

# Model Comparison Games for Generalized Quantifiers

Antti Kuusisto<sup>1</sup>, Miguel Moreno<sup>1</sup>, and Matias Selin<sup>1</sup>

<sup>1</sup>Tampere University, Finland

## Abstract

We introduce two new model comparison games that characterize separability by first-order formulas with generalized quantifiers. One is built on the Ehrenfeucht–Fraïssé game and the other is a formula-size game.

## 1 Introduction

Due to the limited expressive power of first-order logic, several extensions of it have been studied, some of which are obtained by the addition of *generalized quantifiers*. Clearly, by increasing the expressive power of a logic, more differences between models can be expressed, i.e., the logic can separate more models. This leads to studying the characterization of structures separable in the given logic. Some authors have used games to study the separability of models: Kotlatis and Väänänen in [5] with pebble games, Hella and Väänänen in [3] with a formula-size game, and Haber and Shelah in [2] with the Ehrenfeucht–Fraïssé (EF) game.

In this article, we introduce variations of these games that characterize separability in  $\text{FO}(\mathcal{Q})$ , which is first-order logic whose quantifiers belong to a finite set  $\mathcal{Q}$  of generalized quantifiers.<sup>1</sup> We prove that the EF-game characterizes separability by quantifier depth, and that the formula-size game characterizes separability by formula size. When  $\mathcal{Q} = \{\exists, \forall\}$ , our logic coincides with ordinary first-order logic; our games, however, differ from the classical ones, as they are designed to handle arbitrary (and in particular, non-monotone) quantifiers.

Further, instead of using normal forms, as Hella and Väänänen in [3], our version of the formula-size game does not assume negation normal form, but instead treats negation as a regular logical operator that contributes to the length of the formula. We also introduce a notion of *weak separability*, where each pair of structures is separated individually rather than by a single uniform

---

<sup>1</sup>Note that our notation of  $\text{FO}(\mathcal{Q})$  thus differs from the literature [8, 1], since instead of *extending* FO we *replace* the existential and universal quantifiers with the quantifiers in  $\mathcal{Q}$ . Of course  $\mathcal{Q}$  may still contain  $\exists$  and  $\forall$ .

formula. We show that weak and strong separability coincide when separating finitely many structures, and give a game characterization of weak separability.

Some of the ideas for these games originate from the articles [6] and [7]. Using these intuitions, an extension of the *bisimulation game* to cover propositional logic extended with a finite collection of generalized *modalities* that strongly resembles the EF-game described in Section 3 was given in [4].

## 2 Preliminaries

### 2.1 Definitions and Notations

We denote the set of non-negative integers by  $\mathbb{N}$  and the set of positive integers by  $\mathbb{Z}_+$ . We denote the  $m$ -tuple  $(x_1, \dots, x_m)$  by boldface  $\mathbf{x}$  and the  $(m+1)$ -tuple  $(x_1, \dots, x_m, y)$  by  $\mathbf{xy}$ . For all  $1 \leq i \leq m$ , we denote the  $i$ th element of the tuple as  $\mathbf{x}(i) = x_i$ .

We consider relational models with a finite **vocabulary**  $\tau = \{R_1, \dots, R_n\}$ , where each  $R_i$  has an arity of  $k_i \in \mathbb{Z}_+$ . A  $\tau$ -**model** is a pair  $\mathfrak{A} = (A, T)$ , where  $A$  is a non-empty (possibly infinite) set called the **domain** of the model and  $T$  is a function that maps each  $k$ -ary relation symbol  $R \in \tau$  to a set of  $k$ -tuples of the domain, i.e.,  $R^{\mathfrak{A}} := T(R) \in \mathcal{P}(A^k)$ . Models  $\mathfrak{A}, \mathfrak{B}$  are always written in Fraktur typeface, their domains  $A, B$  in regular typeface and classes of models  $\mathcal{A}, \mathcal{B}$  in calligraphic typeface.

We denote the countably infinite set of **variable symbols** or simply **variables** by  $\text{VAR} := \{x_i \mid i \in \mathbb{Z}_+\}$ . Often, we use the usual meta-variables  $x, y, z$  to denote variables. Let  $X \subseteq \text{VAR}$  be finite and  $\mathfrak{A}$  be a model. A (**variable**) **assignment** over  $\mathfrak{A}$  is a (possibly empty) function  $f : X \rightarrow A$ . If  $x \in \text{VAR}$  and  $a \in A$ , then the assignment  $f_x^a : X \cup \{x\} \rightarrow A$  is defined such that  $f_x^a(y) = f(y)$  if  $y \neq x$  and  $f_x^a(y) = a$  if  $y = x$ . By just  $\frac{a}{x}$  we denote the assignment  $\{(x, a)\}$ .

For ease of exposition, in the main section of the paper we only consider the simplest class of generalized quantifiers: those that quantify a single formula that binds a single variable. We can generalize this notion by allowing the quantification of multiple formulas (higher width) that bind multiple variables (higher type). All results in this paper hold for quantifiers with arbitrary, finite width and type; see Appendix A for the formal definitions and proofs.

A **generalized quantifier of width 1 and type 1** is an isomorphism-closed class  $\mathcal{Q}$  of structures  $(D, P)$  where  $P \subseteq D$ . Let  $\mathcal{Q}$  be a finite collection of such generalized quantifiers. The set of  $\tau$ -formulas of **first-order logic with the quantifiers**  $\mathcal{Q}$  ( $\text{FO}(\mathcal{Q})$ ) is generated by the following grammar:

$$\varphi ::= x_1 = x_2 \mid R(x_1, \dots, x_k) \mid \neg\varphi \mid \varphi \wedge \varphi \mid Qx_1 \varphi,$$

where  $x_1, x_2, \dots, x_k \in \text{VAR}$ ,  $R \in \tau$  has arity  $k$  and  $Q \in \mathcal{Q}$ . It is important to keep in mind that the set  $\mathcal{Q}$  *does not necessarily include the existential or universal quantifier*, which is where our notation of  $\text{FO}(\mathcal{Q})$  differs from the literature [8, 1]. Formulas constructed with only the first two rules are called **atomic**, and as usual, disjunction  $\varphi \vee \psi$  is defined as shorthand for  $\neg(\neg\varphi \wedge \neg\psi)$ .

We say that a variable  $x$  is **free** in a formula  $\varphi$  if it does not appear under the scope of a quantifier  $Qx$ . We denote a formula  $\varphi$  that has (at least) the free variables  $x_1, \dots, x_k$  by  $\varphi(x_1, \dots, x_k)$ . A formula with no free variables is called a **sentence**.

Let  $\mathfrak{A}$  be a  $\tau$ -model and  $f$  be an assignment over  $\mathfrak{A}$  whose domain includes all free variables of the formula being evaluated. The semantics of  $\text{FO}(\mathcal{Q})$  are as follows:

1.  $\mathfrak{A}, f \models x_1 = x_2 \iff f(x_1) = f(x_2)$ .
2.  $\mathfrak{A}, f \models R(x_1, \dots, x_k) \iff (f(x_1), \dots, f(x_k)) \in R^{\mathfrak{A}}$ .
3.  $\mathfrak{A}, f \models \neg\varphi \iff \mathfrak{A}, f \not\models \varphi$ .
4.  $\mathfrak{A}, f \models \varphi \wedge \psi \iff \mathfrak{A}, f \models \varphi$  and  $\mathfrak{A}, f \models \psi$ .
5.  $\mathfrak{A}, f \models Qx\varphi \iff (A, \|\varphi\|_x^{\mathfrak{A}, f}) \in Q$ ,

where

$$\|\varphi\|_x^{\mathfrak{A}, f} := \{a \mid \mathfrak{A}, f \frac{a}{x} \models \varphi\}$$

is the **extension** of  $\varphi$  relative to  $(\mathfrak{A}, f)$  and  $x$ , i.e., the set of points that can be interpreted as  $x$  such that  $\varphi$  is satisfied.

We also denote the pair  $(\mathfrak{A}, \emptyset)$  as simply  $\mathfrak{A}$ , which gives rise to denoting  $\mathfrak{A} \models \varphi$  if  $\mathfrak{A}, \emptyset \models \varphi$ ; of course, this notation only makes sense if  $\varphi$  is a sentence. Two formulas  $\varphi$  and  $\psi$  are **equivalent** if  $\mathfrak{A}, f \models \varphi \iff \mathfrak{A}, f \models \psi$  for all models  $\mathfrak{A}$  and assignments  $f$ .

Finally, we say that  $(\mathfrak{A}, f)$  and  $(\mathfrak{B}, g)$  are **separable** in  $\text{FO}(\mathcal{Q})$  if there is a formula  $\varphi \in \text{FO}(\mathcal{Q})$  such that  $\mathfrak{A}, f \models \varphi$  and  $\mathfrak{B}, g \not\models \varphi$ . Otherwise, we say that they are **equivalent** in  $\text{FO}(\mathcal{Q})$  and write  $\mathfrak{A}, f \equiv_{\text{FO}(\mathcal{Q})} \mathfrak{B}, g$ .

## 2.2 Types

In this subsection, we prove two useful results that show that any set of points closed under equivalence in  $\text{FO}(\mathcal{Q})$ , or more generally any finite collection of such sets, can be captured by a formula of  $\text{FO}(\mathcal{Q})$ .

The **quantifier depth** of an  $\text{FO}(\mathcal{Q})$ -formula is the maximum number of nested quantifiers in the formula. We use  $\text{FO}(\mathcal{Q})^d$  to denote the set of formulas of  $\text{FO}(\mathcal{Q})$  that have a quantifier depth of at most  $d$ , and  $\equiv_{\text{FO}(\mathcal{Q})}^d$  to denote equivalence in  $\text{FO}(\mathcal{Q})^d$ .

Let  $\mathfrak{A}$  be a  $\tau$ -model with assignment  $f$ , and let  $x \in \text{VAR}$ . We say that a subset  $X \subseteq A$  of the domain is **closed under**  $\equiv_{\text{FO}(\mathcal{Q})}^d$  relative to  $(\mathfrak{A}, f)$  and  $x$  if  $a \in X$  and  $\mathfrak{A}, f \frac{a}{x} \equiv_{\text{FO}(\mathcal{Q})}^d \mathfrak{A}, f \frac{a'}{x}$  implies  $a' \in X$ , and that  $X$  is **definable** by a formula  $\varphi$  relative to  $(\mathfrak{A}, f)$  and  $x$  if  $\|\varphi\|_x^{\mathfrak{A}, f} = X$ .

