proofs of countable length. We say that a formula ϕ is a *theorem* of $L_{\omega_1\omega}$, denoted by $\vdash_{L_{\omega_1\omega}} \phi$, if there exists a countable sequence of formulas $\langle \phi_\alpha | \alpha \leq \beta \rangle$ such that $\phi_\beta = \phi$ and for each $\alpha \leq \beta$, ϕ_α is either an axiom of $L_{\omega_1\omega}$ or is inferred from earlier formulas $\phi_\gamma, \gamma < \alpha$ by an inference rule. It is easy to see that the set of theorems of $L_{\omega_1\omega}$ is the least set of formulas in $L_{\omega_1\omega}$ which contains all the axioms and is closed under the inference rules. Denote by $\vDash \phi$ the semantic validity of ϕ as usual. The following completeness result is due to Karp [13].

Theorem 10 (The Completeness Theorem for $L_{\omega_1\omega}$)**.** *If ϕ is a sentence of $L_{\omega_1\omega}$, then $\vdash_{L_{\omega_1\omega}} \phi$ if and only if $\vDash \phi$.*

In the logic $L_{\omega_1\omega}$ we are able to express many familiar mathematical concepts which are undefinable in first order logic.

Example 1. Let L be the language of abelian groups. An abelian group G is a *torsion group* if and only if

$$G \models \forall x \bigvee_{n\in\omega} \underbrace{x + \ldots + x}_{n} = 0.$$

Example 2. Let L be the language of graphs, with relation symbol E denoting the edge relation. Let [3]

$$\phi_0(x_0, x_1) = (x_0 = x_1)$$
$$\phi_{n+1}(x_0, x_1) = \exists x_2 \Big(x_0 E x_2 \wedge \exists x_0 \big(x_0 = x_2 \wedge \phi_n(x_0, x_1) \big) \Big).$$

A graph G is *connected* if and only if

$$G \models \forall x_0 \forall x_1 \Big(\bigvee_{n\in\omega} \phi_n(x_0, x_1) \Big).$$

[3]Note that ϕ_n is written in this way so that we use only three variables x_0, x_1 and x_2. If we do not care how many variables are used, we can also define ϕ_n inductively as follows

$$\phi'_{n+1}(x_0, x_1) = \exists x_{n+2} \big(x_0 E x_{n+2} \wedge \phi_n(x_{n+2}, x_1) \big).$$

Example 3. Let L be the language of linear orders. For any ordinal α, well-ordering of type α is definable by a sentence in $L_{\infty\omega}$. When α is a countable ordinal this sentence is in $L_{\omega_1\omega}$. Let

$$\theta_0(x_0) = \neg\exists x_1(x_1 < x_0)$$
$$\theta_\alpha(x_0) = \forall x_1\Big(x_1 < x_0 \leftrightarrow \exists x_0\big(x_0 = x_1 \wedge \bigvee_{\beta<\alpha} \theta_\beta(x_0)\big)\Big).$$

In a linear order $(M, <)$, the formula $\theta_\alpha(x_0)$ has the intended meaning that the initial segment $\{x \in \mathcal{M} | x < x_0\}$ has order type α. Let

$$\phi_\alpha = \Big(\forall x_0 \bigvee_{\beta<\alpha} \theta_\beta(x_0)\Big) \wedge \Big(\bigwedge_{\beta<\alpha} \exists x_0 \theta_\beta(x_0)\Big).$$

It can be proved by transfinite induction on α that for any linear order $(M, <)$, we have

$$(M, <) \models \phi_\alpha \Longleftrightarrow (M, <) \cong (\alpha, <).$$

Note that when α is a countable ordinal, the sentence ϕ_α is in $L_{\omega_1\omega}$. This is not the case when α is uncountable.

A countable well-ordering α can be characterized up to isomorphism by the $L_{\omega_1\omega}$ sentence ϕ_α. The following theorem of Scott [19] shows that this result can be generalized to every countable structure.

Theorem 11 (Scott Isomorphism Theorem)**.** *Let L be a countable vocabulary and $\mathcal{M}$ be a countable L-structure. There exists an L-sentence $\sigma_\mathcal{M}$ in $L_{\omega_1\omega}$, called the Scott sentence of $\mathcal{M}$, such that for every countable L-structure $\mathcal{M}'$*

$$\mathcal{M}' \models \sigma_\mathcal{M} \Longleftrightarrow \mathcal{M}' \cong \mathcal{M}.$$

In the above theorem the requirement of L being countable cannot be dispensed with. It is not reasonable to hope that the theorem remains true when L is an uncountable vocabulary. For an almost trivial example, consider the language $L_0 = \langle P_\alpha, \alpha < \omega_1\rangle$ with uncountably many predicate symbols. Let $\mathcal{M}$ be an L_0 structure with domain the singleton set $\{\emptyset\}$. We claim that there is no $L_{\omega_1\omega}$ sentence ϕ that characterizes $\mathcal{M}$ up to isomorphism. Let ϕ be any L-sentence in $L_{\omega_1\omega}$ that is true in $\mathcal{M}$. It is easy to see that ϕ only involves countably many predicate symbols. Let $\alpha_0 < \omega_1$ be such that P_{α_0} does not appear in ϕ. Let $\mathcal{M}'$ be the L_0 structure with domain $\{\emptyset\}$, and

$$P_\alpha^{\mathcal{M}'} = \begin{cases} P_\alpha^\mathcal{M} & \text{if } \alpha \neq \alpha_0, \\ \{\emptyset\} \setminus P_{\alpha_0}^\mathcal{M} & \text{if } \alpha = \alpha_0. \end{cases}$$

It is clear that $\mathcal{M}' \models \phi$, yet $\mathcal{M} \ncong \mathcal{M}'$.

Another important result in the model theory of $L_{\omega_1\omega}$ is the following theorem of Karp [14]. Let $\mathcal{M} = (M, <)$ and $\mathcal{M}' = (M', <')$ be partially ordered sets, their *product* $\mathcal{M} \times \mathcal{M}'$ is the partially ordered set $(M \times M', <^*)$ where

$$(x, x') < (y, y') \Longleftrightarrow x' <' y' \text{ or } (x' = y' \text{ and } x < y).$$

Considering an ordinal δ as the well-ordered set $(\delta, <)$, we have the following result.

Theorem 12 (Karp)**.** *Suppose an ordinal δ satisfies the condition*

$$\alpha < \delta \Longrightarrow \omega^\alpha < \delta \tag{1}$$

and $\mathcal{M}$ is any linear order with a first element. Then

$$\delta \equiv_\delta \delta \times \mathcal{M}.$$

This theorem can be used to prove the undefinability of well-ordering in $L_{\infty\omega}$ (for details, see [14]). Intuitively, the formula

$$\bigvee_{\alpha \in Ord} \phi_\alpha$$

defines the class of well orderings. The problem is that Ord is not a set but a proper class, therefore this sentence is not in $L_{\infty\omega}$. By the same technique we can prove that the class of countable well-orderings cannot be defined by an $L_{\omega_1\omega}$ sentence. We will come back to this topic in Section 4.

3 What Should an Ehrenfeucht-Fraïssé Game for $L_{\omega_1\omega}$ Look Like?

We may say that there are a broader meaning and a narrower meaning of the term 'Ehrenfeucht-Fraïssé Game'. The narrower notion of Ehrenfeucht-Fraïssé Game refers to a particular game in predicate logic. Depending on the length of the play, this game has three variants EF_n, EF_ω and EFD_α—they are what we call the 'usual' Ehrenfeucht-Fraïssé Games. The basic idea behind these games is the following. Given two structures $\mathcal{A}$ and $\mathcal{B}$, we want to decide how 'similar' these two structures are. For this purpose we invite two mathematicians **I** and **II** to argue over this issue. **I** claims that there is an intrinsic structural difference between the two structures. **II** claims that the two structures are essentially similar. In each round **I** tries to justify himself by calling one more element of one of the structures into question. **II** responds by picking an element from the other structure which is similar to the element that **I** has just picked. The trick of the game is that **I** and **II** only argue over a small piece of the pair of structures at a time. On the other hand, the broader notion of Ehrenfeucht-Fraïssé Game refers to a general technique to determine whether two structures are equivalent (also known as the back-and-forth technique).

In this section we survey Ehrenfeucht-Fraïssé Game in the narrower sense, namely the three usual Ehrenfeucht-Fraïssé Games. The games EF_n, EF_ω and EFD_α characterize a hierarchy of equivalence relations. They provide important information about the question when two structures can be separated in the logic $L_{\omega_1\omega}$. However, the usual Ehrenfeucht-Fraïssé Games do not thoroughly solve the question of separation for $L_{\omega_1\omega}$. Equivalence in the logic $L_{\omega_1\omega}$ does not properly identify with any level of the hierarchy. An analysis of the reason behind this failure will give us some clue what the 'right' Ehrenfeucht-Fraïssé Game for $L_{\omega_1\omega}$ should look like. The new game that we are going to introduce will be an Ehrenfeucht-Fraïssé Game in the broader sense: the game is played in the spirit of Ehrenfeucht-Fraïssé Game, but the mechanism of the game will be substantially different from the usual Ehrenfeucht-Fraïssé Games.

Definition 3. Suppose L is a relational vocabulary and $\mathcal{M}, \mathcal{M}'$ are L-structures. A partial mapping $\pi : M \to M'$ is a *partial isomorphism* if it is an isomorphism between $\mathcal{M} \restriction \mathrm{dom}(\pi)$ and $\mathcal{M}' \restriction \mathrm{ran}(\pi)$.

Definition 4. Suppose L is a relational vocabulary and $\mathcal{M}, \mathcal{M}'$ are L-structures. Suppose also that the domains M, M' of the two structures are disjoint. The *Ehrenfeucht-Fraïssé Game* $\mathrm{EF}_n(\mathcal{M}, \mathcal{M}')$ is the game $\mathcal{G}_n(M \cup M', W_n(\mathcal{M}, \mathcal{M}'))$, where $W_n(\mathcal{M}, \mathcal{M}') \subseteq (M \cup M')^{2n}$ is the set of

$$p = (x_0, y_0, \ldots, x_{n-1}, y_{n-1})$$

such that:

1. For all $i < n : x_i \in M \iff y_i \in M'$.
2. If we denote

$$v_i = \begin{cases} x_i & \text{if } x_i \in M \\ y_i & \text{if } y_i \in M \end{cases} \qquad v_i' = \begin{cases} x_i & \text{if } x_i \in M' \\ y_i & \text{if } y_i \in M', \end{cases}$$

   then

$$f_p = \{(v_0, v_0'), \ldots, (v_{n-1}, v_{n-1}')\}$$

   is a partial isomorphism $\mathcal{M} \to \mathcal{M}'$.

Definition 5. Suppose L is a relational vocabulary and $\mathcal{M}, \mathcal{M}'$ are L-structures such that $M \cap M' = \emptyset$. The infinite game $\mathrm{EF}_\omega(\mathcal{M}, \mathcal{M}')$ is the game $\mathcal{G}_\omega(M \cup M', W_\omega(\mathcal{M}, \mathcal{M}'))$, where $W_\omega(\mathcal{M}, \mathcal{M}')$ is the set of $p = (x_0, y_0, x_1, y_1, \ldots)$ such that for all $n \in \mathbb{N}$ we have $(x_0, y_0, \ldots, x_{n-1}, y_{n-1}) \in W_n(\mathcal{M}, \mathcal{M}')$. The set $W_n(\mathcal{M}, \mathcal{M}')$ is as defined in Definition 4.

Note that $\mathrm{EF}_\omega(\mathcal{M}, \mathcal{M}')$ is a closed game. The set $(M \cup M')^{\mathbb{N}} \setminus W_\omega(\mathcal{M}, \mathcal{M}')$ is open. For suppose $p = (x_0, y_0, x_1, y_1, \ldots) \notin W_\omega(\mathcal{M}, \mathcal{M}')$, then for some $n \in \mathbb{N}$ we have $(x_0, y_0, \ldots, x_{n-1}, y_{n-1}) \notin W_n(\mathcal{M}, \mathcal{M}')$. It follows that

$$(x_0, y_0, \ldots, x_{n-1}, y_{n-1}, x_n', y_n', x_{n+1}', y_{n+1}', \ldots) \notin W_\omega(\mathcal{M}, \mathcal{M}')$$

for all $x_n', y_n', x_{n+1}', y_{n+1}', \ldots \in M \cup M'$. By the Gale-Stewart Theorem, the game $\mathrm{EF}_\omega(\mathcal{M}, \mathcal{M}')$ is determined.

Definition 6. Let L be a relational vocabulary and $\mathcal{M}, \mathcal{M}'$ L-structures such that $M \cap M' = \emptyset$. Let α be an ordinal. The *Dynamic Ehrenfeucht-Fraïssé Game*

$\mathrm{EFD}_\alpha(\mathcal{M},\mathcal{M}')$ is the game $\mathcal{G}_\omega(M\times\alpha\cup M', W_{\omega,\alpha}(\mathcal{M},\mathcal{M}'))$, where $W_{\omega,\alpha}(\mathcal{M},\mathcal{M}')$ is the set of

$$p = ((x_0,\alpha_0), y_0, \ldots, (x_{n-1},\alpha_{n-1}), y_{n-1})$$

such that:

1. For all $i < n : x_i \in M \iff y_i \in M'$.
2. $\alpha > \alpha_0 > \ldots > \alpha_{n-1} = 0$.
3. If we denote

$$v_i = \begin{cases} x_i & \text{if } x_i \in M \\ y_i & \text{if } y_i \in M \end{cases} \qquad v'_i = \begin{cases} x_i & \text{if } x_i \in M' \\ y_i & \text{if } y_i \in M', \end{cases}$$

   then

$$f_p = \{(v_0, v'_0), \ldots, (v_{n-1}, v'_{n-1})\}$$

   is a partial isomorphism $\mathcal{M} \to \mathcal{M}'$.

Note the slight abuse of notation in the above definition. Strictly speaking a play in the game $\mathcal{G}_\omega(M\cup M'\cup\alpha, W_{\omega,\alpha}(\mathcal{M},\mathcal{M}'))$ should be an infinite sequence. Here we may regard

$$p = ((x_0,\alpha_0), y_0, \ldots, (x_{n-1},\alpha_{n-1}), y_{n-1})$$

as representing the infinite sequence beginning with p and followed by constant elements and 0's

$$((x_0,\alpha_0), y_0, \ldots, (x_{n-1},\alpha_{n-1}), y_{n-1}, (x_{n-1}, 0), y_{n-1}, \ldots).$$

In other words, we think of the moves after the first n rounds in the play p as dummy moves. They make no real difference to the outcome of the game. The game $\mathrm{EFD}_\alpha(\mathcal{M},\mathcal{M}')$ is closed.

The three games EF_n, EF_ω and EFD_α are played according to essentially the same set of rules. In each round player **I** picks an element from $\mathcal{M}$ or $\mathcal{M}'$. Player **II** responds with an element in the other structure. The rule is that **II** has to make sure all the finite segments of the play stand in partial isomorphisms. If **II** fails to meet the requirement, she loses. If she manages to meet the requirement until the game ends, she wins.

The difference among the three games lies in the way that the length of the game is measured. The game EF_n lasts for n rounds. The game EF_ω lasts for ω rounds. The case for the game EFD_α is more subtle. The length of the game is measured by an ordinal α. Each position

$$p = ((x_0,\alpha_0), y_0, \ldots, (x_{n-1},\alpha_{n-1}), y_{n-1})$$

in the game is associated with an ordinal α_{n-1}. We call it the *rank* of the position. In each round the rank of the position goes down. But it is up to player **I** to decide how the rank goes down. He can change his mind as the game proceeds, hence the name 'dynamic'. In each move, **I** picks in addition to an element an ordinal that is smaller than the current rank. We may think of the rank as the 'resources' that player **I** possesses. To make every move he has to 'spend' some resources. His objective is to win the game, namely to force **II** into a position where she is unable to continue, before he uses up his resources. Note that the game EFD_α is *not* a game of length α.