For each  $a \in A$ , the  **$d$ -type** of  $a$  relative to  $(\mathfrak{A}, f)$  and  $x$  is the equivalence class

$$[a]_x^d := \{b \in A \mid \mathfrak{A}, f \frac{a}{x} \equiv_{\text{FO}(\mathcal{Q})}^d \mathfrak{A}, f \frac{b}{x}\}.$$

Since the definition of a  $d$ -type is directly based on equivalence in  $\text{FO}(\mathcal{Q})^d$ , it is rather easy to see that all  $d$ -types are indeed definable in  $\text{FO}(\mathcal{Q})^d$ . The only non-obvious part is showing that the defining formula is finitary, which it is for all models with finite index.

**Lemma 2.1.** *Let  $\mathfrak{A}$  be a  $\tau$ -model with assignment  $f$ , where  $\tau$  is finite. Then for each  $d \in \mathbb{N}$  and  $a \in A$ , the  $d$ -type  $[a]_x^d$  is definable by a formula  $\chi_a^d(x) \in \text{FO}(\mathcal{Q})^d$ .*

*Proof.* We proceed by induction on  $d$ .

**Base case.** Let  $\Phi$  be a maximal set of non-equivalent atomic formulas of  $\text{FO}(\mathcal{Q})$  with free variables in  $\text{dom}(f) \cup \{x\}$ . Since  $\text{dom}(f)$  and  $\tau$  are both finite,  $\Phi$  is also finite. Let  $\Phi_a := \{\varphi \in \Phi \mid \mathfrak{A}, f_x^a \models \varphi\}$ . Now, the desired formula is

$$\chi_a^0(x) := \bigwedge_{\varphi \in \Phi_a} \varphi \wedge \bigwedge_{\varphi \in \Phi \setminus \Phi_a} \neg \varphi,$$

since  $\mathfrak{A}, f_x^b \models \chi_a^0(x)$  if and only if  $b \in [a]_x^0$ .

**Induction case.** Suppose the class  $[a]_x^d$  is definable by a formula  $\chi_a^d(x) \in \text{FO}(\mathcal{Q})^d$  for some  $d \in \mathbb{N}$ . Each  $b \in [a]_x^d \setminus [a]_x^{d+1}$  is separated from  $a$  by some formula of depth  $d+1$ , which is, without loss of generality, of the form  $\theta_b(x) := Q_b y \psi_b(x, y)$ , where  $Q_b \in \mathcal{Q}$  and  $\psi_b(x, y) \in \text{FO}(\mathcal{Q})^d$ ; here, “separated” means that  $\mathfrak{A}, f_x^v \models \theta_b(x)$  for each  $v \in [a]_x^{d+1}$  and  $\mathfrak{A}, f_x^{v'} \not\models \theta_b(x)$  for each  $v' \in [b]_x^{d+1}$ .

We verify that  $[a]_x^d$  splits into finitely many  $(d+1)$ -types, so that the construction below is finitary. By the induction hypothesis, the  $d$ -types for pairs  $(x, y)$  are finitely many, and hence there are finitely many non-equivalent formulas  $\psi_b(x, y) \in \text{FO}(\mathcal{Q})^d$ . Combined with the finiteness of  $\mathcal{Q}$ , the number of non-equivalent formulas  $\theta_b(x)$  is finite, so  $[a]_x^d$  contains finitely many  $(d+1)$ -types.

Let  $b_1, \dots, b_k$  be representatives of the  $(d+1)$ -types in  $[a]_x^d \setminus [a]_x^{d+1}$ . The desired formula is thus

$$\chi_a^{d+1}(x) := \chi_a^d(x) \wedge \bigwedge_{i=1}^k \theta_{b_i}(x),$$

since  $\mathfrak{A}, f_x^v \models \chi_a^{d+1}(x)$  if and only if  $v \in [a]_x^d$  (by the induction hypothesis) and  $v \notin [b]_x^{d+1}$  for each  $b \in [a]_x^d \setminus [a]_x^{d+1}$ .  $\square$

**Corollary 2.2.** *Let  $\mathfrak{A}$  be a  $\tau$ -model with assignment  $f$ , where  $\tau$  is finite. Then each  $X \subseteq A$  closed under  $\equiv_{\text{FO}(\mathcal{Q})}^d$  is definable by a formula  $\theta(x) \in \text{FO}(\mathcal{Q})^d$ .*

*Proof.* By the closure assumption,  $X$  is a union of  $d$ -types. Let  $a_1, \dots, a_k \in X$  be representatives of the  $d$ -types in  $X$ . Then  $\theta(x) := \bigvee_{i=1}^k \chi_{a_i}^d(x)$  defines  $X$ .  $\square$

### 3 The EF-Game for Generalized Quantifiers

In this section, we present our EF-game that characterizes separability in  $\text{FO}(\mathcal{Q})$  by quantifier depth. The core idea is to modify the regular EF-game such that

instead of individual points, *sets* of points belonging to the quantifiers in  $\mathcal{Q}$  are chosen.

We say that a pair of assignments  $(f, g)$  with  $\text{dom}(f) = \text{dom}(g)$  **induces a partial isomorphism** between  $\tau$ -models  $\mathfrak{A}$  and  $\mathfrak{B}$  if  $(\mathfrak{A}, f)$  and  $(\mathfrak{B}, g)$  agree on all atomic formulas, i.e.,  $\mathfrak{A}, f \models \alpha$  if and only if  $\mathfrak{B}, g \models \alpha$  for every atomic formula  $\alpha$  with free variables in  $\text{dom}(f)$ .

**Definition 3.1.** Let  $\mathfrak{A}$  and  $\mathfrak{B}$  be  $\tau$ -models, where  $\tau$  is finite, and let  $f$  and  $g$  be (possibly empty) assignments over  $\mathfrak{A}$  and  $\mathfrak{B}$  respectively with  $\text{dom}(f) = \text{dom}(g)$ . The  $\text{EF}\{\mathcal{Q}\}(\mathfrak{A}, \mathfrak{B}, f, g)$ -game is a two-player game that starts from the position  $(\mathfrak{A}, \mathfrak{B}, f, g)$ , with Player  $\text{I}$  starting as the **attacker** and Player  $\text{II}$  starting as the **defender**. The  $i$ th round proceeds from a position  $(\mathfrak{M}, \mathfrak{N}, h, h')$ , where  $h$  and  $h'$  are assignments over  $\mathfrak{M}$  and  $\mathfrak{N}$  with  $\text{dom}(h) = \text{dom}(h')$ , as follows:

1. The attacker chooses a quantifier  $Q \in \mathcal{Q}$  and a variable  $x \in \text{VAR}$ .
2. The attacker chooses a **witness set**  $X$  from the domain of either model (without loss of generality, suppose  $X \subseteq M$ ) such that  $(M, X) \in Q$  and a **spillover set**  $P \subseteq N$  from the domain of the other model.<sup>2</sup>
3. The defender chooses a corresponding witness set  $X' \subseteq N$  from the domain of the other model such that  $(N, X') \in Q$  and  $P \subseteq X'$ .
4. The attacker does one of the following:
  - (a) Chooses  $v' \in N \setminus X'$  and  $v \in X$ . The players swap roles. The next position is  $(\mathfrak{M}, \mathfrak{N}, h \frac{v}{x}, h' \frac{v'}{x})$ .
  - (b) Chooses  $v' \in X'$ . The defender now either
    - chooses  $v \in X$ , after which the next position is  $(\mathfrak{M}, \mathfrak{N}, h \frac{v}{x}, h' \frac{v'}{x})$ ,
    - or
    - chooses  $v \in P$ , after which the next position is  $(\mathfrak{N}, \mathfrak{M}, \frac{v}{x}, \frac{v'}{x})$ .
5. Let  $(\mathfrak{M}', \mathfrak{N}', h_*, h'_*)$  be the position determined in the previous step. If the pair  $(h_*, h'_*)$  does not induce a partial isomorphism between  $\mathfrak{M}'$  and  $\mathfrak{N}'$ , then the game ends and the attacker wins. Otherwise a new round begins from this position.

At any point during steps 2–3, immediately after a set is chosen, the opposing player may **contest** that choice instead of letting the round continue normally. When a contestation occurs, the remaining steps of the round are skipped and replaced as follows:

- **Contesting a witness set.** After  $X$  is chosen in step 2 or  $X'$  is chosen in step 3, the opposing player may contest that the set breaks equivalence by choosing  $u \in Y$  and  $u' \in W \setminus Y$ , where  $Y$  is the contested set and  $W$  is

---

<sup>2</sup>Another way to think about model selection is to give the attacker the option to switch which model is called  $\mathfrak{M}$  and which  $\mathfrak{N}$  before he chooses the sets.

its domain. If the contesting player is the attacker, then the players swap roles. The next position is  $(\mathfrak{W}, \mathfrak{W}, h_W \frac{u}{x}, h_W \frac{u'}{x})$ , where  $h_W = h$  if  $W = M$  and  $h_W = h'$  if  $W = N$ .

- **Contesting a spillover set.** After  $P$  is chosen in step 2, the opposing player may contest that it contains a type realized in  $M$  by choosing  $v' \in P$  and  $v \in M$ . The next position is  $(\mathfrak{M}, \mathfrak{M}, h \frac{v}{x}, h' \frac{v'}{x})$ .

In both cases, the game then proceeds to step 5: the partial isomorphism check is performed on the new position, and if it passes, a new round begins from that position.

Notice that Player I always plays the first move of the first round of the game, but each subsequent round can be started by either Player I or Player III, depending on which one of them is the attacker. If at any point a player must make a move but cannot do so (e.g. there does not exist an  $X \subseteq M$  such that  $(M, X) \in Q$ ), then the other player wins immediately.

The intuition for the game is as follows (cf. Figure 1). First, if the formula separating  $\mathfrak{A}$  and  $\mathfrak{B}$  is  $Qx \varphi$ , then the strategy for the attacker is to choose  $X$  to be precisely the extension of  $\varphi$  in  $\mathfrak{A}$ . The defender cannot respond with an equivalent  $X'$ , since either it is too small (meaning it does not capture enough points that satisfy  $\varphi$ , which the attacker points out with his first option) or too large (meaning that some point in it does not satisfy  $\varphi$ , which the attacker points out with his second option). The point of the spillover set is to force the defender to also include those types that satisfy  $\varphi$  but are not realized in  $\mathfrak{A}$ . The point of the contestation rules is to ensure that neither player cheats by choosing a witness set that is not expressible in  $\text{FO}(\mathcal{Q})$ , and that the attacker's choice of the spillover set includes only types not realized in  $\mathfrak{A}$ .