When α is a finite ordinal n, the game $\mathrm{EFD}_n(\mathcal{M}, \mathcal{M}')$ is essentially the same as the game $\mathrm{EF}_n(\mathcal{M}, \mathcal{M}')$. A moment's reflection shows that **I** has a winning strategy in $\mathrm{EFD}_n(\mathcal{M}, \mathcal{M}')$ if and only if he has a winning strategy in $\mathrm{EF}_n(\mathcal{M}, \mathcal{M}')$. Let us now consider the case where α is an infinite ordinal. Since the ordinals are well-founded, every play of the game is finitely long. However there is no finite upper-bound to the length of the plays. In this sense we say that this game is *potentially infinite*. The game $\mathrm{EFD}_\alpha(\mathcal{M}, \mathcal{M}')$ is easier for **I** to play than any $\mathrm{EF}_n(\mathcal{M}, \mathcal{M}')$ (when α is infinite), but still harder than $\mathrm{EF}_\omega(\mathcal{M}, \mathcal{M}')$.

The following are some straightforward properties of the game EFD_α.

Lemma 13. *If player **II** has a winning strategy in* $\mathrm{EFD}_\alpha(\mathcal{M}, \mathcal{M}')$ *and* $\beta \leq \alpha$, *then she has a winning strategy in* $\mathrm{EFD}_\beta(\mathcal{M}, \mathcal{M}')$.

Proof. **II** pretends to be playing the longer game $\mathrm{EFD}_\alpha(\mathcal{M}, \mathcal{M}')$. The winning strategy in this game would also win the shorter game $\mathrm{EFD}_\beta(\mathcal{M}, \mathcal{M}')$ for her. □

Lemma 14. *If α is a limit ordinal and player **II** has a winning strategy in the game* $\mathrm{EFD}_\beta(\mathcal{M}, \mathcal{M}')$ *for each $\beta < \alpha$, then **II** has a wining strategy in the game* $\mathrm{EFD}_\alpha(\mathcal{M}, \mathcal{M}')$.

Proof. Suppose **I** plays $\alpha_0 < \alpha$ in his opening move. Now **II** pretends that they are actually playing the game $\mathrm{EFD}_{\alpha_0+1}(\mathcal{M}, \mathcal{M}')$. By assumption she has a winning strategy in this game. Since α_0 is arbitrary, player **II** has a winning strategy in the game $\mathrm{EFD}_\alpha(\mathcal{M}, \mathcal{M}')$. □

In particular, **II** has a winning strategy in $\mathrm{EFD}_\omega(\mathcal{M}, \mathcal{M}')$ if and only if she has a winning strategy in every finite game $\mathrm{EFD}_n(\mathcal{M}, \mathcal{M}')$, $n \in \mathbb{N}$.

Definition 7. The *quantifier rank*, denoted by $\mathrm{qr}(\phi)$, of a formula ϕ in $L_{\omega_1\omega}$ is defined inductively as follows:

$$\begin{aligned}
\mathrm{qr}(\phi) &= 0 \text{ if } \phi \text{ is atomic} \\
\mathrm{qr}(\neg\phi) &= \mathrm{qr}(\phi) \\
\mathrm{qr}(\phi \wedge \psi) &= \max(\mathrm{qr}(\phi), \mathrm{qr}(\psi)) \\
\mathrm{qr}(\phi \vee \psi) &= \max(\mathrm{qr}(\phi), \mathrm{qr}(\psi)) \\
\mathrm{qr}(\bigwedge_{i\in\omega} \phi_i) &= \sup_{i\in\omega} \mathrm{qr}(\phi_i) \\
\mathrm{qr}(\bigvee_{i\in\omega} \phi_i) &= \sup_{i\in\omega} \mathrm{qr}(\phi_i) \\
\mathrm{qr}(\exists x\phi) &= \mathrm{qr}(\phi) + 1 \\
\mathrm{qr}(\forall x\phi) &= \mathrm{qr}(\phi) + 1.
\end{aligned} \tag{2}$$

Recall that the game $\mathrm{EF}_n(\mathcal{M}, \mathcal{M}')$ characterizes equivalence in first order logic (Theorem 3). The games EF_ω and EFD_α characterize equivalence in the infinitary logic $L_{\infty\omega}$, the extension of first order logic with arbitrary conjunctions and disjunctions. Let L be a vocabulary. Suppose $\mathcal{A}$ and $\mathcal{B}$ are L-structures. Let $\mathcal{A} \equiv_\alpha \mathcal{B}$ denote that $\mathcal{A}$ and $\mathcal{B}$ satisfy the same L-sentences in $L_{\infty\omega}$ up to quantifier rank α. Let $\mathcal{A} \equiv_{\infty\omega} \mathcal{B}$ denote that $\mathcal{A}$ and $\mathcal{B}$ satisfy the same L-sentences in $L_{\infty\omega}$. The game EF_ω gives a 'global' characterization of equivalence in $L_{\infty\omega}$:

Theorem 15 (Karp)**.** *Let L be a relational vocabulary and $\mathcal{A}, \mathcal{B}$ L-structures. Then the following are equivalent:*

1. $\mathcal{A} \equiv_{\infty\omega} \mathcal{B}$.
2. *Player **II** has a winning strategy in* $\mathrm{EF}_\omega(\mathcal{A}, \mathcal{B})$.

Proof. ($1 \Longrightarrow 2$) Suppose $\mathcal{A} \equiv_{\infty\omega} \mathcal{B}$. We describe the following strategy for player **II**. In each round, she makes sure that the position

$$(\bar{x}, \bar{y}) = (x_0, y_0, \ldots, x_{n-1}, y_{n-1})$$

satisfies

$$(\mathcal{A}, \bar{x}) \equiv_{\infty\omega} (\mathcal{B}, \bar{y}). \tag{3}$$

If this strategy is attainable, it is a winning strategy for **II**: condition (2) guarantees that $\bar{x} \mapsto \bar{y}$ is a partial isomorphism from $\mathcal{A}$ to $\mathcal{B}$. It remains to show that this strategy can indeed be attained by player **II**.

We prove it by induction on n. Suppose the game is in position $(\bar{a}, \bar{b})$ and condition (3) is satisfied:

$$(\mathcal{A}, \bar{a}) \equiv_{\infty\omega} (\mathcal{B}, \bar{b}). \tag{4}$$

Suppose **I** plays $c \in \mathcal{A}$ for the next move. **II** wants to find an element $d \in \mathcal{B}$ such that

$$(\mathcal{A}, \bar{a}c) \equiv_{\infty\omega} (\mathcal{B}, \bar{b}d). \tag{5}$$

Suppose to the contrary that there is no such element. Then for every element $d \in \mathcal{B}$, there is an L-formula $\phi_d(\bar{x}, y)$ in $L_{\infty\omega}$ such that

$$\mathcal{A} \models \phi_d(\bar{a}, c)$$

$$\mathcal{B} \models \neg\phi_d(\bar{b}, d).$$

Take the conjunction $\Phi(\bar{x}, y) = \bigwedge_{d\in\mathcal{B}} \phi_d(\bar{x}, y)$. It is a formula in $L_{\infty\omega}$. We have

$$\mathcal{A} \models \exists y \Phi(\bar{a}, y) \tag{6}$$

$$\mathcal{B} \models \neg\exists y \Phi(\bar{b}, y). \tag{7}$$

A contradiction with condition (4). Therefore there must be an element $d \in \mathcal{B}$ that satisfies (4). Let **II** play that element. Thus she has managed to maintain condition (3) for one more round. The case where **I** plays an element $c \in \mathcal{B}$ is similar.

(2$\Longrightarrow$1) We claim that if $(\bar{a}, \bar{b})$ is a winning position for **II** in the game $\mathrm{EF}_\omega(\mathcal{A}, \mathcal{B})$, then $(\mathcal{A}, \bar{a}) \equiv_{\infty\omega} (\mathcal{B}, \bar{b})$. We prove the claim by induction on the complexity of the formula. The atomic case and the boolean cases are straightforward. The only non-trivial case is with the quantifiers.

Suppose the claim is true for formulas of quantifier rank up to α. Let $\phi(\bar{x}) = \exists y \psi(\bar{x}, y)$, where $\mathrm{qr}(\psi) = \alpha$ and $\mathcal{A} \models \phi(\bar{a})$. Then there is $c \in \mathcal{A}$ such that

$$\mathcal{A} \models \psi(\bar{a}, c).$$

Imagine **I** continues the game from position $(\bar{a}, \bar{b})$ by playing $c \in \mathcal{A}$. By assumption **II** has a winning strategy in the remaining game. Suppose the strategy tells her to play $d \in \mathcal{B}$. Then $(\bar{a}c, \bar{b}d)$ is also a winning position for **II**. By the induction hypothesis, we have

$$\mathcal{B} \models \psi(\bar{b}, d).$$

Therefore

$$\mathcal{B} \models \exists y \psi(\bar{b}, y).$$

And hence the claim. □

The game EFD_α gives a 'local' characterization of equivalence in $L_{\infty\omega}$:

Theorem 16 (Karp)**.** *Let L be a relational vocabulary, $\mathcal{A}$ and $\mathcal{B}$ be L-structures and α be an ordinal number. Then the following are equivalent:*

1. $\mathcal{A} \equiv_\alpha \mathcal{B}$.
2. *Player **II** has a winning strategy in* $\mathrm{EFD}_\alpha(\mathcal{A}, \mathcal{B})$.

Proof. The proof of this theorem is very similar to the previous one. For both directions we prove the assertion by induction on α. For details, the reader is referred to Section 7.3 in [23]. □

The game $\mathrm{EFD}_{\omega_1}(\mathcal{A},\mathcal{B})$ is of particular interest to us. Note that **II** has a winning strategy in this game if and only if she has a winning strategy in $\mathrm{EFD}_{\alpha}(\mathcal{A},\mathcal{B})$ for all countable ordinals $\alpha < \omega_1$. Observe the following connection between $L_{\infty\omega}$ and $L_{\omega_1\omega}$.

Proposition 17. *Let L be a relational vocabulary. For any L-formula $\phi \in L_{\infty\omega}$, if $\phi \in L_{\omega_1\omega}$ then $\mathrm{qr}(\phi) < \omega_1$.*

Proof. ω_1 is closed under successor and countable union. If $\phi \in L_{\omega_1\omega}$ then the quantifier rank of ϕ is bounded to be a countable ordinal. □

Combined with this fact, the game $\mathrm{EFD}_{\omega_1}(\mathcal{A},\mathcal{B})$ gives us important information about the question of separation in the language $L_{\omega_1\omega}$.

Proposition 18. *Let L be a relational vocabulary and $\mathcal{A},\mathcal{B}$ L-structures. If player **II** has a winning strategy in $\mathrm{EFD}_{\omega_1}(\mathcal{A},\mathcal{B})$, then there is no L-formula $\phi \in L_{\omega_1\omega}$ separating $\mathcal{A}$ and $\mathcal{B}$.*

Proof. If there is an $L_{\omega_1\omega}$ formula separating $\mathcal{A}$ and $\mathcal{B}$, it would be of countable quantifier rank. Since player **II** has a winning strategy in $\mathrm{EFD}_{\omega_1}(\mathcal{A},\mathcal{B})$, she wins $\mathrm{EFD}_{\alpha}(\mathcal{A},\mathcal{B})$ for any countable ordinal α. Hence there cannot be a formula of countable quantifier rank separating $\mathcal{A}$ and $\mathcal{B}$. A contradiction. □

A good illustration of the use of this proposition is the following example in [16].

Example 4. Let κ and λ be two uncountable cardinals (viewed as linear orders). Then we have

1. $\kappa \not\equiv_{\infty\omega} \lambda$.
2. $\kappa \equiv_{\omega_1\omega} \lambda$.

To prove that 1 is true, it suffices to consider the Scott sentence σ_κ of κ[4]. It is an $L_{\infty\omega}$ sentence. It is easy to see that $\kappa \models \sigma_\kappa$ and $\lambda \not\models \sigma_\kappa$. As for 2, we resort to Karp's result in Theorem 12. Without loss of generality, assume that $\kappa > \lambda$. The uncountable cardinal κ satisfies condition (1) in the premise of the theorem. Moreover, we have that $\lambda \times \kappa = \kappa$. Therefore by Theorem 12, we have

$$\lambda \equiv_\lambda \lambda \times \kappa \equiv_\lambda \kappa.$$

Hence in particular

$$\lambda \equiv_{\omega_1} \kappa.$$

It follows from Proposition 18 that there is no $L_{\omega_1\omega}$ sentence separating κ and λ.

This example shows that the question of separation in $L_{\omega_1\omega}$ is genuinely different from separation in $L_{\infty\omega}$. Given two uncountable cardinals, there is an $L_{\infty\omega}$ sentence separating the two structures. No such sentence exists in $L_{\omega_1\omega}$.

[4]For the Scott sentence of uncountable structures, see Chapter 7 in [23]

Example 5. For another example, consider countable well-orderings. Recall that well-ordering of type α is definable by a sentence θ_α (Example 3). When α is a countable ordinal, θ_α is an $L_{\omega_1\omega}$ sentence. Therefore the class $\mathcal{C}$ of all countable well-orderings is definable by the following sentence:

$$\theta = \bigvee_{\alpha<\omega_1} \theta_\alpha.$$

However, the sentence θ has an uncountable disjunction and is therefore not in $L_{\omega_1\omega}$. A natural question is: Is there an $L_{\omega_1\omega}$ sentence defining the class $\mathcal{C}$ of all countable ordinals?

The answer is negative. Suppose to the contrary that there is an $L_{\omega_1\omega}$ sentence ϕ defining $\mathcal{C}$. Let $\mathrm{qr}(\phi) = \alpha$. Let $\gamma = \sup_{n\in\omega} \alpha_n$, where $\alpha_0 = \alpha$ and $\alpha_{n+1} = \omega^{\alpha_n}$ for $n \geq 0$. Since ω_1 is regular, $\gamma < \omega_1$. The ordinal γ satisfies condition (1). Hence by Theorem 12 , we have that

$$\gamma \equiv_\gamma \gamma \times \omega_1.$$

Since the quantifier rank of ϕ is less than γ, we have that

$$\gamma \times \omega_1 \models \phi.$$

But $\gamma \times \omega_1$ is not a countable well-ordering.

The usual Ehrenfeucht-Fraïssé Games EF_n, EF_ω and EFD_α provide in a sense lower bounds and upper bounds for the question of separation in $L_{\omega_1\omega}$. Consider two structures $\mathcal{A}$ and $\mathcal{B}$. If $\mathcal{A} \not\equiv \mathcal{B}$, then $\mathcal{A}$ is 'too unsimilar' to $\mathcal{B}$ and it is not interesting from the point of view of $L_{\omega_1\omega}$. If $\mathcal{A} \equiv_{\infty\omega} \mathcal{B}$, then $\mathcal{A}$ is 'too similar' to $\mathcal{B}$ and it is not interesting either. The above result tells us that if $\mathcal{A} \equiv_{\omega_1} \mathcal{B}$, then there is no $L_{\omega_1\omega}$ formula separating $\mathcal{A}$ and $\mathcal{B}$. Therefore all the excitement concentrates in the interval between $\mathcal{A} \equiv \mathcal{B}$ and $\mathcal{A} \equiv_{\omega_1} \mathcal{B}$. In other words, we are mainly interested in the case where the *Scott watershed*[5] of $\mathcal{A}$ and $\mathcal{B}$ is a countable ordinal.

When we are lucky enough, the usual Ehrenfeucht-Fraïssé Games solve the question of separation in $L_{\omega_1\omega}$, as we have seen in the two examples above. However this is not the case in general. The usual Ehrenfeucht-Fraïssé Games are games for quantifiers. A winning strategy of **I** corresponds to a separation formula. The games give information about the *quantifier rank* of the separation formula, but they do not directly give information about the 'boolean size' of the formula. Therefore we may end up with knowing the existence of a formula in $L_{\infty\omega}$ separating two structures, possibly up to a certain quantifier rank, but with EF and EFD we may have no access to knowing whether this formula is in $L_{\omega_1\omega}$. In Example 4 and Example 5 we are lucky in the sense that the game EFD restricted to countable ranks brings a constraint on the boolean size of a separation formula *via* a constraint on its quantifier rank. But the problem is that there are $L_{\infty\omega}$ sentences with a low quantifier rank and a rather

[5]For the definition of Scott watershed, see page 147 in [23].