By restricting the amount of rounds played  $d$ , we obtain a limit for the quantifier depth of the separating formula. The winner of the  $\text{EF}_d\{\mathcal{Q}\}(\mathfrak{A}, \mathfrak{B}, f, g)$ -game is defined recursively as follows:

1. The  $\text{EF}_0\{\mathcal{Q}\}(\mathfrak{A}, \mathfrak{B}, f, g)$ -game is won by Player I if the pair  $(f, g)$  does not induce a partial isomorphism between  $\mathfrak{A}$  and  $\mathfrak{B}$ . Otherwise, it is won by Player III.
2. The  $\text{EF}_{d+1}\{\mathcal{Q}\}(\mathfrak{A}, \mathfrak{B}, f, g)$ -game is won by Player III if Player I has not won the game after  $d + 1$  rounds have been played.

**Lemma 3.2.** *Player I has a winning strategy in the  $\text{EF}_0\{\mathcal{Q}\}(\mathfrak{A}, \mathfrak{B}, f, g)$ -game if and only if  $(\mathfrak{A}, f)$  and  $(\mathfrak{B}, g)$  are separable in  $\text{FO}(\mathcal{Q})^0$ .*

*Proof.* ( $\implies$ ): If Player I wins the 0-round game, then  $(f, g)$  does not induce a partial isomorphism. This means there exists an atomic formula  $\alpha$  with free variables in  $\text{dom}(f)$  such that  $\mathfrak{A}, f \models \alpha$  and  $\mathfrak{B}, g \not\models \alpha$  (or vice versa, in which case  $\neg\alpha$  separates them). Thus  $(\mathfrak{A}, f)$  and  $(\mathfrak{B}, g)$  are separable in  $\text{FO}(\mathcal{Q})^0$ .

( $\impliedby$ ): If  $(f, g)$  induces a partial isomorphism, then  $(\mathfrak{A}, f)$  and  $(\mathfrak{B}, g)$  agree on all atomic formulas, and hence on all Boolean combinations thereof. Thus they are not separable in  $\text{FO}(\mathcal{Q})^0$ .  $\square$

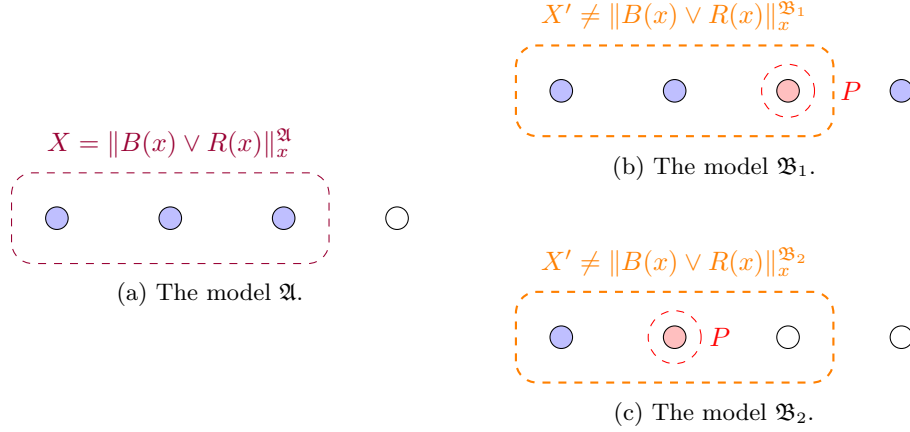


Figure 1: Suppose the quantifier  $\exists_{=3}$ , meaning “there exist exactly three”, is in  $\mathcal{Q}$ , and suppose  $\tau = \{R, B\}$  where  $R$  and  $B$  are both unary. The model  $\mathfrak{A}$  is separated from both  $\mathfrak{B}_1$  and  $\mathfrak{B}_2$  by the formula  $\exists_{=3}x (B(x) \vee R(x))$ , where nodes in the interpretation of  $B$  are coloured in blue and nodes in the interpretation of  $R$  in red. Player II thus has a winning strategy in the respective  $\text{EF}\{\mathcal{Q}\}$ -games by choosing the witness set  $X$  and the spillover set  $P$ , since Player III’s witness set  $X'$  is then either too small in  $\mathfrak{B}_1$  (since it must exclude a point that is blue) or too large in  $\mathfrak{B}_2$  (since it must include a point that is neither blue nor red).

We now make the induction hypothesis that the following two statements are equivalent for some  $d \in \mathbb{N}$ , for all  $\tau$ -models  $\mathfrak{A}, \mathfrak{B}$ , and for all assignments  $f, g$  over  $\mathfrak{A}$  and  $\mathfrak{B}$  with  $\text{dom}(f) = \text{dom}(g)$ :

1. Player II has a winning strategy in the  $\text{EF}_d\{\mathcal{Q}\}(\mathfrak{A}, \mathfrak{B}, f, g)$ -game.
2.  $(\mathfrak{A}, f)$  and  $(\mathfrak{B}, g)$  are separable in  $\text{FO}(\mathcal{Q})^d$ .

**Lemma 3.3.** *If Player II has a winning strategy in the  $\text{EF}_{d+1}\{\mathcal{Q}\}(\mathfrak{A}, \mathfrak{B}, f, g)$ -game, then  $(\mathfrak{A}, f)$  and  $(\mathfrak{B}, g)$  are separable in  $\text{FO}(\mathcal{Q})^{d+1}$ .*

*Proof.* Assume contrapositively that  $(\mathfrak{A}, f)$  and  $(\mathfrak{B}, g)$  are equivalent in  $\text{FO}(\mathcal{Q})^{d+1}$ . Suppose Player II chooses a quantifier  $Q \in \mathcal{Q}$ , a variable  $x \in \text{VAR}$ , a witness set  $X \subseteq A$  and a spillover set  $P \subseteq B$ .

We begin by observing that the first contestation rule forces the players to choose witness sets that are closed under  $\equiv_{\text{FO}(\mathcal{Q})}^d$  relative to their respective models and assignments and  $x$ , i.e., are unions of  $d$ -types. Otherwise, Player III could pick  $v \in X$  and  $v' \in A \setminus X$  such that  $\mathfrak{A}, f \stackrel{v}{x} \equiv_{\text{FO}(\mathcal{Q})}^d \mathfrak{A}, f \stackrel{v'}{x}$ , granting them a winning strategy by the induction hypothesis (and the same applies for Player II contesting  $X'$ ).

By Corollary 2.2, the witness set  $X$  is thus defined by the disjunction of the formulas defining the  $d$ -types contained in  $X$ ; formally,  $X = \|\theta\|_x^{\mathfrak{A}, f}$ , where  $\theta := \bigvee_{1 \leq i \leq k} \chi_{a_i}^d$ , the points  $a_1, \dots, a_k$  are representatives of the  $d$ -types realized

in  $X$ , and each  $\chi_{a_i}^d$  is the defining formula of  $[a_i]_x^d$  as in the proof of Lemma 2.1. Similarly, there exists a formula  $\psi \in \text{FO}(\mathcal{Q})^d$  that defines the closure of  $P$  under  $\equiv_{\text{FO}(\mathcal{Q})}^d$ , i.e., such that

$$\|\psi\|_x^{\mathfrak{B},g} = P \cup \{v \in B \mid \mathfrak{B}, g \frac{v}{x} \equiv_{\text{FO}(\mathcal{Q})}^d \mathfrak{B}, g \frac{v'}{x} \text{ for some } v' \in P\}.$$

By the second contestation rule, the extension of  $\psi$  in  $\mathfrak{A}$  is empty; otherwise there would exist a point  $v \in A$  such that  $\mathfrak{A}, f \frac{v}{x} \equiv_{\text{FO}(\mathcal{Q})}^d \mathfrak{B}, g \frac{v'}{x}$  for some  $v' \in P$ , and Player III could choose the pair  $(v, v')$  and obtain a winning strategy by the induction hypothesis.

Since  $\|\theta\|_x^{\mathfrak{A},f} = X$  and  $\|\psi\|_x^{\mathfrak{A},f} = \emptyset$ , we have  $\|\theta \vee \psi\|_x^{\mathfrak{A},f} = X$ , which combined with the fact that  $(A, X) \in \mathcal{Q}$  implies that  $\mathfrak{A}, f \models Qx(\theta \vee \psi)$ . Recalling the assumption that  $(\mathfrak{A}, f)$  and  $(\mathfrak{B}, g)$  are equivalent in  $\text{FO}(\mathcal{Q})^{d+1}$ , we thus also have  $\mathfrak{B}, g \models Qx(\theta \vee \psi)$ . The strategy for Player III is now to respond with the witness set  $X' = \|\theta \vee \psi\|_x^{\mathfrak{B},g}$ , to which Player I has two possible responses:

1. Suppose Player I chooses  $v' \in B \setminus X'$  and  $v \in X$  and the players swap roles. Since  $\mathfrak{A}, f \frac{v}{x} \models \theta$  and  $\mathfrak{B}, g \frac{v'}{x} \not\models \theta$ , then  $(\mathfrak{A}, f \frac{v}{x})$  and  $(\mathfrak{B}, g \frac{v'}{x})$  are separable in  $\text{FO}(\mathcal{Q})^d$ . By the induction hypothesis, the attacker, who is now Player III, thus has a winning strategy from this position.
2. Suppose Player I chooses  $v' \in X'$ .
  - (a) If  $v' \in \|\psi\|_x^{\mathfrak{B},g}$ , then since  $\|\psi\|_x^{\mathfrak{B},g}$  is the closure of  $P$  under  $\equiv_{\text{FO}(\mathcal{Q})}^d$ , there exists a  $v \in P$  such that  $\mathfrak{B}, g \frac{v'}{x} \equiv_{\text{FO}(\mathcal{Q})}^d \mathfrak{B}, g \frac{v}{x}$ . The defender chooses this  $v$ , and the game continues from  $(\mathfrak{B}, \mathfrak{B}, \frac{v}{x}, \frac{v'}{x})$ . By the induction hypothesis, Player III has a winning strategy from this position.
  - (b) If  $v' \in \|\theta\|_x^{\mathfrak{B},g}$ , then  $v'$  satisfies the defining formula of some  $d$ -type present in  $\theta$ ; denote it by  $\chi_{a_i}^d$ . Then Player III simply selects the point  $v \in X$  that satisfies the same formula  $\chi_{a_i}^d$ . Since these points belong to the same  $d$ -type, we have  $\mathfrak{A}, f \frac{v}{x} \equiv_{\text{FO}(\mathcal{Q})}^d \mathfrak{B}, g \frac{v'}{x}$ , which, by the induction hypothesis, means Player III has a winning strategy from this position.  $\square$

**Lemma 3.4.** *If  $(\mathfrak{A}, f)$  and  $(\mathfrak{B}, g)$  are separable in  $\text{FO}(\mathcal{Q})^{d+1}$ , then Player I has a winning strategy in the  $\text{EF}_{d+1}\{\mathcal{Q}\}(\mathfrak{A}, \mathfrak{B}, f, g)$ -game.*

*Proof.* Without loss of generality, assume that  $\mathfrak{A}, f \models Qx\varphi$  and  $\mathfrak{B}, g \models \neg Qx\varphi$  for some  $Q \in \mathcal{Q}$ ,  $x \in \text{VAR}$  and  $\varphi \in \text{FO}(\mathcal{Q})^d$ . Player I begins the winning strategy by choosing  $Q$ , the variable  $x$ , the witness set  $X := \|\varphi\|_x^{\mathfrak{A},f}$ , and the spillover set

$$P := \{v' \in \|\varphi\|_x^{\mathfrak{B},g} \mid \mathfrak{B}, g \frac{v'}{x} \not\equiv_{\text{FO}(\mathcal{Q})}^d \mathfrak{A}, f \frac{v}{x} \text{ for all } v \in A\}.$$

Clearly  $X$  is closed under  $\equiv_{\text{FO}(\mathcal{Q})}^d$ , which means Player III contesting the choice of  $X$  would give Player II a winning strategy. Moreover, if Player III would contest the choice of  $P$ , then since  $\mathfrak{B}, g_{\frac{v'}{x}} \not\equiv_{\text{FO}(\mathcal{Q})}^d \mathfrak{A}, f_{\frac{v}{x}}$  for all  $v' \in P$  and  $v \in A$ , by the induction hypothesis, Player II would have a winning strategy.

Player III must thus choose a witness set  $X'$  of their own. Since  $\mathfrak{B}, g \models \neg Qx \varphi$ , we know that  $(B, \|\varphi\|_x^{\mathfrak{B}, g}) \notin Q$  and hence  $X' \neq \|\varphi\|_x^{\mathfrak{B}, g}$ . We now have two cases:

1. Suppose there exists a  $v' \in X'$  such that  $\mathfrak{B}, g_{\frac{v'}{x}} \models \neg \varphi$ . By the definition of  $P$ , we have  $v' \in X' \setminus P$ . Player II now chooses  $v'$  and Player III is forced to respond by choosing  $v \in X$  such that  $\mathfrak{A}, f_{\frac{v}{x}} \models \varphi$  or  $v \in P$  such that  $\mathfrak{B}, g_{\frac{v}{x}} \models \varphi$ . Since in both cases  $v$  is separated from  $v'$  by  $\varphi \in \text{FO}(\mathcal{Q})^d$ , by the induction hypothesis, we conclude that the attacker, i.e. Player II, has a winning strategy from this position.
2. Suppose there exists  $v' \in B \setminus X'$  such that  $\mathfrak{B}, g_{\frac{v'}{x}} \models \varphi$ . Since  $P \subseteq X'$ , we have  $v' \notin P$  and by the definition of  $P$ , we have  $\mathfrak{B}, g_{\frac{v'}{x}} \equiv_{\text{FO}(\mathcal{Q})}^d \mathfrak{A}, f_{\frac{v}{x}}$  for some  $v \in A$ . Hence  $\mathfrak{A}, f_{\frac{v}{x}} \models \varphi$ , so  $v \in X$ . Player II now chooses  $v'$  and  $v$  and becomes the defender. By the induction hypothesis, the defender, i.e. Player II, has a winning strategy from this position.  $\square$

Since  $\mathfrak{A}$  and  $\mathfrak{B}$  are separable in  $\text{FO}(\mathcal{Q})$  if and only if there is a  $d \in \mathbb{N}$  such that  $\mathfrak{A}$  and  $\mathfrak{B}$  are separable in  $\text{FO}(\mathcal{Q})^d$ , we now have the desired result.

**Theorem 3.5.** *Player II has a winning strategy in the  $\text{EF}\{\mathcal{Q}\}(\mathfrak{A}, \mathfrak{B}, \emptyset, \emptyset)$ -game if and only if  $\mathfrak{A}$  and  $\mathfrak{B}$  are separable in  $\text{FO}(\mathcal{Q})$ .*

## 4 The Formula-Size Game for Generalized Quantifiers

In this section, we present a game that characterizes a more fine-grained distinction between models by measuring the smallest *size* of their separating formula. This is a modification of the formula-size game originally introduced for FO by Hella and Väänänen [3]. Our game extends this to  $\text{FO}(\mathcal{Q})$  and also considers the effect of negation on the size of the formula.

The **size**  $s(\varphi)$  of an  $\text{FO}(\mathcal{Q})$ -formula  $\varphi$  is defined recursively as follows:

1.  $s(\varphi) := 1$  for atomic  $\varphi$ .
2.  $s(\neg \varphi) := s(\varphi) + 1$ .
3.  $s(\varphi \wedge \psi) := s(\varphi) + s(\psi)$ .
4.  $s(Qx \varphi) := s(\varphi) + 1$  for all  $Q \in \mathcal{Q}$ .

## 4.1 Playing Between Classes

We call a pair  $(\mathfrak{A}, f)$ , where  $\mathfrak{A}$  is a model and  $f$  is an assignment over  $\mathfrak{A}$ , a  $\tau$ -**pair**. Consider *classes* of these pairs. We say that two such classes  $\mathcal{A}$  and  $\mathcal{B}$  are **separable** in  $\text{FO}(\mathcal{Q})$  if there is a formula  $\varphi \in \text{FO}(\mathcal{Q})$  such that  $\mathfrak{A}, f \models \varphi$  for all  $(\mathfrak{A}, f) \in \mathcal{A}$  and  $\mathfrak{B}, g \not\models \varphi$  for all  $(\mathfrak{B}, g) \in \mathcal{B}$ . Notice that  $\mathcal{A}$  and  $\mathcal{B}$  are possibly empty, and that the empty class is separable from any other class, including itself.

**Definition 4.1.** Let  $\mathcal{A}$  and  $\mathcal{B}$  be classes of  $\tau$ -pairs and  $s \in \mathbb{Z}_+$  be a **budget**. The  $\text{EF}_s\{\mathcal{Q}\}(\mathcal{A}, \mathcal{B})$ -game is a two-player game that starts in the position  $(s, \mathcal{A}, \mathcal{B})$ . The  $i$ th round proceeds from the position  $(s_i, \mathcal{A}_i, \mathcal{B}_i)$  as follows:

1. If there is an atomic  $\text{FO}(\mathcal{Q})$ -formula that separates  $\mathcal{A}_i$  and  $\mathcal{B}_i$ , then the game ends and Player I wins.
2. Otherwise, if  $s_i = 1$ , then the game ends and Player III wins.
3. If neither of the above conditions holds, then Player I chooses one of the following three options:
  - (a) **Swap classes (negation).** A new round begins from the position  $(s_i - 1, \mathcal{B}_i, \mathcal{A}_i)$ .
  - (b) **Right splitting move (conjunction).** Player I chooses  $u, v \in \mathbb{Z}_+$  such that  $u+v = s_i$ , and chooses (possibly overlapping) sets  $\mathcal{C}, \mathcal{D} \subseteq \mathcal{B}_i$  such that  $\mathcal{C} \cup \mathcal{D} = \mathcal{B}_i$ . Player III then responds by choosing whether the next round starts from the position  $(u, \mathcal{A}_i, \mathcal{C})$  or from the position  $(v, \mathcal{A}_i, \mathcal{D})$ .
  - (c) **Supplementing move (quantifier).** Player I first chooses a quantifier  $Q \in \mathcal{Q}$  and a variable symbol  $x \in \text{VAR}$ . Player I then chooses a function  $P : \mathcal{A}_i \cup \mathcal{B}_i \rightarrow \bigcup\{\mathcal{P}(C) \mid (C, h) \in \mathcal{A}_i \cup \mathcal{B}_i\}$  such that  $P(C, h) \in \mathcal{P}(C)$  and

$$\begin{aligned} (A, P(\mathfrak{A}, f)) &\in Q \text{ for all } (\mathfrak{A}, f) \in \mathcal{A}_i \text{ and} \\ (B, P(\mathfrak{B}, g)) &\notin Q \text{ for all } (\mathfrak{B}, g) \in \mathcal{B}_i. \end{aligned}$$

(If such a  $P$  does not exist, Player I cannot choose this move.) The next round then starts from the position  $(s_i - 1, \mathcal{C}^+, \mathcal{C}^-)$ , where

$$\begin{aligned} \mathcal{C}^+ &:= \{(C, h \frac{v}{x}) \mid (C, h) \in \mathcal{A}_i \cup \mathcal{B}_i \text{ and } v \in P(C, h)\} \text{ and} \\ \mathcal{C}^- &:= \{(C, h \frac{v}{x}) \mid (C, h) \in \mathcal{A}_i \cup \mathcal{B}_i \text{ and } v \in C \setminus P(C, h)\}. \end{aligned}$$

Intuitively, Player I wants to move precisely those models (from both classes) that satisfy the formula  $Qx\varphi$  to the left and those models that don't to the right. So, for each model, he picks a subset of the domain that can be captured by the quantifier; models extended with an atom from within that subset go to the left

and those extended with an atom from outside that subset go to the right. If  $Qx\varphi$  indeed separates the classes, Player II can pick  $P$  such that the split works as intended. And if Player II can pick  $P$  such that the split works as intended (i.e., if Player II has a winning strategy in the game), by the definitions of the  $\mathcal{C}$ -classes and  $P$ , there must be a formula that separates the classes. See the proof below.

**Theorem 4.2.** *Player II has a winning strategy in the  $\text{EF}_s\{Q\}(\mathcal{A}, \mathcal{B})$ -game if and only if  $\mathcal{A}$  and  $\mathcal{B}$  are separable by a formula of  $\text{FO}(\mathcal{Q})$  of size  $\leq s$ .*

*Proof. Base case.* Let  $s = 1$ . By definition, Player II wins the  $\text{EF}_1\{Q\}(\mathcal{A}, \mathcal{B})$ -game if and only if there is an atomic formula  $\varphi$  (which is thus of size 1) that separates  $\mathcal{A}$  and  $\mathcal{B}$ .