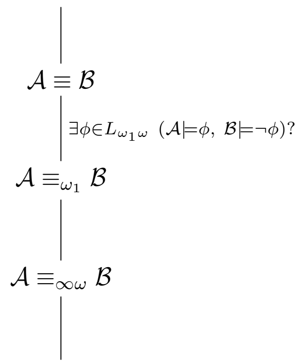

Figure 3: Hierarchy of equivalence relations.

big boolean size. In particular, the usual Ehrenfeucht-Fraïssé Games do not answer the question whether two structures can be separated by a *quantifier-free* $L_{\omega_1\omega}$ formula, yet elimination of quantifiers in favor of connectives is the oldest application of infinitary logic[6].

Example 6. Let $L = \langle P, \{c_n | n \in \mathbb{N}\}\rangle$ be the language with one predicate symbol and countably many constant symbols. For any subset $A \subseteq \mathbb{N}$, let $(\mathbb{N}, A)$ denote the L-structure with P interpreted as A and constant symbol c_n interpreted as the natural number n. Let $\mathcal{A} = \{(\mathbb{N}, A) | A \subseteq \mathbb{N}\}$ be the class of such structures.

Let

$$\phi_A = \bigwedge_{n \in \omega} P_A^n$$

where

$$P_A^n = \begin{cases} P(c_n) & \text{if } n \in A, \\ \neg P(c_n) & \text{if } n \notin A. \end{cases}$$

If $S \subseteq \mathcal{P}(\mathbb{N})$, let

$$\Phi_S = \bigvee_{A \in S} \phi_A.$$

The sentence Φ_S defines the class of structures $\mathcal{A}_S = \{(\mathbb{N}, A) | A \in S\}$ within $\mathcal{A}$. The sentence Φ_S has quantifier rank 0. But when S is an uncountable subset of $\mathcal{P}(\mathbb{N})$, the sentence Φ_S has uncountable size.

In fact, there are only $2^{\aleph_0}$ many L-sentences in $L_{\omega_1\omega}$, but there are $2^{2^{\aleph_0}}$ many $S \subseteq \mathcal{P}(\mathbb{N})$. Therefore there must exist some $S \subseteq \mathcal{P}(\mathbb{N})$ such that the class $\mathcal{A}_S$ cannot be defined by an $L_{\omega_1\omega}$ sentence. Since all the Φ_S's are quantifier free, the usual Ehrenfeucht-Fraïssé Games cannot help us in deciding for which subsets S the class $\mathcal{A}_S$ is $L_{\omega_1\omega}$-definable in $\mathcal{A}$ and for which it is not.

[6] According to Barwise [1], Charles Peirce thought of quantifiers as infinite conjunctions or disjunctions, and this was picked up by Löwenheim, Wittgenstein and others, and used in proof theory by Novikoff already in the 1940s.

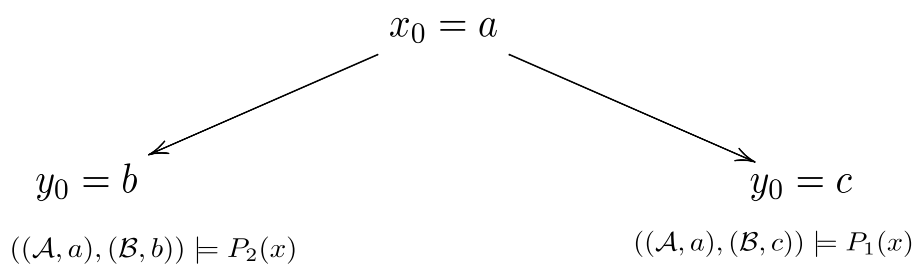

Figure 4: A game tree

If the above example can be seen as a factual analysis of the deficiency of the usual Ehrenfeucht-Fraïssé Games applied to the logic $L_{\omega_1\omega}$, here we also offer a conceptual analysis. In the game $\text{EFD}_\alpha(\mathcal{A}, \mathcal{B})$, a winning strategy of **I** *corresponds* to a separation formula of $\mathcal{A}$ and $\mathcal{B}$. But how does this correspondence work? In other words, given a winning strategy of **I**, how do we find a separation formula? Let us consider the following example.

Example 7. Suppose $L = \langle P_1, P_2 \rangle$. Let $\mathcal{A} = \{a\}$, $P_1^{\mathcal{A}} = P_2^{\mathcal{A}} = \{a\}$. Let $\mathcal{B} = \{b, c\}$, $P_1^{\mathcal{B}} = \{b\}$ and $P_2^{\mathcal{B}} = \{c\}$.

There is a straightforward winning strategy for **I** in $\text{EFD}_1(\mathcal{A}, \mathcal{B})$. He begins by playing $a \in \mathcal{A}$. If **II** plays b, **I** wins by pointing out that $\mathcal{A} \models P_2(a)$ and $\mathcal{B} \models \neg P_2(b)$. If **II** plays c, **I** wins by pointing out that $\mathcal{A} \models P_1(a)$ and $\mathcal{B} \models \neg P_1(c)$. This can be visualized by the game tree in Figure 4. Let $(\mathcal{A}, \mathcal{B}) \models \phi$ be the shorthand for $\mathcal{A} \models \phi$ and $\mathcal{B} \models \neg\phi$.

The separation formula corresponding to this strategy of **I**'s is obviously $\forall x(P_1(x) \wedge P_2(x))$. We have

$$\mathcal{A} \models \forall x(P_1(x) \wedge P_2(x)) \qquad \mathcal{B} \models \neg\forall x(P_1(x) \wedge P_2(x)).$$

The point is that any single play in the game cannot give **I** this separation formula. To get this formula, **I** has to survey the entire game tree (in this case, both branches). This task he does 'effortlessly'. The game EFD assumes a kind of omniscience of player **I**: he is able to know the outcome of all possible plays without actually playing them one by one. Technically, this is related to the fact that the rank of the game does not go down when the players move from a pair of structures corresponding to a conjunction to a pair of structures corresponding to a conjunct. We may say that boolean connectives are 'transparent' to the game EFD. In fact, in proofs about the game EFD the boolean cases are almost always trivial.

In the new game that we are going to define, life will not be as easy for player **I**. In order to survey the outcome of different branches, he now has to play. Technically this means that the rank of the game goes down when the players move from a pair of structures corresponding to a conjunction to a pair of structures corresponding to a conjunct. In this way the boolean size of a formula is taken into consideration. Moreover, the new game has the advantage of being defined on classes of structures. This game is at the same time a game for propositional logic. All this talk might seem vague to the reader. We will make it precise in the next section.

4 An Ehrenfeucht-Fraïssé Game for $L_{\omega_1\omega}$

4.1 A Game for First Order Logic

Hella and Väänänen introduced a game for propositional and first order logic in [9]. This game measures a first order formula not according to its quantifier rank, but according to its size. This game will be the blueprint of our Ehrenfeucht-Fraïssé Game for $L_{\omega_1\omega}$.

We need some notation. Our vocabulary is relational. Throughout this section we assume formulas to be in negation normal form. The universe of a structure $\mathcal{A}$ is denoted by A, of $\mathcal{B}$ by B. We use $x_j, j \in \mathbb{N}$ to denote variables. A variable assignment for a structure $\mathcal{A}$ is a finite partial mapping $\alpha : \mathbb{N} \to A$. The finite domain of α is denoted by $\mathrm{dom}(\alpha)$.

We consider classes $\mathfrak{A}$ of structures $(\mathcal{A}, \alpha)$, where $\mathcal{A}$ is a model and α is a variable assignment. We assume that whenever $(\mathcal{A}, \alpha), (\mathcal{B}, \beta) \in \mathfrak{A}$, then $\mathcal{A}$ and $\mathcal{B}$ have the same vocabulary, and α and β have the same domain, which we denote by $\mathrm{dom}(\mathfrak{A})$. If α is an assignment on $\mathcal{A}$, $a \in A$ and $j \in \mathbb{N}$, then $\alpha(a/j)$ is the assignment that maps j to a and agrees with α otherwise. If F is a choice function on $\mathfrak{A}$, namely that F is a function defined on $\mathfrak{A}$ such that $F(\mathcal{A}, \alpha) \in A$ for all $(\mathcal{A}, \alpha) \in \mathfrak{A}$, then $\mathfrak{A}(F/j)$ is defined as $\{(\mathcal{A}, \alpha(F(\mathcal{A}, \alpha)/j) | (\mathcal{A}, \alpha) \in \mathfrak{A}\}$. Finally, $\mathfrak{A}(\star/j) = \{(\mathcal{A}, \alpha(a/j)) | (\mathcal{A}, \alpha) \in \mathfrak{A}, a \in A)\}$.

Let $\mathfrak{A}$ and $\mathfrak{B}$ be classes of structures of a relational vocabulary. Assume further that $\mathrm{dom}(\mathfrak{A}) = \mathrm{dom}(\mathfrak{B})$, and ϕ is a formula such that $j \in \mathrm{dom}(\mathfrak{A})$ for all free variables in ϕ. We say that ϕ *separates the classes* $\mathfrak{A}$ *and* $\mathfrak{B}$, denoted by $(\mathfrak{A}, \mathfrak{B}) \models \phi$, if $(\mathcal{A}, \alpha) \models \phi$ for all $(\mathcal{A}, \alpha) \in \mathfrak{A}$ and $(\mathcal{B}, \beta) \not\models \phi$ for all $(\mathcal{B}, \beta) \in \mathfrak{B}$.

Definition 8. Let L be a relational vocabulary and $\mathfrak{A}$, $\mathfrak{B}$ classes of L-structures. Let n be a positive integer. The game $\mathrm{EFB}_n(\mathfrak{A}, \mathfrak{B})$ has two players. The number n is called the *rank* of the game. The positions in the game are of the form $(\mathfrak{A}_m, \mathfrak{B}_m, r_m)$ where $\mathfrak{A}_m, \mathfrak{B}_m$ are classes of L-structures and $r_m \in \mathbb{N}$. The game begins from position $(\mathfrak{A}, \mathfrak{B}, n)$. Suppose the position after m moves is $(\mathfrak{A}_m, \mathfrak{B}_m, r_m)$. There are the following possibilities for the next move in the game.

Left splitting move: Player **I** first chooses positive numbers u and v such that $u + v = r_m$. Then **I** represents $\mathfrak{A}$ as a union $\mathfrak{C} \cup \mathfrak{D}$. Now the game continues from the position $(\mathfrak{C}, \mathfrak{B}_m, u)$ or from the position $(\mathfrak{D}, \mathfrak{B}_m, v)$, and player **II** can choose which.

Right splitting move: Player **II** first chooses positive numbers u and v such that $u + v = r_m$. Then **I** represents $\mathfrak{B}$ as a union $\mathfrak{C} \cup \mathfrak{D}$. Now the game continues from the position $(\mathfrak{A}_m, \mathfrak{C}, u)$ or from the position $(\mathfrak{A}_m, \mathfrak{D}, v)$, and player **II** can choose which.

Left supplementing move: Player **I** picks an element from each structure $(\mathcal{A}, \alpha) \in \mathfrak{A}_m$. More precisely, **I** chooses a natural number j and a choice function F for $\mathfrak{A}_m$. Then the game continues from the position $(\mathfrak{A}_m(F/j), \mathfrak{B}_m(\star/j), r_m - 1)$.

Right supplementing move: Player **I** picks an element from each structure $(\mathcal{B}, \beta) \in \mathfrak{B}_m$. More precisely, **I** chooses a natural number j and a choice function F for $\mathfrak{B}_m$. Then the game continues from the position $(\mathfrak{A}_m(\star/j), \mathfrak{B}_m(F/j), r_m - 1)$.

The game ends in a position $(\mathfrak{A}_m, \mathfrak{B}_m, r_m)$ and player **I** wins if there is an atomic or a negated atomic formula ϕ such that $(\mathfrak{A}_m, \mathfrak{B}_m) \models \phi$. Player **II** wins the game if they reach a position $(\mathfrak{A}_m, \mathfrak{B}_m, r_m)$ such that $r_m = 1$ and **I** does not win in this position.

We have defined the game $\mathrm{EFB}_n(\mathfrak{A}, \mathfrak{B})$ by describing the rules and winning positions of the game. We can convert this definition into the form of $\mathcal{G}_\omega(A, W)$, but the latter would be more obscure. This game is closed, and therefore determined by the Gale-Stewart Theorem.

We define the following measure of the size of a formula.

Definition 9. The *size*, denoted by $\mathrm{s}(\phi)$, of a formula ϕ in first order logic is defined inductively as follows:

$$
\begin{aligned}
\mathrm{s}(\phi) &= 1 \text{ if } \phi \text{ is an atomic or negated atomic formula} \\
\mathrm{s}(\phi \wedge \psi) &= \mathrm{s}(\phi) + \mathrm{s}(\psi) \\
\mathrm{s}(\phi \vee \psi) &= \mathrm{s}(\phi) + \mathrm{s}(\psi) \\
\mathrm{s}(\exists x \phi) &= \mathrm{s}(\phi) + 1 \\
\mathrm{s}(\forall x \phi) &= \mathrm{s}(\phi) + 1.
\end{aligned}
$$

Note that the size of a first order formula is a natural number. The size of a formula is always bigger or equal to its quantifier rank.

The game EFB_n characterizes separation in first order logic up to size n:

Theorem 19 (Hella-Väänänen)**.** *Let L be a relational vocabulary, $\mathfrak{A}$ and $\mathfrak{B}$ classes of L-structures, and n be a positive integer. Then the following are equivalent.*

1. *Player **I** has a winning strategy in the game* $\mathrm{EFB}_n(\mathfrak{A}, \mathfrak{B})$.
2. *There is a first order L-formula of size $\leq n$ such that* $(\mathfrak{A}, \mathfrak{B}) \models \phi$.

Proof. We prove the theorem by induction on n. The case where $n = 1$ follows directly from the definition of the game. Suppose 1 and 2 are equivalent for all $k < n$. Now consider $k = n$.

(1 ⇒ 2) Suppose **I** has a winning strategy in the game $\mathrm{EFB}_n(\mathfrak{A}, \mathfrak{B})$. We look at his first move in this strategy. Depending on which type of move it is, there are the following four cases.

Case 1. Player **I** first makes a left splitting move. He represents $\mathfrak{A}$ as a union $\mathfrak{C} \cup \mathfrak{D}$. He chooses natural numbers u and v such that $u+v = n$. Now since this is a winning strategy, both $(\mathfrak{C}, \mathfrak{B}, u)$ and $(\mathfrak{D}, \mathfrak{B}, v)$ are winning positions for **I**. By the induction hypothesis, there are formulas ψ and θ such that $\mathrm{s}(\psi) \leq u$, $\mathrm{s}(\theta) \leq v$, $(\mathfrak{C}, \mathfrak{B}) \models \psi$ and $(\mathfrak{D}, \mathfrak{B}) \models \theta$. Let ϕ be the formula $\psi \vee \theta$. Then we have that $\mathrm{s}(\phi) = \mathrm{s}(\psi) + \mathrm{s}(\theta) \leq u + v = n$, and that $(\mathfrak{A}, \mathfrak{B}) \models \phi$.

Case 2. Player **I** first makes a right splitting move. He represents $\mathfrak{B}$ as a union $\mathfrak{C} \cup \mathfrak{D}$. He chooses natural numbers u and v such that $u + v = n$. Now since this is a winning strategy, both $(\mathfrak{A}, \mathfrak{C}, u)$ and $(\mathfrak{A}, \mathfrak{D}, v)$ are winning positions for **I**. By the induction hypothesis, there are formulas ψ and θ such that $\mathrm{s}(\psi) \leq u$, $\mathrm{s}(\theta) \leq v$, $(\mathfrak{A}, \mathfrak{C}) \models \psi$ and $(\mathfrak{A}, \mathfrak{D}) \models \theta$. Let ϕ be the formula $\psi \wedge \theta$. Then we have that $\mathrm{s}(\phi) = \mathrm{s}(\psi) + \mathrm{s}(\theta) \leq u + v = n$, and that $(\mathfrak{A}, \mathfrak{B}) \models \phi$.

Case 3. Player **I** first makes a left supplementing move. He chooses a natural number j and a choice function F for $\mathfrak{A}$. The game continues from the position $(\mathfrak{A}(F/j), \mathfrak{B}(\star/j), n-1)$. Since this is a winning strategy for **I**, the position $(\mathfrak{A}(F/j), \mathfrak{B}(\star/j), n-1)$ is a winning position for him too. By the induction hypothesis, there is a formula ψ such that $\mathrm{s}(\psi) \leq n-1$, and $(\mathfrak{A}(F/j), \mathfrak{B}(\star/j)) \models \psi$. Let ϕ be the formula $\exists x_j \psi$. Then we have that $\mathrm{s}(\phi) = \mathrm{s}(\psi) + 1 \leq n$, and that $(\mathfrak{A}, \mathfrak{B}) \models \phi$.

Case 4. Player **I** first makes a right supplementing move. He chooses a natural number j and a choice function F for $\mathfrak{B}$. The game continues from the position $(\mathfrak{A}(\star/j), \mathfrak{B}(F/j), n-1)$. Since this is a winning strategy for **I**, the position $(\mathfrak{A}(\star/j), \mathfrak{B}(F/j), n-1)$ is a winning position for him too. By the induction hypothesis, there is a formula ψ such that $\mathrm{s}(\psi) \leq n-1$, and $(\mathfrak{A}(\star/j), \mathfrak{B}(F/j)) \models \psi$. Let ϕ be the formula $\forall x_j \psi$. Then we have that $\mathrm{s}(\phi) = \mathrm{s}(\psi) + 1 \leq n$, and that $(\mathfrak{A}, \mathfrak{B}) \models \phi$.

Now for the converse direction (2 ⇒ 1). Suppose there is a first order formula ϕ with size $\leq n$ such that $(\mathfrak{A}, \mathfrak{B}) \models \phi$. Depending on the shape of ϕ, there are the following possibilities.

Case 1. ϕ is an atomic formula. By definition **I** wins the game $\mathrm{EFB}_1(\mathfrak{A}, \mathfrak{B})$.

Case 2. ϕ is $\psi \vee \theta$. Let $\mathfrak{C}$ be the class of structures $(\mathcal{A}, \alpha) \in \mathfrak{A}$ such that $(\mathcal{A}, \alpha) \models \psi$, $\mathfrak{D}$ be the class of structures $(\mathcal{A}, \alpha) \in \mathfrak{A}$ such that $(\mathcal{A}, \alpha) \models \theta$. Since $(\mathcal{A}, \alpha) \models \phi$ for all $(\mathcal{A}, \alpha) \in \mathfrak{A}$, we have $\mathfrak{A} = \mathfrak{C} \cup \mathfrak{D}$. Moreover since $(\mathcal{B}, \beta) \not\models \phi$ for all $(\mathcal{B}, \beta) \in \mathfrak{B}$, we have $(\mathfrak{C}, \mathfrak{B}) \models \psi$ and $(\mathfrak{D}, \mathfrak{B}) \models \theta$. Finally, as $\mathrm{s}(\phi) \leq n$, there are numbers u and v such that $\mathrm{s}(\psi) \leq u$, $\mathrm{s}(\theta) \leq v$, and $u + v = n$. By the induction hypothesis, **I** has a winning strategy in both $(\mathfrak{C}, \mathfrak{B}, u)$ and $(\mathfrak{D}, \mathfrak{B}, v)$. Therefore he has a winning strategy in the position $(\mathfrak{A}, \mathfrak{B}, n)$ by first making a left splitting move, and then follow the winning strategy in $(\mathfrak{C}, \mathfrak{B}, u)$ or $(\mathfrak{D}, \mathfrak{B}, v)$.

Case 3. ϕ is $\psi \wedge \theta$. Let $\mathfrak{C}$ be the class of structures $(\mathcal{B}, \beta) \in \mathfrak{B}$ such that $(\mathcal{B}, \beta) \not\models \psi$, $\mathfrak{D}$ be the class of structures $(\mathcal{B}, \beta) \in \mathfrak{B}$ such that $(\mathcal{B}, \beta) \not\models \theta$. Since $(\mathcal{B}, \alpha) \not\models \phi$ for all $(\mathcal{B}, \beta) \in \mathfrak{B}$, we have $\mathfrak{B} = \mathfrak{C} \cup \mathfrak{D}$. Moreover since $(\mathcal{A}, \alpha) \models \phi$ for all $(\mathcal{A}, \alpha) \in \mathfrak{A}$, we have $(\mathfrak{A}, \mathfrak{C}) \models \psi$ and $(\mathfrak{A}, \mathfrak{D}) \models \theta$. Finally, as $\mathrm{s}(\phi) \leq n$, there are numbers u and v such that $\mathrm{s}(\psi) \le u$, $\mathrm{s}(\theta) \le v$, and $u + v = n$. By the induction hypothesis, **I** has a winning strategy in both $(\mathfrak{A}, \mathfrak{C}, u)$ and $(\mathfrak{A}, \mathfrak{D}, v)$. Therefore he has a winning strategy at $(\mathfrak{A}, \mathfrak{B}, n)$ by first making a right splitting move, and then follow the winning strategy at $(\mathfrak{C}, \mathfrak{B}, u)$ or $(\mathfrak{D}, \mathfrak{B}, v)$.

Case 4. ϕ is $\exists x_j \psi$. Since $(\mathcal{A}, \alpha) \models \phi$ for all $(\mathcal{A}, \alpha) \in \mathfrak{A}$, there is a choice function F for $\mathfrak{A}$ such that $(\mathcal{A}, \alpha(F(\mathcal{A}, \alpha)/j)) \models \psi$ for all $(\mathcal{A}, \alpha) \in \mathfrak{A}$. Thus $(\mathcal{A}, \alpha^*) \models \psi$ for all $(\mathcal{A}, \alpha^*) \in \mathfrak{A}(F/j)$. On the other hand, we have that $(\mathfrak{B}, \beta^*) \not\models \psi$ for all $(\mathfrak{B}, \beta^*) \in \mathfrak{B}(\star/j)$. Therefore $(\mathfrak{A}(F/j), \mathfrak{B}(\star/j)) \models \psi$. By the induction hypothesis, player **I** has a winning strategy in the position $(\mathfrak{A}(F/j), \mathfrak{B}(\star/j), \mathrm{s}(\psi))$. Note that $\mathrm{s}(\phi) = \mathrm{s}(\psi) + 1$. Hence **I** has a winning strategy in the position $(\mathfrak{A}, \mathfrak{B}, \mathrm{s}(\phi))$ by first making a left supplementing move, and then follow the winning strategy in $(\mathfrak{A}(F/j), \mathfrak{B}(\star/j), \mathrm{s}(\psi))$.

Case 5. ϕ is $\forall x_j \psi$. Since $(\mathcal{B}, \beta) \not\models \phi$ for all $(\mathcal{B}, \beta) \in \mathfrak{B}$, there is a choice function F for $\mathfrak{B}$ such that $(\mathcal{B}, \beta(F(\mathcal{B}, \beta)/j)) \not\models \psi$ for all $(\mathcal{B}, \beta) \in \mathfrak{B}$. Thus $(\mathcal{B}, \beta^*) \not\models \psi$ for all $(\mathcal{B}, \beta^*) \in \mathfrak{B}(F/j)$. On the other hand, we have that $(\mathfrak{A}, \alpha^*) \models \psi$ for all $(\mathfrak{A}, \alpha^*) \in \mathfrak{A}(\star/j)$. Therefore $(\mathfrak{A}(\star/j), \mathfrak{B}(F/j)) \models \psi$. By the induction hypothesis, player **I** has a winning strategy in the position $(\mathfrak{A}(\star/j), \mathfrak{B}(F/j), \mathrm{s}(\psi))$. Note that $\mathrm{s}(\phi) = \mathrm{s}(\psi) + 1$. Hence **I** has a winning strategy in $(\mathfrak{A}, \mathfrak{B}, \mathrm{s}(\phi))$ by first making a right supplementing move, and then follow the winning strategy in $(\mathfrak{A}(\star/j), \mathfrak{B}(F/j), \mathrm{s}(\psi))$.

□

It can be clearly seen from the above proof that the four kinds of moves in the game EFB correspond to four kinds of formula forming operations in first order logic: left splitting move to disjunction, right splitting move to conjunction, left supplementing move to existential quantification and right supplementing move to universal quantification. This correspondence enables us to consider various restrictions of the full game EFB_n. These games characterize separation in various fragments of first order logic. The following results can be found in [9].

Corollary 20. *Let* $\mathrm{EFB}^P_n(\mathfrak{A}, \mathfrak{B})$ *be the restriction of the game* $\mathrm{EFB}_n(\mathfrak{A}, \mathfrak{B})$ *where player* ***I*** *can only make splitting moves, but no supplementing moves. Then the following are equivalent.*

1. *Player* ***I*** *has a winning strategy in the game* $\mathrm{EFB}^P_n(\mathfrak{A}, \mathfrak{B})$.
2. *There is a quantifier-free L-formula of size* $\le n$ *such that* $(\mathfrak{A}, \mathfrak{B}) \models \phi$.

The game $\mathrm{EFB}^P_n(\mathfrak{A}, \mathfrak{B})$ is a essentially game for propositional logic.

Corollary 21. *Let* $\mathrm{EFB}^\exists_n(\mathfrak{A}, \mathfrak{B})$ *be the restriction of the game* $\mathrm{EFB}_n(\mathfrak{A}, \mathfrak{B})$ *where player* ***I*** *cannot make right supplementing moves. Then the following are equivalent.*

1. *Player* ***I*** *has a winning strategy in the game* $\mathrm{EFB}^\exists_n(\mathfrak{A}, \mathfrak{B})$.
2. *There is an existential L-formula of size* $\le n$ *such that* $(\mathfrak{A}, \mathfrak{B}) \models \phi$.

In Section 5 we will give an application of the game $\mathrm{EFB}_n^{\exists}$ in describing the length of linear orders. Further applications of these games can be found in [9]. Results include a proof using EFB_n^{P} that the minimal size of propositional formula defining the parity of binary strings of length n is n^2.

4.2 Extending the Game to $L_{\omega_1\omega}$

We are now in a position to define an Ehrenfeucht-Fraïssé Game for the logic $L_{\omega_1\omega}$. This game will be a proper extension of the game $\mathrm{EFB}_n(\mathfrak{A}, \mathfrak{B})$. We introduce a new type of moves into the game to account for infinite conjunctions and disjunctions of formulas in $L_{\omega_1\omega}$. In short, in the game $\mathrm{EFB}_n(\mathfrak{A}, \mathfrak{B})$ player **I** can split a class of structures into two pieces; he now also has the option to split it into *countably many* pieces.

The way that the rank of the game is measured needs to be modified accordingly. In the new game the rank needs no longer to be a natural number, instead it can be any countable ordinal. Here we encounter the interesting phenomenon that the arithmetic of infinite ordinals behaves rather differently from the arithmetic of finite ordinals. To name a few, addition of finite ordinals is always commutative, which is not the case for infinite ordinals; addition of finite non-zero ordinals is always strictly monotonic, in the sense that the sum of two non-zero finite ordinals is strictly greater than any one of the summands, which is in general not true for infinite ordinals. For these reasons, we find the usual addition of ordinals not suitable for measuring the rank of an Ehrenfeucht-Fraïssé Game for $L_{\omega_1\omega}$ or for measuring the size of infinitary formulas.

We need another way to add up infinite ordinals. The operation that we turn to is the *natural sum* of ordinals. Recall that every ordinal can be written in Cantor normal form (for details, see Chapter 2 in [12]).

Theorem 22. *Every ordinal $\gamma > 0$ can be represented uniquely in the form*

$$\gamma = \omega^{\alpha_1} \cdot k_1 + \ldots + \omega^{\alpha_n} \cdot k_n$$

where $n \geq 1$, $\gamma \geq \alpha_1 > \ldots > \alpha_n$, and $k_1, \ldots, k_n$ are non-zero natural numbers.

Using the Cantor normal form, one can define the natural sum of ordinals.

Definition 10. Let γ_1 and γ_2 be ordinals. One can represent γ_1 and γ_2 uniquely in the form

$$\gamma_1 = \omega^{\alpha_1} \cdot k_1 + \ldots + \omega^{\alpha_n} \cdot k_n$$

$$\gamma_2 = \omega^{\alpha_1} \cdot j_1 + \ldots + \omega^{\alpha_n} \cdot j_n$$

where $\alpha_1 > \ldots > \alpha_n$ is a sequence of ordinals, $k_1, \ldots, k_n$ and $j_1, \ldots, j_n$ are natural numbers satisfying $k_i + j_i > 0$ for all i. Define the *natural sum*, also called the *Hessenberg sum* of γ_1 and γ_2, denoted by $\gamma_1 \# \gamma_2$, of γ_1 and γ_2 as

$$\gamma_1 \# \gamma_2 = \omega^{\alpha_1} \cdot (k_1 + j_1) + \ldots + \omega^{\alpha_n} \cdot (k_n + j_n).$$

The natural sum of finite ordinals is just their usual ordinal sum. As soon as we step into the realm of infinite ordinals, the natural sum and the usual sum part ways. The natural sum is always greater or equal to the usual sum, but it may be greater. Consider the case $\gamma_1 = 1$ and $\gamma_2 = \omega$. We have

$$\gamma_1 + \gamma_2 = 1 + \omega = \omega,$$

while

$$\gamma_1 \# \gamma_2 = \omega + 1.$$

The natural sum of ordinals enjoys a group of desirable properties. It is commutative and associative. Moreover, the natural sum of non-zero ordinals is strictly monotonic. Given ordinals $\gamma_1, \gamma_2 > 0$, we always have $\gamma_1 \# \gamma_2 > \gamma_1$ and $\gamma_1 \# \gamma_2 > \gamma_2$.

We also introduce the natural sum of a countable sequence of ordinals.

Definition 11. Let $\{\gamma_i | i \in \omega\}$ be a sequence of ordinals. Let S_n denote the natural sum of the first n items in the sequence

$$S_n = \gamma_0 \# \ldots \# \gamma_{n-1}.$$

Define the *infinite natural sum*, denoted by $\sum_{i \in \omega}^{\#} \gamma_i$, of the sequence $\{\gamma_i | i \in \omega\}$ as

$$\sum_{i \in \omega}^{\#} \gamma_i = \sup_{n \in \mathbb{N}} S_n.$$

The infinite natural sum is invariant under permutations of the summands, so we may say it is commutative in a generalized sense. More precisely, if p is a permutation of ω, we have

$$\sum_{i \in \omega}^{\#} \gamma_i = \sum_{i \in \omega}^{\#} \gamma_{p(i)}.$$

To see that this is true, it suffices to realize that any finite natural sum S_n of the sequence $\{\gamma_i | i \in \omega\}$ is subsumed by some finite natural sum of the sequence $\{\gamma_{p(i)} | i \in \omega\}$, and vice versa.

The infinite natural sum of non-zero ordinals is also strictly monotonic. Given a sequence $\{\gamma_i | i \in \omega\}$ of non-zero ordinals, we have that $\sum_{i \in \omega}^{\#} \gamma_i > \gamma_i, i \in \omega$. This property is important for our purposes.

We can now define an Ehrenfeucht-Fraïssé Game for the logic $L_{\omega_1 \omega}$.

Definition 12. Let L be a relational vocabulary and $\mathfrak{A}$, $\mathfrak{B}$ classes of L-structures. Let α be a countable ordinal. The game $\mathrm{EFB}_\alpha(\mathfrak{A}, \mathfrak{B})$ has two players. The number α is called the *rank* of the game. The positions in the game are of the form $(\mathfrak{A}_m, \mathfrak{B}_m, \gamma)$ where $\mathfrak{A}_m, \mathfrak{B}_m$ are classes of L-structures and γ is an ordinal. The game begins from position $(\mathfrak{A}, \mathfrak{B}, \alpha)$. Suppose the position after m moves is $(\mathfrak{A}_m, \mathfrak{B}_m, \gamma)$. There are the following possibilities for the next move in the game.

Finite left splitting move: Player **I** first chooses non-zero ordinals γ_1 and γ_2 such that $\gamma_1 \# \gamma_2 \leq \gamma$. Then **I** represents $\mathfrak{A}_m$ as a union $\mathfrak{C} \cup \mathfrak{D}$. Now the game continues from the position $(\mathfrak{C}, \mathfrak{B}_m, \gamma_1)$ or from the position $(\mathfrak{D}, \mathfrak{B}_m, \gamma_2)$, and player **II** can choose which.

Finite right splitting move: Player **I** first chooses non-zero ordinals γ_1 and γ_2 such that $\gamma_1 \# \gamma_2 \leq \gamma$. Then **I** represents $\mathfrak{B}_m$ as a union $\mathfrak{C} \cup \mathfrak{D}$. Now the game continues from the position $(\mathfrak{A}_m, \mathfrak{C}, \gamma_1)$ or from the position $(\mathfrak{A}_m, \mathfrak{D}, \gamma_2)$, and player **II** can choose which.