Suppose now that  $s > 1$  and that the statement holds for all  $l < s$ .

**Induction case: negation.** Suppose Player II has a winning strategy for the  $\text{EF}_s\{Q\}(\mathcal{A}, \mathcal{B})$ -game that begins by swapping the classes. This is true if and only if Player II has a winning strategy in the  $\text{EF}_{s-1}\{Q\}(\mathcal{B}, \mathcal{A})$ -game. By the induction hypothesis, this is equivalent to  $\mathcal{B}$  and  $\mathcal{A}$  being separated by some formula  $\varphi$  of size  $\leq s - 1$ . By the definition of separation, this is equivalent to  $\mathcal{A}$  and  $\mathcal{B}$  being separated by  $\neg\varphi$ , which is of size  $\leq s$ .

**Induction case: conjunction.** We first prove the  $\implies$  direction. Suppose Player II has a winning strategy for the  $s$ -game that begins by choosing  $u, v \in \mathbb{Z}_+$  such that  $u + v = s$  and  $\mathcal{C}, \mathcal{D} \subseteq \mathcal{B}$  such that  $\mathcal{C} \cup \mathcal{D} = \mathcal{B}$ . Since the strategy is winning, Player II has a winning strategy in both  $\text{EF}_u\{Q\}(\mathcal{A}, \mathcal{C})$  and  $\text{EF}_v\{Q\}(\mathcal{A}, \mathcal{D})$ . By the induction hypothesis, there thus exist formulas  $\varphi$  and  $\psi$  such that  $s(\varphi) \leq u$  and  $s(\psi) \leq v$ , and such that  $\mathcal{A}$  and  $\mathcal{C}$  are separated by  $\varphi$  and  $\mathcal{A}$  and  $\mathcal{D}$  are separated by  $\psi$ . By the definition of separation, this means  $\mathfrak{A}, f \models \varphi$  and  $\mathfrak{A}, f \models \psi$ , and hence  $\mathfrak{A}, f \models \varphi \wedge \psi$  for all  $(\mathfrak{A}, f) \in \mathcal{A}$ . Moreover,  $\mathfrak{B}, g \not\models \varphi$  for all  $(\mathfrak{B}, g) \in \mathcal{C}$  and  $\mathfrak{B}, g \not\models \psi$  for all  $(\mathfrak{B}, g) \in \mathcal{D}$ . Thus  $\mathfrak{B}, g \not\models \varphi \wedge \psi$  for all  $(\mathfrak{B}, g) \in \mathcal{B}$ . This means  $\mathcal{A}$  and  $\mathcal{B}$  are separated by  $\varphi \wedge \psi$  of length  $s(\varphi \wedge \psi) = s(\varphi) + s(\psi) \leq u + v = s$ .

We then prove the  $\impliedby$  direction. Suppose  $\varphi \wedge \psi$  of size  $\leq s$  separates  $\mathcal{A}$  and  $\mathcal{B}$ . Let  $\mathcal{C} := \{(\mathfrak{B}, g) \in \mathcal{B} \mid \mathfrak{B}, g \not\models \varphi\}$  and  $\mathcal{D} := \{(\mathfrak{B}, g) \in \mathcal{B} \mid \mathfrak{B}, g \not\models \psi\}$ . Since  $\mathfrak{B}, g \not\models \varphi \wedge \psi$  for all  $(\mathfrak{B}, g) \in \mathcal{B}$ , we have  $\mathcal{B} = \mathcal{C} \cup \mathcal{D}$ . Moreover, since  $\mathfrak{A}, f \models \varphi \wedge \psi$  for all  $(\mathfrak{A}, f) \in \mathcal{A}$ , we see that  $\varphi$  separates  $\mathcal{A}$  and  $\mathcal{C}$  and  $\psi$  separates  $\mathcal{A}$  and  $\mathcal{D}$ . Also, since  $s(\varphi \wedge \psi) \leq s$ , there exist  $u, v \in \mathbb{Z}_+$  such that  $u \geq s(\varphi)$ ,  $v \geq s(\psi)$  and  $u + v = s$ . By the induction hypothesis, Player II has a winning strategy in the games  $\text{EF}_u(\mathcal{A}, \mathcal{C})$  and  $\text{EF}_v(\mathcal{A}, \mathcal{D})$ . Player II thus has a winning strategy in the  $\text{EF}_s\{Q\}(\mathcal{A}, \mathcal{B})$ -game by choosing  $u, v, \mathcal{C}$  and  $\mathcal{D}$  as his first move.

**Induction case: quantifier.** We first prove the  $\implies$  direction. Suppose Player II has a winning strategy for the  $s$ -game that begins by choosing  $Q \in \mathcal{Q}$ , a variable  $x \in \text{VAR}$  and a function  $P$ , inducing the sets  $\mathcal{C}^+$  and  $\mathcal{C}^-$ . Since Player II now has a winning strategy for the  $\text{EF}_{s-1}\{Q\}(\mathcal{C}^+, \mathcal{C}^-)$ -game, by the induction hypothesis, there exists a formula  $\varphi$  of size  $\leq s - 1$  that separates  $\mathcal{C}^+$  and  $\mathcal{C}^-$ .

We first show that  $P(\mathfrak{C}, h) = \|\varphi\|_x^{\mathfrak{C}, h}$  for all  $(\mathfrak{C}, h) \in \mathcal{A} \cup \mathcal{B}$ . If  $v \in P(\mathfrak{C}, h)$ , then  $(\mathfrak{C}, h \frac{v}{x}) \in \mathcal{C}^+$  by definition, so  $\mathfrak{C}, h \frac{v}{x} \models \varphi$  by separation. Conversely, if  $v \notin P(\mathfrak{C}, h)$ , then  $(\mathfrak{C}, h \frac{v}{x}) \in \mathcal{C}^-$  by definition, so  $\mathfrak{C}, h \frac{v}{x} \not\models \varphi$  by separation.

Together, we thus have  $v \in P(\mathfrak{C}, h) \iff \mathfrak{C}, h \frac{v}{x} \models \varphi$ , establishing the desired equality.

It follows that for all  $(\mathfrak{A}, f) \in \mathcal{A}$ , we have  $P(\mathfrak{A}, f) = \|\varphi\|_x^{\mathfrak{A}, f}$ , and since  $(A, P(\mathfrak{A}, f)) \in Q$  by the definition of  $P$ , we conclude  $\mathfrak{A}, f \models Qx\varphi$ . Similarly, for all  $(\mathfrak{B}, g) \in \mathcal{B}$ , we have  $P(\mathfrak{B}, g) = \|\varphi\|_x^{\mathfrak{B}, g}$ , and since  $(B, P(\mathfrak{B}, g)) \notin Q$  by the definition of  $P$ , we conclude  $\mathfrak{B}, g \not\models Qx\varphi$ . Thus  $Qx\varphi$  of size  $\leq s$  separates  $\mathcal{A}$  and  $\mathcal{B}$ .

We then prove the  $\Leftarrow$  direction. Suppose  $Qx\varphi$  of size  $\leq s$  separates  $\mathcal{A}$  and  $\mathcal{B}$ . Player  $\text{II}$  chooses  $P$  such that  $P(\mathfrak{C}, h) := \|\varphi\|_x^{\mathfrak{C}, h}$  for all  $(\mathfrak{C}, h) \in \mathcal{A} \cup \mathcal{B}$ . This is a valid move, since for all  $(\mathfrak{A}, f) \in \mathcal{A}$ , we have  $\mathfrak{A}, f \models Qx\varphi$ , so  $(A, P(\mathfrak{A}, f)) \in Q$ , and for all  $(\mathfrak{B}, g) \in \mathcal{B}$ , we have  $\mathfrak{B}, g \not\models Qx\varphi$ , so  $(B, P(\mathfrak{B}, g)) \notin Q$ .

The game continues from the position  $(s-1, \mathcal{C}^+, \mathcal{C}^-)$ . Now, for any  $(\mathfrak{C}, h) \in \mathcal{A} \cup \mathcal{B}$  and  $v \in C$ , if  $v \in P(\mathfrak{C}, h)$ , then  $\mathfrak{C}, h \frac{v}{x} \models \varphi$  and  $(\mathfrak{C}, h \frac{v}{x}) \in \mathcal{C}^+$ , and if  $v \notin P(\mathfrak{C}, h)$ , then  $\mathfrak{C}, h \frac{v}{x} \not\models \varphi$  and  $(\mathfrak{C}, h \frac{v}{x}) \in \mathcal{C}^-$ . Thus  $\mathcal{C}^+$  and  $\mathcal{C}^-$  are separated by  $\varphi$ , and since  $\varphi$  is of size  $\leq s-1$ , by the induction hypothesis, Player  $\text{II}$  has a winning strategy from this position.  $\square$

## 4.2 Playing Between Models

There are multiple ways to interpret the statement “ $\mathcal{A}$  and  $\mathcal{B}$  are separable”, and the one we have used so far results in a rather strong condition. In this subsection, we investigate an alternative, weaker notion of separability. We say that two classes of  $\tau$ -pairs  $\mathcal{A}$  and  $\mathcal{B}$  are **weakly separable** in  $\text{FO}(\mathcal{Q})$  if for all  $(\mathfrak{A}, f) \in \mathcal{A}$  and  $(\mathfrak{B}, g) \in \mathcal{B}$ , there is a formula  $\varphi \in \text{FO}(\mathcal{Q})$  such that  $\mathfrak{A}, f \models \varphi$  and  $\mathfrak{B}, g \not\models \varphi$ . It is not hard to see that when the classes are infinite, separability implies weak separability but weak separability does not necessarily imply separability.

**Example 4.3.** Let  $\mathcal{Q} = \{\exists, \forall\}$  and  $\tau = \{P\}$ , where  $P$  is unary. For all  $n \in \mathbb{Z}_+$ , let  $\mathfrak{A}_n$  be the model with domain  $\mathbb{N}$  such that  $\mathfrak{A}_n \models \exists_{=n}x P(x)$ . (Note that  $\exists_{=n}$  can be expressed using ordinary first-order quantifiers.) Clearly the classes  $\{\mathfrak{A}_n \mid n \text{ is odd}\}$  and  $\{\mathfrak{A}_n \mid n \text{ is even}\}$  are weakly separable but not separable.