Infinite left splitting move: Player **I** first chooses a sequence of non-zero ordinals $\{\gamma_i | i \in \omega\}$ such that $\sum_{i\in\omega}^{\#} \gamma_i \leq \gamma$. Then **I** represents $\mathfrak{A}_m$ as a countable union $\bigcup_{i\in\omega} \mathfrak{C}_i$ of L-structures. Now the game continues from the position $(\mathfrak{C}_i, \mathfrak{B}_m, \gamma_i)$ for some $i \in \omega$, and player **II** can choose which.

Infinite right splitting move: Player **I** first chooses a sequence of non-zero ordinals $\{\gamma_i | i \in \omega\}$ such that $\sum_{i\in\omega}^{\#} \gamma_i \leq \gamma$. Then **I** represents $\mathfrak{B}_m$ as a countable union $\bigcup_{i\in\omega} \mathfrak{D}_i$ of L-structures. Now the game continues from the position $(\mathfrak{A}_m, \mathfrak{D}_i, \gamma_i)$ for some $i \in \omega$, and player **II** can choose which.

Left supplementing move: Player **I** picks an element from each structure $(\mathcal{A}, \alpha) \in \mathfrak{A}_m$. More precisely, **I** chooses a natural number j and a choice function F for $\mathfrak{A}_m$. He also chooses an ordinal $\beta < \gamma$. Then the game continues from the position $(\mathfrak{A}_m(F/j), \mathfrak{B}_m(\star/j), \beta)$.

Right supplementing move: Player **I** picks an element from each structure $(\mathcal{B}, \beta) \in \mathfrak{B}_m$. More precisely, **I** chooses a natural number j and a choice function F for $\mathfrak{B}_m$. He also chooses an ordinal $\beta < \gamma$. Then the game continues from the position $(\mathfrak{A}_m(\star/j), \mathfrak{B}_m(F/j), \beta)$.

The game ends in a position $(\mathfrak{A}_m, \mathfrak{B}_m, \gamma)$ and player **I** wins if there is an atomic or a negated atomic formula ϕ such that $(\mathfrak{A}_m, \mathfrak{B}_m) \models \phi$. Player **II** wins the game if they reach a position $(\mathfrak{A}_m, \mathfrak{B}_m, \gamma)$ such that $\gamma = 1$ and **I** does not win in this position.

Note that when α is a finite ordinal, **I** cannot play infinite splitting moves. Therefore the game EFB_α restricted to finite ranks is essentially the same as the game EFB_n introduced in the Section 4.1. This is the reason why we keep the name EFB for the new game. It is easy to see that $\mathrm{EFB}_\alpha(\mathfrak{A}, \mathfrak{B})$ is a closed game. It is therefore determined by the Gale-Stewart Theorem.

We extend the definition of the size of a formula to include formulas in $L_{\omega_1\omega}$. Assume that all formulas are in negation normal form.

Definition 13. The *size*, denoted by $\mathrm{s}(\phi)$, of a formula ϕ in $L_{\omega_1\omega}$ is defined inductively as follows:

$$\begin{aligned}
\mathrm{s}(\phi) &= 1 \text{ if } \phi \text{ is an atomic or negated atomic formula}\\
\mathrm{s}(\phi \wedge \psi) &= \mathrm{s}(\phi)\#\mathrm{s}(\psi)\\
\mathrm{s}(\phi \vee \psi) &= \mathrm{s}(\phi)\#\mathrm{s}(\psi)\\
\mathrm{s}(\exists x\phi) &= \mathrm{s}(\phi) + 1\\
\mathrm{s}(\forall x\phi) &= \mathrm{s}(\phi) + 1\\
\mathrm{s}(\bigwedge_{i\in\omega} \phi_i) &= \sum_{i\in\omega}^{\#} \mathrm{s}(\phi_i)\\
\mathrm{s}(\bigvee_{i\in\omega} \phi_i) &= \sum_{i\in\omega}^{\#} \mathrm{s}(\phi_i).
\end{aligned}$$

The size of an $L_{\omega_1\omega}$ formula is a countable ordinal.

The properties of the natural sum of ordinals pass on to properties of the measurement of size for $L_{\omega_1\omega}$ formulas. When restricted to first order formulas, this measurement of size coincides with the measurement in Definition 9. This measurement is strictly monotonic, in the sense that the size of a formula is always strictly greater than its proper subformulas. This measurement is also invariant under permutations of conjuncts and disjuncts, in both the finite and infinite cases. More precisely, for formulas ϕ and ψ we always have $\mathrm{s}(\phi \vee \psi) = \mathrm{s}(\psi \vee \phi)$. For a sequence of formulas $\phi_i, i \in \omega$ and a permutation p of ω, we always have

$$s\Big(\bigvee_{i\in\omega} \phi_i\Big) = s\Big(\bigvee_{i\in\omega} \phi_{p(i)}\Big).$$

However this measurement of size is *not* invariant under logical equivalence in general.

One last remark before we proceed to prove the adequacy theorem of the game EFB_α. It might seem *ad hoc* to the reader that we treat finite conjunctions/disjunctions and infinite ones as different types of operations: in the game we have different rules for finite and infinite splitting moves, and in the measurement of the size of formulas finite and infinite conjunctions/disjunctions are treated as different cases. We claim that this is a price that has to be paid, given that we want to measure the size of infinitary formulas meaningfully. The usual approach in infinitary logic takes countable conjunction and disjunction as primitive operations in $L_{\omega_1\omega}$. Finite conjunction and disjunction are considered as abbreviations. Let ϕ and ψ be two formulas. The disjunction $\phi \vee \psi$ is taken as a shorthand for

$$\phi \vee \psi \vee \bot \vee \bot \vee \dots \tag{8}$$

This convention runs into trouble when we consider the size of the formula. Suppose ϕ and ψ are atomic formulas. Intuitively the formula $\phi \vee \psi$ should have size 2. In any case its size should be a finite number. On the other hand, formula (8) has infinite size under any reasonable measurement of size. More precisely, we consider a measurement of the size of $L_{\omega_1\omega}$ formulas to be reasonable if (1) it is an extension of the measurement of first order formulas in Definition 9, (2) it is monotonic. It is easy to see that if a measurement s satisfies these two conditions, then formula (8) has infinite size under this measurement. Therefore the formula $\phi \vee \psi$ should not be identified with formula (8), but rather be treated differently.

The game EFB_α characterizes separation in $L_{\omega_1\omega}$ up to size α. The following theorem is the central result in this paper.

Theorem 23 (Adequacy Theorem for EFB_α)**.** *Let L be a relational vocabulary, $\mathfrak{A}$ and $\mathfrak{B}$ classes of L-structures, and α a countable ordinal. Then the following are equivalent.*

1. *Player **I** has a winning strategy in the game* $\mathrm{EFB}_\alpha(\mathfrak{A}, \mathfrak{B})$.
2. *There is an L-formula ϕ in $L_{\omega_1\omega}$ of size $\leq \alpha$ such that* $(\mathfrak{A}, \mathfrak{B}) \models \phi$.

Proof. The strategy in proving this theorem is very much similar to Theorem 19. We prove the claim by induction on the rank of the game α. Again, when $\alpha = 1$ the proposition is obvious. Suppose the proposition is true for all ordinals $\gamma < \alpha$. Now consider the case $\gamma = \alpha$.

($1 \Longrightarrow 2$) Suppose **I** has a winning strategy in the game $\mathrm{EFB}_n(\mathfrak{A}, \mathfrak{B})$. We look at the first move in this strategy. Depending on which type of move it is, there are the following cases.

Case 1. Player **I** first makes a finite left splitting move. He represents $\mathfrak{A}$ as a union $\mathfrak{C} \cup \mathfrak{D}$. He chooses non-zero ordinals γ_1 and γ_2 such that $\gamma_1 \# \gamma_2 \leq \gamma$. Now since this is a winning strategy, both $(\mathfrak{C}, \mathfrak{B}, \gamma_1)$ and $(\mathfrak{D}, \mathfrak{B}, \gamma_2)$ are winning positions for **I**. Note that both γ_1 and γ_2 are strictly smaller than γ. By the induction hypothesis, there are L-formulas ψ and θ in $L_{\omega_1\omega}$ such that $\mathrm{s}(\psi) \leq \gamma_1$, $\mathrm{s}(\theta) \leq \gamma_2$, $(\mathfrak{C}, \mathfrak{B}) \models \psi$ and $(\mathfrak{D}, \mathfrak{B}) \models \theta$. Let ϕ be the formula $\psi \vee \theta$. Then we have that $\mathrm{s}(\phi) = \mathrm{s}(\psi) \# \mathrm{s}(\theta) = \gamma_1 \# \gamma_2 \leq \gamma$, and that $(\mathfrak{A}, \mathfrak{B}) \models \phi$.

Case 2. Player **I** first makes a finite right splitting move. He represents $\mathfrak{B}$ as a union $\mathfrak{C} \cup \mathfrak{D}$. He chooses non-zero ordinals γ_1 and γ_2 such that $\gamma_1 \# \gamma_2 \leq \gamma$. Now since this is a winning strategy, both $(\mathfrak{A}, \mathfrak{C}, \gamma_1)$ and $(\mathfrak{A}, \mathfrak{D}, \gamma_2)$ are winning positions for **I**. Note that both γ_1 and γ_2 are strictly smaller than γ. By the induction hypothesis, there are formulas ψ and θ such that $\mathrm{s}(\psi) \leq \gamma_1$, $\mathrm{s}(\theta) \leq \gamma_2$, $(\mathfrak{A}, \mathfrak{C}) \models \psi$ and $(\mathfrak{A}, \mathfrak{D}) \models \theta$. Let ϕ be the formula $\psi \wedge \theta$. Then we have that $\mathrm{s}(\phi) = \mathrm{s}(\psi) \# \mathrm{s}(\theta) \leq \gamma_1 \# \gamma_2 \leq \gamma$, and that $(\mathfrak{A}, \mathfrak{B}) \models \phi$.

Case 3. Player **I** first makes an infinite left splitting move. He represents $\mathfrak{A}$ as a union $\bigcup_{i \in \omega} \mathfrak{C}_i$. He chooses ordinals $\{\gamma_i | i \in \omega\}$ such that $\sum^{\#}_{i \in \omega} \gamma_i \leq \gamma$. Note that every γ_i is strictly smaller than γ. Now since this is a winning strategy, for every $i \in \omega$ $(\mathfrak{C}_i, \mathfrak{B}, \gamma_i)$ is a winning position for him. By the induction hypothesis, there are L-formulas ϕ_i in $L_{\omega_1\omega}$ such that $\mathrm{s}(\phi_i) \leq \gamma_i$ and $(\mathfrak{C}_i, \mathfrak{B}) \models \phi_i$ for all $i \in \omega$. Let ϕ be the formula $\bigvee_{i \in \omega} \phi_i$. It is an $L_{\omega_1\omega}$ formula. We have that $\mathrm{s}(\phi) = \sum^{\#}_{i \in \omega} \mathrm{s}(\phi_i) = \sum^{\#}_{i \in \omega} \gamma_i \leq \gamma$, and that $(\mathfrak{A}, \mathfrak{B}) \models \phi$.

Case 4. Player **I** first makes an infinite right splitting move. He represents $\mathfrak{B}$ as a union $\bigcup_{i \in \omega} \mathfrak{C}_i$. He chooses ordinals $\{\gamma_i | i \in \omega\}$ such that $\sum^{\#}_{i \in \omega} \gamma_i \leq \gamma$. Note that every γ_i is strictly smaller than γ. Now since this is a winning strategy, for every $i \in \omega$ $(\mathfrak{A}, \mathfrak{C}_i, \gamma_i)$ is a winning position for him. By the induction hypothesis, there are L-formulas ϕ_i in $L_{\omega_1\omega}$ such that $\mathrm{s}(\phi_i) \leq \gamma_i$ and $(\mathfrak{A}, \mathfrak{C}_i) \models \phi_i$ for all $i \in \omega$. Let ϕ be the formula $\bigwedge_{i\in\omega} \phi_i$. It is an $L_{\omega_1\omega}$ formula. We have that $\mathrm{s}(\phi) = \sum^{\#}_{i\in\omega} \mathrm{s}(\phi_i) = \sum^{\#}_{i\in\omega} \gamma_i \leq \gamma$, and that $(\mathfrak{A}, \mathfrak{B}) \models \phi$.

Case 5. Player **I** first makes a left supplementing move. He chooses a natural number j and a choice function F for $\mathfrak{A}$. He also chooses an ordinal $\beta < \gamma$. The game continues from $(\mathfrak{A}(F/j), \mathfrak{B}(\star/j), \beta)$. Since this is a winning strategy for **I**, the position $(\mathfrak{A}(F/j), \mathfrak{B}(\star/j), \beta)$ is a winning position for him too. By the induction hypothesis, there is a formula ψ such that $\mathrm{s}(\psi) \leq \beta$, and $(\mathfrak{A}(F/j), \mathfrak{B}(\star/j)) \models \psi$. Let ϕ be the formula $\exists x_j \psi$. Then we have that $\mathrm{s}(\phi) = \mathrm{s}(\psi) + 1 \leq \beta + 1 \leq \alpha$, and that $(\mathfrak{A}, \mathfrak{B}) \models \phi$.

Case 6. Player **I** first makes a right supplementing move. He chooses a natural number j and a choice function F for $\mathfrak{B}$. He also chooses an ordinal $\beta < \gamma$. The game continues from $(\mathfrak{A}(\star/j), \mathfrak{B}(F/j), \beta)$. Since this is a winning strategy for **I**, the position $(\mathfrak{A}(\star/j), \mathfrak{B}(F/j), \beta)$ is a winning position for him too. By the induction hypothesis, there is a formula ψ such that $\mathrm{s}(\psi) \leq \beta$, and $(\mathfrak{A}(\star/j), \mathfrak{B}(F/j)) \models \psi$. Let ϕ be the formula $\forall x_j \psi$. Then we have that $\mathrm{s}(\phi) = \mathrm{s}(\psi) + 1 \leq \beta + 1 \leq \alpha$, and that $(\mathfrak{A}, \mathfrak{B}) \models \phi$.

Now for the converse direction ($2 \Longrightarrow 1$). Suppose there is an $L_{\omega_1\omega}$ formula ϕ of size $\leq \alpha$ such that $(\mathfrak{A}, \mathfrak{B}) \models \phi$. Depending on the shape of ϕ, there are the following possibilities. Note that the game EFB_α is perfectly symmetric with respect to left and right. Therefore in what follows we omit the proof of several cases when their duals have already been proven.

Case 1. ϕ is an atomic formula. By definition player **I** wins the game $\mathrm{EFB}_1(\mathfrak{A}, \mathfrak{B})$.

Case 2. ϕ is $\psi \vee \theta$. Let $\mathfrak{C}$ be the class of structures $(\mathcal{A}, \alpha) \in \mathfrak{A}$ such that $(\mathcal{A}, \alpha) \models \psi$, $\mathfrak{D}$ be the class of structures $(\mathcal{A}, \alpha) \in \mathfrak{A}$ such that $(\mathcal{A}, \alpha) \models \theta$. Since $(\mathcal{A}, \alpha) \models \phi$ for all $(\mathcal{A}, \alpha) \in \mathfrak{A}$, we have $\mathfrak{A} = \mathfrak{C} \cup \mathfrak{D}$. Moreover since $(\mathcal{B}, \beta) \not\models \phi$ for all $(\mathcal{B}, \beta) \in \mathfrak{B}$, we have $(\mathfrak{C}, \mathfrak{B}) \models \psi$ and $(\mathfrak{D}, \mathfrak{B}) \models \theta$. Finally, as $\mathrm{s}(\phi) \leq \gamma$, there are ordinals γ_1 and γ_2 such that $\mathrm{s}(\psi) \leq \gamma_1$, $\mathrm{s}(\theta) \leq \gamma_2$, and $\gamma_1 \# \gamma_2 \leq \gamma$. By the induction hypothesis, **I** has a winning strategy in both $(\mathfrak{C}, \mathfrak{B}, \gamma_1)$ and $(\mathfrak{D}, \mathfrak{B}, \gamma_2)$. Therefore he has a winning strategy in $(\mathfrak{A}, \mathfrak{B}, \gamma)$ by first making a finite left splitting move, and then follow the winning strategy in $(\mathfrak{C}, \mathfrak{B}, \gamma_1)$ or $(\mathfrak{D}, \mathfrak{B}, \gamma_2)$.

Case 3. ϕ is $\psi \wedge \theta$. This case is completely dual to Case 2. We omit the details of the argument.

Case 4. ϕ is $\bigvee_{i\in\omega} \phi_i$. For each i, let $\mathfrak{C}_i$ be the class of structures $(\mathcal{A}, \alpha) \in \mathfrak{A}$ such that $(\mathcal{A}, \alpha) \models \phi_i$. Since $(\mathcal{A}, \alpha) \models \phi$ for all $(\mathcal{A}, \alpha) \in \mathfrak{A}$, we have $\mathfrak{A} = \bigcup_{i\in\omega} \mathfrak{C}_i$. Moreover since $(\mathcal{B}, \beta) \not\models \phi$ for all $(\mathcal{B}, \beta) \in \mathfrak{B}$, we have $(\mathfrak{C}_i, \mathfrak{B}) \models \phi_i$ for every i. Finally, as $\mathrm{s}(\phi) \leq \gamma$, there are ordinals $\{\gamma_i | i \in \omega\}$ such that $\mathrm{s}(\phi_i) \leq \gamma_i$ and $\sum^{\#}_{i\in\omega} \gamma_i \leq \gamma$. By the induction hypothesis, **I** has a winning strategy in both $(\mathfrak{C}_i, \mathfrak{B}, \gamma_i)$ for every i. Therefore he has a winning strategy in $(\mathfrak{A}, \mathfrak{B}, \gamma)$ by first making an infinite left splitting move, and then follow the winning strategy in some $(\mathfrak{C}_i, \mathfrak{B}, \gamma_i)$.

Case 5. ϕ is $\bigwedge_{i\in\omega} \phi_i$. This case is dual to Case 4.

Case 6. ϕ is $\exists x_j \psi$. Since $(\mathcal{A}, \alpha) \models \phi$ for all $(\mathcal{A}, \alpha) \in \mathfrak{A}$, there is a choice function F for $\mathfrak{A}$ such that $(\mathcal{A}, \alpha(F(\mathcal{A}, \alpha)/j)) \models \psi$. Thus $(\mathcal{A}, \alpha^*) \models \psi$ for all $(\mathcal{A}, \alpha^*) \in \mathfrak{A}(F/j)$. On the other hand, we have that $(\mathfrak{B}, \beta^*) \not\models \psi$ for all $(\mathfrak{B}, \beta^*) \in \mathfrak{B}(\star/j)$. Therefore $(\mathfrak{A}(F/j), \mathfrak{B}(\star/j)) \models \psi$. By the induction hypothesis, player **I** has a winning strategy in $(\mathfrak{A}(F/j), \mathfrak{B}(\star/j), \mathrm{s}(\psi))$. Note that $\mathrm{s}(\phi) = \mathrm{s}(\psi) + 1$. Hence **I** has a winning strategy in $(\mathfrak{A}, \mathfrak{B}, \gamma)$ by first making a left supplementing move, chooses the ordinal $\mathrm{s}(\psi) < \mathrm{s}(\phi) \leq \gamma$ and then follow the winning strategy in $(\mathfrak{A}(F/j), \mathfrak{B}(\star/j), \mathrm{s}(\psi))$.

Case 7. ϕ is $\forall x_j \psi$. This case is dual to Case 6.

□

With this theorem we complete the missing leg in the Strategic Balance of $L_{\omega_1\omega}$.

In practice, we can make the game easier for **I** to play by discarding some of the structures at hand. Given two classes of L-structures $\mathfrak{A}$ and $\mathfrak{A}'$, let $\mathfrak{A} \subseteq \mathfrak{A}'$ denote that every structure in $\mathfrak{A}$ is contained in $\mathfrak{A}'$. The following result will be useful in applications of the game.

Corollary 24. *Let $\mathfrak{A}, \mathfrak{A}', \mathfrak{B}, \mathfrak{B}'$ be classes of L-structures and α be a countable ordinal. Suppose $\mathfrak{A}' \subseteq \mathfrak{A}$ and $\mathfrak{B}' \subseteq \mathfrak{B}$. If player **I** has a winning strategy in* $\mathrm{EFB}_\alpha(\mathfrak{A}, \mathfrak{B})$*, then he also has a winning strategy in* $\mathrm{EFB}_\alpha(\mathfrak{A}', \mathfrak{B}')$.

Proof. Suppose **I** has a winning strategy in $\mathrm{EFB}_\alpha(\mathfrak{A}, \mathfrak{B})$. By Theorem 23, there is a formula ϕ with $\mathrm{s}(\phi) \leq \alpha$ such that $(\mathfrak{A}, \mathfrak{B}) \models \phi$. Since $\mathfrak{A}' \subseteq \mathfrak{A}$ and $\mathfrak{B}' \subseteq \mathfrak{B}$, we also have that $(\mathfrak{A}', \mathfrak{B}') \models \phi$. By Theorem 23 again, **I** has a winning strategy in $\mathrm{EFB}_\alpha(\mathfrak{A}', \mathfrak{B}')$. □

Like in the case of the finite game EFB_n, we are also interested in various restrictions of the full game EFB_α. These games correspond to fragments of the logic $L_{\omega_1\omega}$.

Corollary 25. *Let* $\mathrm{EFB}^P_\alpha(\mathfrak{A}, \mathfrak{B})$ *be the restriction of the game* $\mathrm{EFB}_\alpha(\mathfrak{A}, \mathfrak{B})$ *where player **I** can only make finite and infinite splitting moves, but no supplementing moves. Then the following are equivalent.*

1. *Player **I** has a winning strategy in the game* $\mathrm{EFB}^P_\alpha(\mathfrak{A}, \mathfrak{B})$.
2. *There is a quantifier-free $L_{\omega_1\omega}$ formula ϕ of size $\leq \alpha$ such that $(\mathfrak{A}, \mathfrak{B}) \models \phi$.*

The game EFB^P_α is essentially a game for propositional logic with countable conjunctions and disjunctions. In the final part of this paper we will give an application of this game.

5 Applications

The game EFB is useful in answering the following type of questions.

Question Given a property $\mathcal{P}$ of structures, what is the size of the smallest formula in L^* defining $\mathcal{P}$?

Depending on the situation, the logic L^* may vary: propositional, first order, existential, $L_{\omega_1\omega}$, etc. In this section we look at some of these applications. We first prove, using the game $\mathrm{EFB}_n^{\exists}(\mathfrak{A}, \mathfrak{B})$, that the minimal size of an existential formula defining the property 'a linear order has length at least n' is $2n-1$. This proof is contained in [9]. We then proceed to give some applications of the propositional fragment of the new infinite game EFB_α. Throughout this section we assume formulas to be in negation normal form.

5.1 The Existential Complexity of the Length of Linear Order

For the first application, consider the property that the length of a linear order is at least n, for n a natural number. Let us denote this property by $\mathcal{P}_n$. The property $\mathcal{P}_n$ is expressible in first order logic. Moreover, suppose ϕ is a first order formula defining this property, then ϕ is preserved under embeddings of structures. By Łoś-Tarski Theorem ([18], [21]), ϕ is logically equivalent to an *existential* formula. Henceforth we focus on existential formulas. We ask the question: What is the minimal existential formula defining this property?

The answer to this question depends on how we interpret the meaning of 'minimal'. If we take it to mean minimal in terms of quantifier rank, then the answer is well-known. There is a formula of logarithm quantifier rank expressing property $\mathcal{P}_n$. Let

$$A_0(x,y) = x < y$$
$$A_{k+1}(x,y) = \exists z (A_k(x,z) \wedge A_k(z,y) \wedge x < z \wedge z < y).$$

The formula $A_k(x,y)$ has quantifier rank k. It has the meaning in a linear order $(M,<)$ that the interval $(x,y]$ contains at least 2^k elements. Given a natural number n, represent $n-1$ as the sum of powers of 2:

$$n - 1 = 2^{k_1} + \ldots + 2^{k_m}$$

where $k_1, \ldots, k_m \in \mathbb{N}, k_1 > \ldots > k_m \geq 0$. Note that this representation is unique. Let $\psi_n(x_0)$ be the formula

$$\psi_n(x_0) = \exists z_1 \Bigg(x_0 < z_1 \wedge A_{k_1}(x_0, z_1) \wedge \exists z_2 \Bigg(z_1 < z_2 \wedge A_{k_2}(z_1, z_2) \\ \wedge \ldots \exists z_m \big(z_{m-1} < z_m \wedge A_{k_m}(z_{m-1}, z_m) \big) \ldots \Bigg) \Bigg).$$

The formula $\psi_n(x_0)$ says that there are at least $n-1$ elements above x_0 in the linear order. Let $\theta_n = \exists x_0 \psi_n(x_0)$. For any linear order $(M,<)$, we have

$$(M,<) \models \theta_n \iff \text{the length of } (M,<) \text{ is at least } n.$$

That is to say, θ_n defines property $\mathcal{P}_n$.

The formula θ_n has quantifier rank $\lceil \log(n-1) \rceil + 1$. Note that $\lceil \log(n) \rceil + 1 \geq \lceil \log(n-1) \rceil + 1 \geq \lceil \log(n) \rceil$. With a clever use of the Ehrenfeucht-Fraïssé Game, it can be shown that this result cannot be essentially improved (for details, see [8]). No formula of quantifier rank less than $\lceil \log(n) \rceil$ can define the property $\mathcal{P}_n$.

On the other hand, if we consider the minimal formula in terms of size that defines the property $\mathcal{P}_n$, the problem becomes quite different. To get started, let us calculate the size of the formula θ_n. An easy induction shows that the formula $A_k(x, y)$ has size $4 \cdot 2^k - 3$ in the sense of Definition 9. We have

$$\begin{aligned} \mathrm{s}(\theta_n) &= 1 + m + m + \mathrm{s}(A_{k_1}) + \ldots + \mathrm{s}(A_{k_m}) \\ &= 1 + 2m + 4(2^{k_1} + \ldots + 2^{k_m}) - 3m \\ &= 1 - m + 4(n-1) \\ &= 4n - m - 3 \\ &\geq 3n - 3. \end{aligned}$$

We claim that although the formula θ_n is minimal in terms of quantifier rank, it is not minimal in terms of size. The following formula also defines property $\mathcal{P}_n$:

$$\phi_n = \exists x_1 \ldots \exists x_n (x_1 < \ldots < x_n).$$

It is easy to see that ϕ_n has size $2n-1$. When $n > 2$, we have that $\mathrm{s}(\phi_n) < \mathrm{s}(\psi_n)$. In the rest of this section we show that the formula ϕ_n is indeed minimal in size in defining property $\mathcal{P}_n$.

In order to prove this claim, we use the game $\mathrm{EFB}_n^{\exists}$. Let $\mathfrak{A}_0 = \{(\mathcal{A}, \emptyset)\}$, where $\mathcal{A}$ is a linear order of length n, and let $\mathfrak{B}_0 = \{(\mathcal{B}, \emptyset)\}$, where $\mathcal{B}$ is a linear order of length $n-1$. Our aim is to show that player **II** has a winning strategy in the game $\mathrm{EFB}_m^{\exists}(\mathfrak{A}_0, \mathfrak{B}_0)$ for any $m < 2n-1$. If so, the claim would then follow from Corollary 21.

The idea of the proof is the following. We define a measure N for the 'distance from separation' between $\mathfrak{A}$ and $\mathfrak{B}$. In the starting position $(\mathfrak{A}_0, \mathfrak{B}_0)$ the distance count is $2n-1$. We prove that in every move of the game the rank of the position always goes down faster than the distance count N does. Therefore in every game of rank less than $2n-1$, there is still distance from separation left when the rank of the game runs up. This means that player **II** has a winning strategy.

Let us make this idea more precise. Consider a position $(\mathfrak{A}, \mathfrak{B}, u)$ in the game $\mathrm{EFB}_m^{\exists}(\mathfrak{A}_0, \mathfrak{B}_0)$. Since in the existential game no right supplementing move can be made, $\mathfrak{A}$ always consists of a single structure $(\mathcal{A}, \alpha)$. Let $a_1 < \ldots a_l$ be the elements in $\mathrm{ran}(\alpha)$, and let a_0 and a_{l+1} be the least and the largest element in $\mathfrak{A}$, respectively. We say that a variable assignment β in $\mathfrak{B}$ is *acceptable* if $\mathrm{dom}(\alpha)$=$\mathrm{dom}(\beta)$, and the mapping $\alpha(j) \mapsto \beta(j), j \in \mathrm{dom}(\alpha)$ preserves the relation $\leq$. More exactly, there are elements $b_1 \leq \ldots \leq b_l$ such that $\mathrm{ran}(\beta) = \{b_1, \ldots, b_l\}$ and for all $i \in \{1, \ldots, l\}$ and all $j \in \mathrm{dom}(\alpha)$

$$\alpha(j) = a_i \Longleftrightarrow \beta(j) = b_i. \tag{9}$$

We say that an assignment β is *nice* with respect to α if it is acceptable and in addition $|\{i \leq l : d(a_i, a_{i+1}) \neq d(b_i, b_{i+1})\}| = 1$, where $d(x, y)$ is the distance between x and y in the given linear order, and b_0 and b_{l+1} are the least and largest element in $\mathcal{B}$, respectively. We define the 'distance from defect' of β to be $\delta(\beta) = d(b_i, b_{i+1})$, where $i \leq l$ is the unique index such that $d(a_i, a_{i+1} \neq d(b_i, b_{i+1})$. We denote this index by $i(\beta)$. Note that $d(b_i, b_{i+1}) = d(a_i, a_{i+1}) - 1$ for $i = i(\beta)$. Intuitively, an assignment β is nice if the distance between adjacent elements in $\text{ran}(\beta)$ differs from the distance between corresponding elements in $\text{ran}(\alpha)$ at exactly one place. The function $i(\beta)$ denotes this place and the function $\delta(\beta)$ denotes this distance.

Given the singleton set $\mathfrak{A} = \{(\mathcal{A}, \alpha)\}$ and $\mathfrak{B}$ a set of structures of the form $(\mathcal{B}, \beta)$, we define the *niceness measure* of $\mathfrak{B}$ to be

$$N(\mathfrak{B}) = \sum_{\beta \in \mathcal{N}} (2\delta(\beta) + 1),$$

where $\mathcal{N}$ is the set of all nice variable assignments β such that $(\mathcal{B}, \beta) \in \mathfrak{B}$.

Whenever the niceness measure is non-zero, $\mathfrak{B}$ cannot be separated from $\mathfrak{A}$ by an atomic formula.

Lemma 26. *If $N(\mathfrak{B}) > 1$, then there is no atomic formula ϕ such that $(\mathfrak{A}, \mathfrak{B}) \models \phi$ or $(\mathfrak{A}, \mathfrak{B}) \models \neg\phi$.*

Proof. Suppose $N(\mathfrak{B}) > 1$, then either there is a nice assignment $\beta \in \mathcal{N}$ such that $\delta(\beta) \geq 1$, or there are two distinct assignments $\beta, \beta' \in \mathcal{N}$. Consider first the former case. Let β be such an assignment. There are elements $a_1, \ldots, a_l, b_1, \ldots, b_l$ such that $\text{ran}(\alpha) = \{a_1, \ldots, a_l\}$, $\text{ran}(\beta) = \{b_1, \ldots, b_l\}$, $a_1 < \ldots < a_l$ and $b_1 \leq \ldots \leq b_l$. The only place that $d(a_i, a_{i+1})$ may differ from $d(b_i, b_{i+1})$ is at $i = i(\beta)$. Since $\delta(\beta) \geq 1$, we have that $d(b_i, b_{i+1}) \geq 1$. Therefore in fact $b_1 < \ldots < b_l$. If follows from condition 9 that $(\mathcal{A}, \alpha)$ and $(\mathcal{B}, \beta)$ satisfy the same atomic formulas. Hence no atomic formula separates $\mathfrak{A}$ and $\mathfrak{B}$.