The reason why we played with possibly infinite model classes in the previous section was because we wanted to characterize specifically the stronger notion of separability. If we restrict to the finite case, these two notions of separability coincide. One way of showing this is simply to define a naïve version of the formula-size game where Player  $\text{III}$  first picks the model pair to separate, and the game is then played on singletons formed from that pair.

**Definition 4.4.** Let  $\mathcal{A}$  and  $\mathcal{B}$  be classes of  $\tau$ -pairs. The  $\text{EF}_s^*\{\mathcal{Q}\}(\mathcal{A}, \mathcal{B})$ -game is a two-player game that starts in the position  $(s, \mathcal{A}, \mathcal{B})$  as follows:

1. Player  $\text{III}$  chooses  $(\mathfrak{A}, f) \in \mathcal{A}$  and  $(\mathfrak{B}, g) \in \mathcal{B}$ .
2. Then, the  $\text{EF}_s\{\mathcal{Q}\}(\{(\mathfrak{A}, f)\}, \{(\mathfrak{B}, g)\})$ -game is played (with Player  $\text{II}$  making the first move) to determine the winner.

**Corollary 4.5.** *Player I has a winning strategy in the  $\text{EF}_s^*\{\mathcal{Q}\}(\mathcal{A}, \mathcal{B})$ -game if and only if  $\mathcal{A}$  and  $\mathcal{B}$  are weakly separable by formulas of  $\text{FO}(\mathcal{Q})$  of size  $\leq s$ .*

*Proof.* The following statements are equivalent, with the 2.  $\iff$  3. equivalence being an implication of Theorem 4.2:

1. Player I has a winning strategy in the  $\text{EF}_s^*\{\mathcal{Q}\}(\mathcal{A}, \mathcal{B})$ -game.
2. For each  $(\mathfrak{A}, f) \in \mathcal{A}$  and  $(\mathfrak{B}, g) \in \mathcal{B}$ , Player I has a winning strategy for the  $\text{EF}_s\{\mathcal{Q}\}(\{(\mathfrak{A}, f)\}, \{(\mathfrak{B}, g)\})$ -game.
3. For each  $(\mathfrak{A}, f) \in \mathcal{A}$  and  $(\mathfrak{B}, g) \in \mathcal{B}$ , the classes  $\{(\mathfrak{A}, f)\}$  and  $\{(\mathfrak{B}, g)\}$  are separable by a formula of  $\text{FO}(\mathcal{Q})$  of size  $\leq s$ .
4. For each  $(\mathfrak{A}, f) \in \mathcal{A}$  and  $(\mathfrak{B}, g) \in \mathcal{B}$ , there exists a formula  $\varphi$  of  $\text{FO}(\mathcal{Q})$  of size  $\leq s$  such that  $\mathfrak{A}, f \models \varphi$  and  $\mathfrak{B}, g \not\models \varphi$ .
5.  $\mathcal{A}$  and  $\mathcal{B}$  are weakly separable by formulas of  $\text{FO}(\mathcal{Q})$  of size  $\leq s$ .  $\square$

**Theorem 4.6.** *Let  $\mathcal{A}$  and  $\mathcal{B}$  be finite classes of  $\tau$ -pairs. Then they are separable in  $\text{FO}(\mathcal{Q})$  if and only if they are weakly separable in  $\text{FO}(\mathcal{Q})$ .*

*Proof.* If  $\mathcal{A}$  and  $\mathcal{B}$  are separable by a formula  $\varphi \in \text{FO}(\mathcal{Q})$ , then they are clearly weakly separable in  $\text{FO}(\mathcal{Q})$ , since choosing  $\varphi$  for each pair suffices.

Suppose then that  $\mathcal{A}$  and  $\mathcal{B}$  are weakly separable in  $\text{FO}(\mathcal{Q})$ . For each  $(\mathfrak{A}, f) \in \mathcal{A}$  and  $(\mathfrak{B}, g) \in \mathcal{B}$ , let  $\varphi_{\mathfrak{B}, g}^{\mathfrak{A}, f}$  be a formula that separates the  $\tau$ -pairs. Consider now the disjunction

$$\psi := \bigvee_{(\mathfrak{A}, f) \in \mathcal{A}} \bigwedge_{(\mathfrak{B}, g) \in \mathcal{B}} \varphi_{\mathfrak{B}, g}^{\mathfrak{A}, f},$$

which is a finitary formula of  $\text{FO}(\mathcal{Q})$  by the finiteness of  $\mathcal{A}$  and  $\mathcal{B}$ . Since by the definition of weak separation each  $(\mathfrak{A}, f) \in \mathcal{A}$  satisfies  $\varphi_{\mathfrak{B}, g}^{\mathfrak{A}, f}$  for every  $(\mathfrak{B}, g) \in \mathcal{B}$ , we know that each  $(\mathfrak{A}, f) \in \mathcal{A}$  satisfies the conjunction  $\bigwedge_{(\mathfrak{B}, g) \in \mathcal{B}} \varphi_{\mathfrak{B}, g}^{\mathfrak{A}, f}$  and hence the disjunction  $\psi$ . Conversely, for each  $(\mathfrak{B}, g) \in \mathcal{B}$  and  $(\mathfrak{A}, f) \in \mathcal{A}$ , the formula  $\varphi_{\mathfrak{B}, g}^{\mathfrak{A}, f}$  is false in  $(\mathfrak{B}, g)$ , and hence the conjunction  $\bigwedge_{(\mathfrak{B}, g) \in \mathcal{B}} \varphi_{\mathfrak{B}, g}^{\mathfrak{A}, f}$  is false in  $(\mathfrak{B}, g)$ . Since this holds for every  $(\mathfrak{A}, f) \in \mathcal{A}$ , the formula  $\psi$  is false in  $(\mathfrak{B}, g)$ . Thus  $\psi$  separates  $\mathcal{A}$  and  $\mathcal{B}$ .  $\square$

The game we used to prove this theorem is a rather heavy hammer, and does not give Player III much agency. We now present a more natural version of the game that is played directly between a pair of models. This game is more suited for showing lower-bound separability results between individual models.

**Definition 4.7.** Let  $(\mathfrak{A}, f)$  and  $(\mathfrak{B}, g)$  be  $\tau$ -pairs and  $s \in \mathbb{Z}_+$  be a budget. The  $\text{EF}_s\{\mathcal{Q}\}(\mathfrak{A}, f, \mathfrak{B}, g)$ -game is a two-player game that starts in the position  $(s, \mathfrak{A}, f, \mathfrak{B}, g)$ . The  $i$ th round proceeds from the position  $(s_i, \mathfrak{M}_i, h_i, \mathfrak{N}_i, h'_i)$  as follows:

1. If there is an atomic  $\text{FO}(\mathcal{Q})$ -formula that separates  $(\mathfrak{M}_i, h_i)$  and  $(\mathfrak{N}_i, h'_i)$ , then the game ends and Player I wins.

2. Otherwise, if  $s_i = 1$ , then the game ends and Player III wins.
3. If neither of the above conditions holds, then Player I chooses one of the following three options:
  - (a) **Swap models (negation).** A new round begins from the position  $(s_i - 1, \mathfrak{N}_i, h'_i, \mathfrak{M}_i, h_i)$ .
  - (b) **Split budget (conjunction).** Player I chooses  $u, v \in \mathbb{Z}_+$  such that  $u + v = s_i$ . Player III then responds by choosing whether the next round starts from the position  $(u, \mathfrak{M}_i, h_i, \mathfrak{N}_i, h'_i)$  or from the position  $(v, \mathfrak{M}_i, h_i, \mathfrak{N}_i, h'_i)$ .
  - (c) **Supplementing move (quantifier).** Player I first chooses a quantifier  $Q \in \mathcal{Q}$  and a variable symbol  $x \in \text{VAR}$ . Player I then chooses subsets  $M' \subseteq M$  and  $N' \subseteq N$  such that  $(M, M') \in Q$  and  $(N, N') \notin Q$ . (If no such choice exists, Player I cannot choose this move.) Player III then picks one of the following:
    - An element  $a \in M'$  and an element  $b \in N \setminus N'$ . The next position is  $(s_i - 1, \mathfrak{M}_i, h_i \frac{a}{x}, \mathfrak{N}_i, h'_i \frac{b}{x})$ .
    - An element  $a \in M \setminus M'$  and an element  $b \in N'$ . The next position is  $(s_i - 1, \mathfrak{N}_i, h'_i \frac{b}{x}, \mathfrak{M}_i, h_i \frac{a}{x})$ .
    - Two elements  $a \in M'$  and  $a' \in M \setminus M'$ . The next position is  $(s_i - 1, \mathfrak{M}_i, h_i \frac{a}{x}, \mathfrak{M}_i, h_i \frac{a'}{x})$ .
    - Two elements  $b \in N'$  and  $b' \in N \setminus N'$ . The next position is  $(s_i - 1, \mathfrak{N}_i, h'_i \frac{b}{x}, \mathfrak{N}_i, h'_i \frac{b'}{x})$ .

Note that in this game, Player I never benefits from choosing the budget splitting move. This is due to the fact that if  $(\mathfrak{A}, f)$  and  $(\mathfrak{B}, g)$  are separable by a formula  $\varphi \wedge \psi$ , then they are also separable by either  $\varphi$  or  $\psi$ .

**Lemma 4.8.** *If  $(\mathfrak{A}, f)$  and  $(\mathfrak{B}, g)$  are separable by a formula of  $\text{FO}(\mathcal{Q})$  of size  $\leq s$ , then Player I has a winning strategy in the  $\text{EF}_s\{\mathcal{Q}\}(\mathfrak{A}, f, \mathfrak{B}, g)$ -game.*

*Proof.* Induction on  $s$ . The base case as well as the negation and conjunction induction cases are as in the proof of Lemma 4.2, so it is sufficient to check the quantifier case.