Now assume there are two distinct assignments $\beta, \beta' \in \mathcal{N}$. There are elements $a_1, \ldots, a_l, b_1, \ldots, b_l, b'_1, \ldots, b'_n$ such that $\text{ran}(\alpha) = \{a_1, \ldots, a_l\}$, $\text{ran}(\beta) = \{b_1, \ldots, b_l\}$, $\text{ran}(\beta') = \{b'_1, \ldots, b'_l\}$, $a_1 < \ldots < a_l$, $b_1 \leq \ldots \leq b_l$ and $b'_1 \leq \ldots \leq b'_l$. The fact that β and β' are different assignments implies that $i(\beta) \neq i(\beta')$. Hence for any $i, j \leq l$, we have either $(a_i < a_j \iff b_i < b_j)$ or $(a_i < a_j \iff b'_i < b'_j)$. It follows from condition (9) that no atomic formula separates $\mathfrak{A}$ from $\mathfrak{B}$. □

Lemma 27. *(a) If $\mathfrak{B} = \mathfrak{C} \cup \mathfrak{D}$, then $N(\mathfrak{C}) + N(\mathfrak{D}) \geq N(\mathfrak{B})$.*
(b) If $\mathfrak{A}' = \mathfrak{A}(F/j)$ and $\mathfrak{B}' = \mathfrak{B}(\star/j)$, then $N(\mathfrak{B}') \geq N(\mathfrak{B}) - 1$.

Proof. (a) Suppose $\mathfrak{B} = \mathfrak{C} \cup \mathfrak{D}$. Let $\mathcal{N}_B$ denote the set of nice assignments such that $(\mathcal{B}, \beta) \in \mathfrak{B}$, $\mathcal{N}_C$ of that in $\mathfrak{C}$ and $\mathcal{N}_D$ of that in $\mathfrak{D}$. We have that $\mathcal{N}_B \subseteq \mathcal{N}_C \cup \mathcal{N}_D$. Hence

$$N(\mathfrak{B}) = \sum_{\beta \in \mathcal{N}_B} (2\delta(\beta)+1) \leq \sum_{\beta \in \mathcal{N}_C} (2\delta(\beta)+1) + \sum_{\beta \in \mathcal{N}_D} (2\delta(\beta)+1) = N(\mathfrak{C}) + N(\mathfrak{D}).$$

(b) Let $\mathcal{N}$ and $\mathcal{N}'$ be the sets of nice assignments such that $(\mathcal{B},\beta)$ is in $\mathfrak{B}$ and $\mathfrak{B}$', respectively. Let $\mathrm{ran}(\alpha) = \{a_1,\ldots,a_l\}$ with $a_1 < \ldots < a_l$, and let a_0 and a_{l+1} be the least and the largest element in $\mathcal{A}$, respectively. For any nice assignment β, let $\mathrm{ran}(\beta) = \{b_1,\ldots,b_l\}$ with $b_1 \leq \ldots \leq b_l$. Furthermore we denote $F(\mathcal{A},\alpha)$ by c. We distinguish between the following cases.

Assume that $c = a_k$ for some $0 \leq k \leq l+1$. It is easy to see that for any assignment β, $\beta \in \mathcal{N}$ if and only if $\beta(b_k/j) \in \mathcal{N}'$. Moreover $\delta(\beta) = \delta(\beta(b_k/j))$. Note also that if $\beta \neq \beta'$, then $\beta(b_k/j) \neq \beta'(b_k/j)$. Hence we conclude that $N(\beta') \geq N(\beta)$.

Assume that c falls in the gap between a_h and a_{h+1} for some $0 \leq h \leq l$. Let β be a nice assignment in $\mathcal{N}$. If $i(\beta) \neq h$, then $d(b_h,b_{h+1}) = d(a_h,a_{h+1})$, hence there is an element $b_h < d < b_{h+1}$ such that $d(b_h,d) = d(c_h,c)$ and $d(d,b_{h+1}) = d(c,c_{h+1})$. Clearly $\beta(d/j) \in \mathcal{N}'$, moreover we have $\delta(\beta(d/j)) = \delta(\beta)$.

On the other hand, it may happen that $i(\beta) = h$. Note that there is exactly one such assignment β, let us denote it by β_h. Now $d(b_h,b_{h+1}) = d(a_h,a_{h+1})-1$. There are elements d and e such that $d(b_h,d) = d(a_h,c)-1$, $d(d,b_{h+1}) = d(c,a_{h+1})$, $d(b_h,e) = d(a_h,c)$ and $d(e,b_{h+1}) = d(c,a_{h+1})-1$. Let $\beta' = \beta_h(d/j)$ and $\beta'' = \beta_h(e/j)$. Then $\beta,\beta' \in \mathcal{N}'$, and we have

$$\delta(\beta') + \delta(\beta'') = d(b_h,d) + d(e,b_{h+1}) = d(a_h,a_{h+1}) - 2 = \delta(\beta_h) - 1.$$

Hence

$$(2\delta(\beta')+1) + (2\delta(\beta'')+1) = 2(\delta(\beta_h)-1) + 2 = (2\delta(\beta_h)+1) - 1.$$

Therefore we have

$$N(\mathfrak{B}') = \sum_{\beta'\in\mathcal{N}'}(2\delta(\beta')+1) = \sum_{\beta\in\mathcal{N}'\setminus\{b_h\}}(2\delta(\beta)+1)+((2\delta(\beta_h)+1)-1) = N(\mathfrak{B})-1.$$

□

Lemma 28. *If $m < N(\mathfrak{B})$, then player **II** has a winning strategy in* $\mathrm{EFB}^{\exists}_m(\mathfrak{A}_0,\mathfrak{B}_0)$.

Proof. We prove this by induction on m. The case $m = 1$ follows directly from Lemma 26. Now suppose the claim is true for all $k < m$. We play the game $\mathrm{EFB}^{\exists}_m(\mathfrak{A},\mathfrak{B})$. Since the game is existential, he cannot make right supplementing moves, therefore $\mathfrak{A}$ is always a singleton set. Since $\mathfrak{A}$ is a singleton set, there is no point for **I** to make a left splitting move. He is left with two options. Let us look at them in turn.

Assume **I** first makes a right splitting move. He splits $\mathfrak{B}$ into $\mathfrak{C}$ and $\mathfrak{D}$, he chooses numbers u and v such that $u+v = m$. Since $u+v = m < N(\mathfrak{B}) \leq N(\mathfrak{C}) + N(\mathfrak{D})$, we have either $u < N(\mathfrak{C})$ or $v < N(\mathfrak{D})$. By the induction hypothesis, player **II** has a winning strategy either in the position $(\mathfrak{A},\mathfrak{C},u)$ or in the position $(\mathfrak{A},\mathfrak{D},v)$. Let her choose this piece and apply the winning strategy in the rest of the match. She wins this game.

Assume **I** first makes a left supplementing move. He picks a choice function F on $\mathfrak{A}$. The game continues from the position $(\mathfrak{A}(F/j),\mathfrak{B}(\star/j),m-1))$. Let us denote $\mathfrak{B}(\star/j)$ by $\mathfrak{B}'$. By Lemma 27, we have that $m-1 < N(\mathfrak{B})-1 = N(\mathfrak{B}')-1$. By the induction hypothesis, player **II** has a winning strategy in the remaining game. □

With these lemmas we can now prove the claim at the beginning of this section. Consider the sets $\mathfrak{A}_0 = \{(\mathcal{A}, \emptyset)\}$ and $\mathfrak{B}_0 = \{(\mathcal{B}, \emptyset)\}$. Clearly $\delta(\emptyset) = n = 1$, therefore $N(\mathfrak{B}_0) = 2n-1$. By Lemma 28 player **II** has a winning strategy in the game $\mathrm{EFB}^{\exists}_m(\mathfrak{A}_0, \mathfrak{B}_0)$ for all $m < 2n-1$. We may now apply Corollary 21 to establish the following result.

Proposition 29 (Hella-Väänänen)**.** *If ϕ is an existential first order sentence expressing the property that the length of a linear order is at least n, then the size of ϕ is at least $2n-1$.*

5.2 The Propositional Complexity of Finiteness

In this section we give an application of the propositional fragment EFB^P_α of the new game EFB_α. Let us recall Example 6 in Section 3. Here we consider a special case of this example. Let $\mathfrak{A}$ be the class of structures $(\mathbb{N}, A), A \subset \mathbb{N}$. Let S be the following subset of $\mathbb{P}(\mathbb{N})$:

$$S = \{A \subseteq \mathbb{N} | A \text{ is finite}\}.$$

Consider the class of structures

$$\mathfrak{A}_S = \{(\mathbb{N}, A) | A \in S\}$$

and its complement in $\mathfrak{A}$

$$\mathfrak{B}_S = \{(\mathbb{N}, A) | A \notin S\}.$$

The classes of structures $\mathfrak{A}_S$ and $\mathfrak{B}_S$ are separated by the following sentence

$$\Phi_S = \bigvee_{A \in S} \phi_A,$$

where

$$\phi_A = \bigwedge_{n \in \omega} P^n_A$$

$$P^n_A = \begin{cases} P(c_n) & \text{if } n \in A, \\ \neg P(c_n) & \text{if } n \notin A. \end{cases}$$

In other words, the sentence Φ_S defines the property 'P is a finite set' in $\mathfrak{A}$. Since S is a countable set, the sentence Φ_S is in $L_{\omega_1\omega}$. Each sentence ϕ_A has size ω. The sentence Φ_S has size

$$\mathrm{s}(\Phi_S) = \sum_{A \in S}^{\#} \mathrm{s}(\phi_A) = \sup_{n \in \omega} \omega \cdot n = \omega \cdot \omega.$$

Note that Φ_S is a quantifier-free formula. A natural question is: Is Φ_S the minimal quantifier-free formula separating $\mathfrak{A}_S$ and $\mathfrak{B}_S$? With the help of the game EFB^P_α, we can give an affirmative answer to this question.

To make the notation easier, we translate the problem into another language. In what follows we identify a structure $(\mathbb{N}, A)$ with the characteristic function $h_A \in 2^\omega$ of A:

$$h_A(n) = \begin{cases} 1 & \text{if } n \in A \\ 0 & \text{if } n \notin A. \end{cases}$$

A function $h_A \in 2^\omega$ is an *infinite binary string*. In the context of binary strings, the class $\mathfrak{A}_S$ is identified with the class

$$\{h \in 2^\omega | h^{-1}(1) \text{ is finite}\}$$

and $\mathfrak{B}_S$ with the class

$$\{h \in 2^\omega | h^{-1}(1) \text{ is infinite}\}.$$

Let $2^{<\omega}$ denote the set of partial functions $\omega \to 2$ with a finite domain. For an infinite binary string $h \in 2^\omega$, we call a finite binary string $g \in 2^{<\omega}$ a *finite segment* of h if $g \subset h$. In such a case we also say that h *agrees* with g and h *extends* g. In the rest of this section we consider the game EFB^P_α played on classes of infinite binary strings.

In an infinite splitting move, player **I** splits $\mathfrak{A}$ (or $\mathfrak{B}$) into infinitely many pieces. We say that an infinite splitting move is a *proper infinite splitting move* if there are *infinitely* many pieces that are non-empty. If player **I** splits $\mathfrak{A}$ (or $\mathfrak{B}$) into infinitely many pieces among which only finitely many are non-empty, we call it a *degenerate infinite splitting move*. Strange as it seems, there is nothing in the rules that forbids **I** playing in this way. A degenerate infinite splitting move is closer in nature to a finite series of finite splitting moves. Henceforth we refer to both a degenerate infinite splitting move and a finite series of finite splitting moves also as a finite splitting move, as long as ambiguity does not arise.

Theorem 30. *Player **II** has a winning strategy in* $\mathrm{EFB}^P_\alpha(\mathfrak{A}_S, \mathfrak{B}_S)$ *for all* $\alpha < \omega \cdot \omega$.

Let α be an ordinal less than $\omega \cdot \omega$. We describe the following strategy for player **II** in the game $\mathrm{EFB}^P_\alpha(\mathfrak{A}_S, \mathfrak{B}_S)$. Note that in this game only splitting moves can be played, not supplementing moves.

Player II's strategy: Let $(\mathfrak{A}, \mathfrak{B}, \gamma)$ denote a game position. In each round, player **II** makes sure that

$$\begin{aligned} \exists f \in 2^{<\omega} \Bigg(& \forall h \in \mathfrak{A}(h \supset f) \wedge \\ & \forall h \in \mathfrak{B}(h \supset f) \wedge \\ & \forall g \in 2^{<\omega} \Big(g \supset f \to \big(\exists h \in \mathfrak{A}(h \supset g) \wedge \exists h' \in \mathfrak{B}(h' \supset g)\big)\Big)\Bigg). \end{aligned} \tag{10}$$

She maintains this strategy until she sees an opportunity to finish off the game directly.

Intuitively, condition (10) says that at any stage of the game, we can always find a finite segment f such that all structures in $\mathfrak{A}$ and $\mathfrak{B}$ agree with f. Moreover, for any finite extension g of f, there are structure h and h' in $\mathfrak{A}$ and $\mathfrak{B}$ respectively that agree with g. We will specify later what counts as an opportunity for **II** to win the game directly.

There are several things that we need to prove in order to establish Theorem 30.

Lemma 31. *Condition (10) is true in the starting position* $(\mathfrak{A}_S, \mathfrak{B}_S, \alpha)$.

Proof. Let f be the empty function. □

Lemma 32. *If condition (10) is true in position* $(\mathfrak{A}, \mathfrak{B}, \gamma)$*, then no atomic formula separates* $\mathfrak{A}$ *and* $\mathfrak{B}$*.*

Proof. There is an atomic formula separating $\mathfrak{A}$ and $\mathfrak{B}$ if and only if there exists $g \in 2^{<\omega}$ such that $|\mathrm{dom}(g)| = 1$, $g \subset h$ for all $h \in \mathfrak{A}$, and $g \not\subset h'$ for all $h' \in \mathfrak{B}$. We claim that condition (10) guarantees that this situation does not occur. There are the following three possibilities, let us consider them one by one. Firstly, if $g \subset f$, then by condition (10) all structures in $\mathfrak{A}$ and $\mathfrak{B}$ agree with g. Secondly, if g disagrees with f, namely if $\mathrm{dom}(g) \subseteq \mathrm{dom}(f)$ yet $g \not\subset f$, then by condition (10) again all structures in $\mathfrak{A}$ and $\mathfrak{B}$ disagree with g. Finally, suppose $\mathrm{dom}(g) \cap \mathrm{dom}(f) = \emptyset$. The finite segment $g \cup f$ is an extension of f. Therefore by condition (10) there are $h \in \mathfrak{A}$ and $h' \in \mathfrak{B}$ that agree with $g \cup f$ respectively. In particular they agree with g. □

Lemma 33. *If player **I** makes a finite splitting move in the generalized sense, namely if player **I** makes a usual finite splitting move or a degenerate infinite splitting move, player **II** can maintain condition (10) to the next round.*

Proof. Suppose the game is in position $(\mathfrak{A}, \mathfrak{B}, \gamma)$ and condition (10) holds. Suppose player **I** makes a finite left splitting move in the generalized sense. He represents $\mathfrak{A}$ as a union $\mathfrak{A}_1 \cup \ldots \cup \mathfrak{A}_n$ of n pieces. He also picks ordinals $\gamma_1, \ldots, \gamma_n$. We claim that there must exist a finite segment $f' \supset f$ and a piece $\mathfrak{A}_k$ such that

$$\forall g \in 2^{<\omega}\big(g \supset f' \to \exists h \in \mathfrak{A}_k(h \supset g)\big). \tag{11}$$

Suppose this is not the case for $\mathfrak{A}_1, \ldots, \mathfrak{A}_{n-1}$. Then we have

$$\forall f' \supset f \exists g \in 2^{<\omega}\left(g \supset f \wedge \forall h \in 2^{\omega}(h \supset g \to h \notin \mathfrak{A}_i)\right) \tag{12}$$

for $i = 1, \ldots, n-1$. Now let $f' = f$, apply condition (12) to $\mathfrak{A}_1$:

$$\exists g_1 \supset f \forall h \in 2^{\omega}\big(h \supset g_1 \to h \notin \mathfrak{A}_1\big).$$

This means that

$$\exists g_1 \supset f \forall h \in 2^\omega \Big((h \supset g_1 \wedge h \in \mathfrak{A}) \to h \in \mathfrak{A}_2 \cup \ldots \cup \mathfrak{A}_n \Big).$$

Now let $f' = g_1$, apply condition (12) to $\mathfrak{A}_2$:

$$\exists g_2 \supset g_1 \forall h \in 2^\omega \Big((h \supset g_2 \wedge h \in \mathfrak{A}) \to h \in \mathfrak{A}_3 \cup \ldots \cup \mathfrak{A}_n \Big).$$

Iterate this process, we get

$$\exists g_{n-1} \supset g_{n-2} \forall h \in 2^\omega \Big((h \supset g_{n-1} \wedge h \in \mathfrak{A}) \to h \in \mathfrak{A}_n \Big).$$

This means that condition (11) is true for $\mathfrak{A}_n$ when we take g_{n-1} for f'. Therefore the previous claim is true.