Suppose  $Qx\varphi$  of size  $\leq s$  separates  $(\mathfrak{A}, f)$  from  $(\mathfrak{B}, g)$ . Player I chooses  $Q$ ,  $x$ ,  $M' := \|\varphi\|_x^{\mathfrak{A}, f}$  and  $N' := \|\varphi\|_x^{\mathfrak{B}, g}$ . This is a valid move since  $(A, M') \in Q$  and  $(B, N') \notin Q$ . Player III now has four options:

- If Player III picks  $a \in M'$  and  $b \in N \setminus N'$ , then  $\mathfrak{A}, f \frac{a}{x} \models \varphi$  and  $\mathfrak{B}, g \frac{b}{x} \not\models \varphi$ , so  $\varphi$  separates them. By the induction hypothesis, Player I wins at budget  $s - 1$ .
- If Player III picks  $a \in M \setminus M'$  and  $b \in N'$ , then  $\mathfrak{B}, g \frac{b}{x} \models \varphi$  and  $\mathfrak{A}, f \frac{a}{x} \not\models \varphi$ , so  $\varphi$  separates  $(\mathfrak{B}, g \frac{b}{x})$  from  $(\mathfrak{A}, f \frac{a}{x})$ . By the induction hypothesis, Player I wins at budget  $s - 1$  (note the swap of left/right).

- If Player III picks  $a \in M'$  and  $a' \in M \setminus M'$ , then  $\mathfrak{A}, f \frac{a}{x} \models \varphi$  and  $\mathfrak{A}, f \frac{a'}{x} \not\models \varphi$ , so  $\varphi$  separates them within  $\mathfrak{A}$ . By the induction hypothesis, Player I wins at budget  $s - 1$ .
- If Player III picks  $b \in N'$  and  $b' \in N \setminus N'$ , then  $\mathfrak{B}, g \frac{b}{x} \models \varphi$  and  $\mathfrak{B}, g \frac{b'}{x} \not\models \varphi$ , so  $\varphi$  separates them within  $\mathfrak{B}$ . By the induction hypothesis, Player I wins at budget  $s - 1$ .  $\square$

The converse of Lemma 4.8 does not hold in general: Player I may have a winning strategy at budget  $s$  even though the smallest separating formula is strictly larger than  $s$ . The issue is that in the quantifier move, Player I's choice of  $M'$  and  $N'$  determines the intended extension of the subformula  $\varphi$ , but the induction hypothesis only yields a *different* separating formula for each of Player III's responses. Assembling a single  $\varphi$  with  $\|\varphi\|_x^{\mathfrak{A},f} = M'$  and  $\|\varphi\|_x^{\mathfrak{B},g} = N'$  may require a formula much larger than  $s - 1$ , as demonstrated by the following example.

**Example 4.9.** Let  $\tau = \{P_1, P_2, P_3\}$  consist of three unary relations, let  $\mathcal{Q} = \{\exists_{=3}\}$ , and consider the models  $\mathfrak{A}$  with domain  $\{a_1, a_2, a_3, a_4\}$  where  $P_i^{\mathfrak{A}} = \{a_i\}$  for  $i \leq 3$ , and  $\mathfrak{B}$  with domain  $\{b\}$  where  $P_1^{\mathfrak{B}} = P_2^{\mathfrak{B}} = P_3^{\mathfrak{B}} = \emptyset$ .

Player I wins the  $\text{EF}_2\{\mathcal{Q}\}(\mathfrak{A}, \emptyset, \mathfrak{B}, \emptyset)$ -game by choosing  $\exists_{=3}$ , some variable  $x$ ,  $M' = \{a_1, a_2, a_3\}$  and  $N' = \emptyset$ . For Player III's third option (the only non-trivial one, since  $N'$  is empty), Player III picks some  $a_i \in M'$  and  $a_4 \in A \setminus M'$ . The atomic formula  $P_i(x)$  then separates them at budget 1.

However, no formula of size  $\leq 1$  defines  $M' = \{a_1, a_2, a_3\}$  in  $\mathfrak{A}$ : the formula  $P_1(x) \vee P_2(x) \vee P_3(x)$  is needed, which has size strictly greater than 1 (even if we treat disjunction as a primitive in the logical language, like conjunction). Thus the smallest formula separating  $\mathfrak{A}$  from  $\mathfrak{B}$  is strictly larger than the budget at which Player I wins.

## Acknowledgements

The authors were supported by the project *Perspectives on computational logic*, funded by the Research Council of Finland, project number 36942.

## References

- [1] Heinz-Dieter Ebbinghaus and Jörg Flum. *Finite Model Theory*. Springer Science & Business Media, 2005. 1, 2
- [2] Simi Haber and Saharon Shelah. An extension of the Ehrenfeucht-Fraïssé game for first order logics augmented with Lindström quantifiers. In *Fields of logic and computation. II*, volume 9300 of *Lecture Notes in Comput. Sci.*, pages 226–236. Springer, Cham, 2015. 1
- [3] Lauri Hella and Jouko Väänänen. The size of a formula as a measure of complexity. In *Logic without Borders: Essays on Set Theory, Model Theory*,

*Philosophical Logic and Philosophy of Mathematics*, pages 193–214. Walter De Gruyter, 2015. 1, 9

- [4] Reijo Jaakkola, Tomi Janhunen, Antti Kuusisto, Magdalena Ortiz, Matias Selin, and Mantas Šimkus. Graph Learning via Logic-Based Weisfeiler-Leman Variants and Tabularization. *arXiv preprint arXiv:2508.10651*, 2025. 2
- [5] Phokion G. Kolaitis and Jouko A. Väänänen. Generalized quantifiers and pebble games on finite structures. *Ann. Pure Appl. Logic*, 74(1):23–75, 1995. 1
- [6] Antti Kuusisto. Some Turing-Complete Extensions of First-Order Logic. In Adriano Peron and Carla Piazza, editors, *Proceedings Fifth International Symposium on Games, Automata, Logics and Formal Verification, GandALF 2014, Verona, Italy, September 10-12, 2014*, volume 161 of *EPTCS*, pages 4–17, 2014. doi:10.4204/EPTCS.161.4. 2
- [7] Antti Kuusisto. A double team semantics for generalized quantifiers. *Journal of Logic, Language and Information*, 24(2):149–191, 2015. 2
- [8] Per Lindström. First order predicate logic with generalized quantifiers. *Theoria*, 32(3), 1966. 1, 2

## A Quantifiers of Arbitrary Width and Type

In this section, we consider generalized quantifiers of arbitrary, finite width and type, which can quantify multiple formulas that bind multiple variables.

Now that we deal with tuples of variables, we might encounter the case where the same variable appears multiple times in a single tuple, e.g., in the formula  $Q(xxy)R(x, x, y)$ . We handle this by requiring that to extend the assignment with such a quantifier, the interpreted point tuple must *respect* this variable repetition. Let  $\mathbf{x} \in \text{VAR}^n$  and  $\mathbf{a} \in A^n$  for some  $n \in \mathbb{Z}_+$ . We say that  $\mathbf{a}$  **respects  $\mathbf{x}$ -repetitions** if when  $\mathbf{x}$  repeats a variable,  $\mathbf{a}$  repeats the corresponding value, i.e., if  $\mathbf{x}(i) = \mathbf{x}(j)$  for some  $1 \leq i, j \leq n$  then  $\mathbf{a}(i) = \mathbf{a}(j)$ . If each member  $\mathbf{a} \in X \subseteq A^n$  of a set of tuples respects  $\mathbf{x}$ -repetitions, then we say that the set  $X$  respects  $\mathbf{x}$ -repetitions.

Recall that an assignment over  $\mathfrak{A}$  is a function  $f : X \rightarrow A$ , where  $X \subseteq \text{VAR}$  is finite. If  $f$  is an assignment over  $\mathfrak{A}$  and  $\mathbf{a}$  respects  $\mathbf{x}$ -repetitions, then  $f \frac{\mathbf{a}}{\mathbf{x}}$  is defined such that  $f \frac{\mathbf{a}}{\mathbf{x}}(y) = f(y)$  if  $y \neq \mathbf{x}(i)$  for all  $1 \leq i \leq n$ , and  $f \frac{\mathbf{a}}{\mathbf{x}}(y) = \mathbf{a}(i)$  if  $y = \mathbf{x}(i)$  for some  $1 \leq i \leq n$ . By just  $\frac{\mathbf{a}}{\mathbf{x}}$  we denote the assignment  $\{(\mathbf{x}(i), \mathbf{a}(i)) \mid 1 \leq i \leq n\}$  (again requiring that  $\mathbf{a}$  respects  $\mathbf{x}$ -repetitions).

The **extension** of a formula  $\varphi$  in relation to  $\mathbf{x} \in \text{VAR}^n$  in the model  $\mathfrak{A}$  with assignment  $f$  is the set of tuples of points that, when interpreted as  $\mathbf{x}$ , satisfy  $\varphi$ , i.e., the set

$$\|\varphi\|_{\mathbf{x}}^{\mathfrak{A}, f} := \{\mathbf{a} \mid \mathfrak{A}, f \frac{\mathbf{a}}{\mathbf{x}} \models \varphi\}.$$

A **generalized quantifier of width  $k \in \mathbb{Z}_+$  and type  $\mathbf{n} \in \mathbb{Z}_+^k$**  is an isomorphism-closed class  $Q$  of structures  $(D, P_1, \dots, P_k)$  where  $P_i \subseteq D^{\mathbf{n}(i)}$  for all  $1 \leq i \leq k$ . Let  $\mathcal{Q}$  be a finite collection of such generalized quantifiers. The set of  $\tau$ -formulas of  $\text{FO}(\mathcal{Q})$  is generated by replacing the quantifier rule in Section 2.1 with the following rule: if  $\varphi_1, \dots, \varphi_k$  are formulas of  $\text{FO}(\mathcal{Q})$ , then so is  $Q\mathbf{x}_1, \dots, \mathbf{x}_k(\varphi_1, \dots, \varphi_k)$  for each  $Q \in \mathcal{Q}$  of width  $k$  and type  $\mathbf{n}$ , where  $\mathbf{x}_i \in \text{VAR}^{\mathbf{n}(i)}$  for all  $1 \leq i \leq k$ .