Now let player **II** pick this piece $\mathfrak{A}_k$ and the corresponding ordinal γ_k. What remains to be done is a cosmetic surgery on $\mathfrak{A}_k$ and $\mathfrak{B}$. We throw away the structures that do not agree with f'. Let

$$\mathfrak{A}' = \{h \in \mathfrak{A}_k | h \supset f'\}$$

and let

$$\mathfrak{B}' = \{h' \in \mathfrak{B} | h' \supset f'\}.$$

By Corollary 24, this can always be done. If **II** has a winning strategy in this new position, she also has one in the old position. It is easy to check that condition (10) holds for $\mathfrak{A}'$ and $\mathfrak{B}'$. Let the game continue from the position $(\mathfrak{A}', \mathfrak{B}', \gamma_k)$.

The case where **I** plays a finite right splitting move is similar. □

Lemma 34. *Suppose condition (10) is true in position* $(\mathfrak{A}, \mathfrak{B}, \gamma)$ *with* $\gamma < \omega \cdot \omega$. *If player* ***I*** *makes a proper infinite splitting move, then player* ***II*** *has a winning strategy in the rest of the game.*

Proof. Suppose the game is in position $(\mathfrak{A}, \mathfrak{B}, \gamma)$ and condition (10) is true. Suppose player **I** makes a proper infinite splitting move. He represents $\mathfrak{A}$ as a union $\bigcup_{i \in \omega} \mathfrak{A}_i$, in which there are infinitely many non-empty pieces. He also chooses ordinals $\gamma_i, i \in \omega$ such that $\sum^{\#}_{i \in \omega} \gamma_i \leq \gamma < \omega \cdot \omega$. Note that there can only be finitely many $i \in \omega$ such that γ_i is an infinite ordinal, for otherwise the natural sum $\sum^{\#}_{i \in \omega} \gamma_i$ would be greater or equal to $\omega \cdot \omega$, which is a contradiction. On the other hand there are infinitely many $i \in \omega$ such that $\mathfrak{A}_i$ is non-empty. Therefore there must exist $k \in \omega$ such that $\mathfrak{A}_k$ is non-empty and γ_k is finite.

If this situation occurs, we ask player **II** to jump out of the the strategy prescribed above and go on to win the game directly. Let player **II** pick this piece $\mathfrak{A}_k$ and γ_k. The game continues from the position $(\mathfrak{A}_k, \mathfrak{B}, \gamma_k)$. Apply Corollary 24 again and suppose $\mathfrak{A}_k$ consists of a single structure h_A. Note that since γ_k is a finite number, in the rest of the game **I** can only make finite right splitting moves.

Suppose **I** makes a finite right splitting move. He represents $\mathfrak{B}$ as a union $\mathfrak{B}_1 \cup \ldots \cup \mathfrak{B}_n$ of n pieces. For any natural number m, let h_m be the function such that h_m agrees with h_A on the first m elements of ω, and that $h_m(j) = 1$ for $j > m$. It is clear that the function h_m is in $\mathfrak{B}$. In other words, we have

$$\forall m \in \omega \; \exists h_m \in \mathfrak{B}(h_m \restriction m = h_A \restriction m). \tag{13}$$

Among the n pieces $\mathfrak{B}_1, \ldots, \mathfrak{B}_n$, there must be one piece $\mathfrak{B}_l$ that contains h_m for infinitely many $m \in \omega$. Let player **II** pick this piece $\mathfrak{B}_l$. She uses the same strategy if **I** makes another finite right splitting move. She makes sure that at any stage of the remaining game, the piece $\mathfrak{B}_k$ at hand always contain h_m for infinitely many $m \in \omega$. This guarantees that no atomic formula separates $\mathfrak{A}_k$ and $\mathfrak{B}$. Player **II** wins the game when the rank runs up. □

Now we are in a position to prove Theorem 30. If player **II** sees an opportunity as in Lemma 34, she goes on to win the game as described in the lemma. If she doesn't, she hangs on in the game by maintaining condition (10). This keeps her away from losing. The game terminates after finitely many rounds, and **II** will eventually prevail.

Theorem 35. *If ϕ is an quantifier-free L-formula in $L_{\omega_1\omega}$ separating $\mathfrak{A}_S$ and $\mathfrak{B}_S$, then the size of ϕ is at least $\omega \cdot \omega$.*

Proof. The claim follows immediately from Corollary 25 and Theorem 30. □

Note however that if quantifiers are allowed, then $\mathfrak{A}_S$ and $\mathfrak{B}_S$ can be separated by a much smaller formula:

$$\theta = \bigvee_{n\in\omega} \exists x_1 \ldots \exists x_n \forall y \big(P(y) \to y = x_1 \vee \ldots \vee y = x_n\big).$$

This formula has size ω only.

The result in Theorem 35 can be generalized to a wider range of properties. The only place in the above proof where we use the exact construction of $\mathfrak{A}_S$ and $\mathfrak{B}_S$ is in Lemma 31, namely in showing that condition (10) is satisfied in the initial position. Everything else follows from condition (10) itself, and we no longer need to care about what exactly $\mathfrak{A}_S$ and $\mathfrak{B}_S$ are. Let us consider all properties that validate Lemma 31.

Definition 14. Let $\mathcal{P} \subseteq 2^\omega$ be a property of infinite binary strings. We say that $\mathcal{P}$ is *dense* if for all $g \in 2^{<\omega}$, there is $h \in \mathcal{P}$ that extends g.

Note that 'f has finitely many zeros' ($\mathfrak{A}_S$) and 'f has infinitely many zeros' ($\mathfrak{B}_S$) are both dense properties. Theorem 35 holds for all dense properties.

Theorem 36. *Let $\mathcal{P}_1$ and $\mathcal{P}_2$ be two dense properties of infinite binary strings in 2^ω. If ϕ is an quantifier-free L-formula in $L_{\omega_1\omega}$ separating $\mathcal{P}_1$ and $\mathcal{P}_2$, then the size of ϕ is at least $\omega \cdot \omega$.*

Proof. It follows from the definition of dense properties that Lemma 31 holds for $\mathcal{P}_1$ and $\mathcal{P}_2$. The rest of the proof is *verbatim* the same as the proof of Theorem 35. □

Examples of dense properties of infinite binary strings include the following.

Example 8. 1. $\mathcal{P}_1 = \{f \in 2^\omega | f^{-1}(1) \text{ is finite}\}$ 'f has finitely many ones'.

$\mathcal{P}_2 = \{f \in 2^\omega | f^{-1}(0) \text{ is finite}\}$ 'f' has finitely many zeros'.

2. $\mathcal{P}_1 = \{f \in 2^\omega || f^{-1}(1)| \text{ is an odd number}\}$ 'f has an odd number of ones'.

$\mathcal{P}_2 = \{f \in 2^\omega || f^{-1}(1)| \text{ is an even number}\}$ 'f' has an even number of ones'.

By Theorem 36, none of these pairs of properties can be separated by an $L_{\omega_1\omega}$ sentence of size less than $\omega \cdot \omega$.

What we have presented above is an application of the game EFB_α to the quantifier-free fragment of predicate logic $L_{\omega_1\omega}$. This story can equally be told in pure propositional logic with countable conjunctions and disjunctions. Let L be the propositional language with countably many propositional variables $p_i, i \in \omega$. Let $L^P_{\omega_1\omega}$ denote the propositional logic with countable conjunctions and disjunctions of formulas. For an infinite binary string $f \in 2^\omega$, we say that $f \models p_i$ if $f(i) = 1$ and $f \not\models p_i$ if $f(i) = 0$. The satisfaction relation for boolean connectives is defined as usual. For two classes of infinite binary strings $\mathfrak{A}$, $\mathfrak{B}$ and a propositional formula ϕ, we say that ϕ separates $\mathfrak{A}$ and $\mathfrak{B}$, denoted by $(\mathfrak{A}, \mathfrak{B}) \models \phi$, if $f \models \phi$ for all $f \in \mathfrak{A}$ and $g \not\models \phi$ for all $g \in \mathfrak{B}$. The size of an $L^P_{\omega_1\omega}$ formula is defined as in Definition 13, disregarding the clauses for quantifiers. The results that we have proved in this section can be easily translated into the propositional language $L^P_{\omega_1\omega}$.

Theorem 37. *Let $\mathcal{P}_1$ and $\mathcal{P}_2$ be two dense properties of infinite binary strings in 2^ω. If ϕ is a propositional L-formula in $L^P_{\omega_1\omega}$ separating $\mathcal{P}_1$ and $\mathcal{P}_2$, then the size of ϕ is at least $\omega \cdot \omega$.*

6 Conclusions and Further Questions

In this thesis we introduced an Ehrenfeucht-Fraïssé Game for the infinitary logic $L_{\omega_1\omega}$. Let us recall the position from which we started this quest. The usual Ehrenfeucht-Fraïssé Games fail to characterize separation in $L_{\omega_1\omega}$. The reason is that, we claim, the usual Ehrenfeucht-Fraïssé Games only reflect our understanding of the behavior of quantifiers in logic. Consequently the usual Ehrenfeucht-Fraïssé Games most naturally characterize separation in the logic $L_{\infty\omega}$, where there is basically no constraint on the boolean size of formulas.

In the logic $L_{\omega_1\omega}$ the situation is different. The logic $L_{\omega_1\omega}$ is defined not only by reference to countability of the quantifier rank but also by reference to the countability of *conjunctions* and *disjunctions* in the formula. Based on a game for propositional and first order logic in [9], we defined a new Ehrenfeucht-Fraïssé Game EFB_α for $L_{\omega_1\omega}$. The supplementing moves in this game can be seen as the heritage of the usual Ehrenfeucht-Fraïssé Games. The splitting moves are an innovation, aiming at reflecting the nature of boolean connectives. Correspondingly, we introduced a measure of the size of $L_{\omega_1\omega}$ formulas. We proved that the game EFB_α characterizes separation in $L_{\omega_1\omega}$ up to size α.

Moreover, we also want to show that the game EFB_α is indeed useful. In fact, if we define a game just for the sake of proving an adequacy theorem for this game, it would not be of much value. Games are supposed to help us solving problems in logic. The complexity result about infinite binary strings in Section 5.2 is an application of the propositional fragment of EFB_α. This is the first propositional complexity result for $L_{\omega_1\omega}$. An interesting topic for future research would be to search for applications of the full game. It would be also interesting to get higher propositional complexity results for $L_{\omega_1\omega}$, e.g. on the level of ω^3.

Acknowledgements

Foremost, I would like to thank my thesis supervisor Jouko Väänänen for his active supervision of this thesis project. I own much of this thesis to his insights, some of which took me several months to realize their importance. I am especially grateful for the effort he spent on reading and commenting on my drafts. I also thank him for his patience and encouragement during this thesis project. Working with Jouko I feel like more than just another student. I thank my supervisor Benedikt Löwe for his encouragement and his advices on improving the draft. I also thank him for the wonderful lectures I had with him on set theory.

I thank my academic mentor Dick de Jongh for the many conversations we had about my study plan and about academic life in general. I thank Johan van Benthem and Yanjing Wang, whom I met back in 2009 while I was an undergraduate mathematics student knowing almost nothing about logic. Without them I would most probably not have come to study logic in Amsterdam.

My sincere thanks also go to my fellow students in the MoL program: Haitao Cai, Katya Garmash, Aleks Knoks, Adam Lesnikowski, Zhenhao Li, Alessandra Marra, Riccardo Pinosio, Fabio Zanasi, and many others. Studying together with them has broaden my horizon. Interacting with them has made my stay in Amsterdam a memorable experience.

References

[1] Jon Barwise. The role of the omitting types theorem in infinitary logic. *Arch. Math. Logik Grundlag*, 21(1-2):55–68, 1981.

[2] Evert W. Beth. Remarks on natural deduction. *Indagationes Mathematicae*, 17:322–325, 1955.

[3] Evert W. Beth. *Semantic Entailment and Formal Derivability*. Mededelingen der koninklijke Nederlandse Akademie van Wetenschappen, afd. Letterkunde. Nieuwe Reeks, Deel 18, No.13. N. V. Noord-Hollandsche Uitgevers Maatschappij, Amsterdam, 1955.

[4] Andrzej Ehrenfeucht. Application of games to some problems of mathematical logic. *Bulletin de l'Acadámie Polonaise des Science, Série des Sciences Mathématiques, Astronomiques et Physiques Cl. III.,5*, IV:35–37, 1957.

[5] Andrzej Ehrenfeucht. An application of games to the completeness problem for formalized theories. *Fundamenta Mathematicae*, 49:129–141, 1960/1961.

[6] Roland Fraïssé. Sur quelques classifications des relations, basées sur des isomorphismes restreints. ii. application aux relations d'ordre, et construction d'exemples montrant que ces classifications sont distinctes. *Publ. Sci. Univ. Alger. Sér. A.*, 2:273–295, 1957.

[7] David Gale and F. M. Stewart. Infinite games with perfect information. In *Contributions to the Theory of Games*, number 2 in Annals of Mathematics Studies, pages 245–266. Princeton University Press, Princeton, N. J., 1953.

[8] Yuri Gurevich. Toward logic tailored for computational complexity. In *Computation and Proof Theory*, volume 1104 of *Lecture Notes in Mathematics*, pages 175–216. Springer-Verlag, Berlin, 1984.

[9] Lauri Hella and Jouko Väänänen. The size of a formula as a measure of complexity. `http://arxiv.org/abs/1208.4803`, 2012.

[10] Leon Henkin. Some remarks on infinitely long formulas. In *Infinitistic Methods (Proc. Sympos. Foundations of Math., Warsaw, 1959)*, pages 167–183. Pergamon, Oxford, 1961.

[11] Wilfrid Hodges. *Model Theory*, volume 42 of *Encyclopedia of Mathematics and its Applications*. Cambridge University Press, Cambridge, 1993.

[12] Thomas Jech. *Set Theory*. Perspectives in Mathematical Logic. Springer-Verlag, Berlin, second edition, 1997.

[13] Carol Karp. *Languages with Expressions of Infinite Length*. North-Holland, Amsterdam, 1964.

[14] Carol Karp. *Finite-quantifier equivalence*, pages 407–412. North-Holland, Amsterdam, 1965.

[15] Jerome Keisler. *Model Theory for Infinitary Logic*, volume 62 of *Studies in Logic and the Foundations of Mathematics*. North-Holland, Amsterdam, 1971.

[16] David Kueker. Back-and-forth arguments and infinitary logic. In *Infinitary Logic: in Memoriam Carol Karp*, volume 492 of *Lecture Notes in Mathematics*, pages 17–71. Springer-Verlag, Berlin, 1975.

[17] E. G. K. Lopez-Escobar. An interpolation theorem for denumerably long formulas. *Fundamenta Mathematicae*, 57:253–272, 1965.

[18] Jerzy Łoś. On the extending of models. *Fundamenta Mathematicae*, 42:38–54, 1955.

[19] Dana Scott. *Logic with denumerably long formulas and finite strings of quantifiers*, pages 329–341. North-Holland, Amsterdam, 1965.

[20] Dana Scott and Alfred Tarski. The sentential calculus with infinitely long expressions. *Colloquium Mathematicum*, 6:165–170, 1958.

[21] Alfred Tarski. Contributions to the theory of models, i. *Indagationes Mathematicae*, 16:572–581, 1954.

[22] Alfred Tarski. Remarks on predicate logic with infinitely long expressions. *Colloquium Mathematicum*, 6:171–176, 1958.

[23] Jouko Väänänen. *Models and Games*. Number 132 in Cambridge Studies in Advanced Mathematics. Cambridge University Press, Cambridge, 2011.

Index