A variable  $x \in \text{VAR}$  is **free** in  $Q\mathbf{x}_1, \dots, \mathbf{x}_k(\varphi_1, \dots, \varphi_k)$  if it is free in some  $\varphi_i$  and does not appear in the corresponding  $\mathbf{x}_i$ . Let  $\mathfrak{A}$  be a  $\tau$ -model and  $f$  be an assignment over  $\mathfrak{A}$  whose domain includes all free variables of the formula being evaluated. The semantics of  $\text{FO}(\mathcal{Q})$  are defined as in Section 2.1, replacing clause 5 with:

$$\mathfrak{A}, f \models Q\mathbf{x}_1, \dots, \mathbf{x}_k(\varphi_1, \dots, \varphi_k) \iff (A, \|\varphi_1\|_{\mathbf{x}_1}^{\mathfrak{A}, f}, \dots, \|\varphi_k\|_{\mathbf{x}_k}^{\mathfrak{A}, f}) \in Q.$$

**Example A.1.** *The existential quantifier  $\exists$ , used as  $\exists x \varphi$ , has width 1 and type (1). It quantifies a single formula that binds a single variable, and checks whether  $\|\varphi\|_x^{\mathfrak{A}, f}$  is nonempty.*

*The Hartig quantifier  $I$ , used as  $I x_1, x_2(\varphi, \psi)$ , has width 2 and type (1, 1). It quantifies two formulas, each binding a single variable, and checks whether  $\|\varphi\|_{x_1}^{\mathfrak{A}, f}$  and  $\|\psi\|_{x_2}^{\mathfrak{A}, f}$  have the same cardinality.*

*The quantifier Ham, used as  $\text{Ham}(x_1 x_2) \varphi$ , has width 1 and type (2). It quantifies a single formula that binds a pair of variables, and checks whether the binary relation  $\|\varphi\|_{(x_1 x_2)}^{\mathfrak{A}, f} \subseteq A^2$  contains a Hamiltonian path through the domain.*

Let  $\mathfrak{A}$  be a  $\tau$ -model with assignment  $f$ , and let  $\mathbf{x} \in \text{VAR}^n$ . We say that a set  $X \subseteq A^n$  is **closed under**  $\equiv_{\text{FO}(\mathcal{Q})}^d$  relative to  $(\mathfrak{A}, f)$  and  $\mathbf{x}$  if  $\mathbf{a} \in X$  and  $\mathfrak{A}, f \frac{\mathbf{a}}{\mathbf{x}} \equiv_{\text{FO}(\mathcal{Q})}^d \mathfrak{A}, f \frac{\mathbf{a}'}{\mathbf{x}}$  implies  $\mathbf{a}' \in X$ , and that  $X$  is **definable** by a formula  $\varphi$  relative to  $(\mathfrak{A}, f)$  and  $\mathbf{x}$  if  $\|\varphi\|_{\mathbf{x}}^{\mathfrak{A}, f} = X$ .

For each  $n \in \mathbb{Z}_+$  and  $\mathbf{a} \in A^n$ , we define the  $d$ -**type** of  $\mathbf{a}$  relative to  $(\mathfrak{A}, f)$  and the  $n$ -tuple of variables  $\mathbf{x} \in \text{VAR}^n$  to be

$$[\mathbf{a}]_{\mathbf{x}}^d := \{\mathbf{b} \in A^n \mid \mathfrak{A}, f \frac{\mathbf{a}}{\mathbf{x}} \equiv_{\text{FO}(\mathcal{Q})}^d \mathfrak{A}, f \frac{\mathbf{b}}{\mathbf{x}}\}.$$

**Lemma A.2.** *Lemma 2.1 and Corollary 2.2 hold also when  $\mathcal{Q}$  is a finite collection of quantifiers of arbitrary, finite width and type.*

*Proof. Base case.* Let  $\Phi$  be a maximal set of non-equivalent atomic formulas of  $\text{FO}(\mathcal{Q})$  with free variables in  $\text{dom}(f) \cup \{\mathbf{x}\}$  and let  $\Phi_{\mathbf{a}} := \{\varphi \in \Phi \mid \mathfrak{A}, f \frac{\mathbf{a}}{\mathbf{x}} \models \varphi\}$ . The set  $\Phi$  is finite by similar reasoning as in Lemma 2.1. Now, the desired formula is

$$\chi_{\mathbf{a}}^0(\mathbf{x}) := \bigwedge_{\varphi \in \Phi_{\mathbf{a}}} \varphi \wedge \bigwedge_{\varphi \in \Phi \setminus \Phi_{\mathbf{a}}} \neg \varphi,$$

since  $\mathfrak{A}, f \frac{\mathbf{b}}{\mathbf{x}} \models \chi_{\mathbf{a}}^0(\mathbf{x})$  if and only if  $\mathbf{b} \in [\mathbf{a}]_{\mathbf{x}}^0$ .

**Induction case.** For  $\mathbf{b} \in [\mathbf{a}]_{\mathbf{x}}^d \setminus [\mathbf{a}]_{\mathbf{x}}^{d+1}$ , the separating formula is, without loss of generality, of the form

$$\theta_{\mathbf{b}}(\mathbf{x}) := Q_{\mathbf{b}} \mathbf{y}_1, \dots, \mathbf{y}_{k_{\mathbf{b}}} (\psi_1(\mathbf{x}, \mathbf{y}_1), \dots, \psi_{k_{\mathbf{b}}}(\mathbf{x}, \mathbf{y}_{k_{\mathbf{b}}}),$$

where  $Q_{\mathbf{b}}$  has width  $k_{\mathbf{b}}$  and type  $(n_1, \dots, n_{k_{\mathbf{b}}})$ , and each  $\psi_i(\mathbf{x}, \mathbf{y}_i)$  is a formula of  $\text{FO}(\mathcal{Q})^d$ .

We now verify that  $[\mathbf{a}]_{\mathbf{x}}^d$  splits into finitely many  $(d+1)$ -types. By the induction hypothesis, each  $\psi_i(\mathbf{x}, \mathbf{y}_i)$  is a depth- $d$  formula with free variables among  $\mathbf{x}$  and  $\mathbf{y}_i$ , and since the  $d$ -types for tuples of length  $|\mathbf{x}| + n_i$  are finitely many (by the induction hypothesis applied at the appropriate tuple length), there are finitely many non-equivalent such formulas. Combined with the finiteness of  $\mathcal{Q}$  and the finitely many choices of  $k$ -tuples of formulas, the number of non-equivalent formulas  $\theta_{\mathbf{b}}(\mathbf{x})$  is finite. Hence  $[\mathbf{a}]_{\mathbf{x}}^d$  contains finitely many  $(d+1)$ -types.

Let  $\mathbf{b}_1, \dots, \mathbf{b}_k$  be representatives of the  $(d+1)$ -types in  $[\mathbf{a}]_{\mathbf{x}}^d \setminus [\mathbf{a}]_{\mathbf{x}}^{d+1}$ . The desired formula is thus

$$\chi_{\mathbf{a}}^{d+1}(\mathbf{x}) := \chi_{\mathbf{a}}^d(\mathbf{x}) \wedge \bigwedge_{i=1}^k \theta_{\mathbf{b}_i}(\mathbf{x}).$$

Corollary 2.2 now follows as before.  $\square$

## A.1 EF-Game

**Definition A.3.** Let  $\mathfrak{A}$  and  $\mathfrak{B}$  be  $\tau$ -models, where  $\tau$  is finite, and let  $f$  and  $g$  be (possibly empty) assignments over  $\mathfrak{A}$  and  $\mathfrak{B}$  respectively, with  $\text{dom}(f) =$

$\text{dom}(g)$ . The  $\text{EF}\{Q\}(\mathfrak{A}, \mathfrak{B}, f, g)$ -game is a two-player game that starts from the position  $(\mathfrak{A}, \mathfrak{B}, f, g)$ , with Player I starting as the **attacker** and Player III starting as the **defender**. The  $i$ th round proceeds from a position  $(\mathfrak{M}, \mathfrak{N}, h, h')$ , where  $h$  and  $h'$  are assignments over  $\mathfrak{M}$  and  $\mathfrak{N}$  respectively with  $\text{dom}(h) = \text{dom}(h')$ , as follows:

1. The attacker chooses a quantifier  $Q \in \mathcal{Q}$ . Suppose it has width  $k \in \mathbb{Z}_+$  and type  $\mathbf{n} \in \mathbb{Z}_+^k$ . The attacker also chooses  $k$  tuples of variable symbols  $\mathbf{x}_1, \dots, \mathbf{x}_k$ , where  $\mathbf{x}_j \in \text{VAR}^{\mathbf{n}(j)}$  for each  $1 \leq j \leq k$ .
2. The attacker chooses  $k$  **witness sets**  $X_1, \dots, X_k$  from the domain of either model (without loss of generality, suppose they are chosen from  $M$ ), where each  $X_j \subseteq M^{\mathbf{n}(j)}$  respects  $\mathbf{x}_j$ -repetitions, such that  $(M, X_1, \dots, X_k) \in Q$ . The attacker also chooses  $k$  **spillover sets**  $P_1, \dots, P_k$ , where each  $P_j \subseteq N^{\mathbf{n}(j)}$  respects  $\mathbf{x}_j$ -repetitions.
3. The defender chooses corresponding witness sets  $X'_1, \dots, X'_k$ , where each  $X'_j \subseteq N^{\mathbf{n}(j)}$  respects  $\mathbf{x}_j$ -repetitions, such that  $(N, X'_1, \dots, X'_k) \in Q$  and  $P_j \subseteq X'_j$  for all  $1 \leq j \leq k$ .
4. The attacker chooses  $j \in \{1, \dots, k\}$  and one of the following:
  - (a) Chooses a  $\mathbf{w}' \in N^{\mathbf{n}(j)} \setminus X'_j$  that respects  $\mathbf{x}_j$ -repetitions and  $\mathbf{w} \in X_j$ . The players swap roles. The next position is  $(\mathfrak{M}, \mathfrak{N}, h \frac{\mathbf{w}}{\mathbf{x}_j}, h' \frac{\mathbf{w}'}{\mathbf{x}_j})$ .
  - (b) Chooses  $\mathbf{w}' \in X'_j$ . The defender now either:
    - chooses  $\mathbf{w} \in X_j$ , after which the next position is  $(\mathfrak{M}, \mathfrak{N}, h \frac{\mathbf{w}}{\mathbf{x}_j}, h' \frac{\mathbf{w}'}{\mathbf{x}_j})$ ,  
or
    - chooses  $\mathbf{w} \in P_j$ , after which the next position is  $(\mathfrak{N}, \mathfrak{M}, \frac{\mathbf{w}}{\mathbf{x}_j}, \frac{\mathbf{w}'}{\mathbf{x}_j})$ .
5. Let  $(\mathfrak{M}', \mathfrak{N}', h_*, h'_*)$  be the position determined in the previous step. If the pair  $(h_*, h'_*)$  does not induce a partial isomorphism between  $\mathfrak{M}'$  and  $\mathfrak{N}'$ , then the game ends and the attacker wins. Otherwise a new round begins from this position.

At any point during steps 2–3, immediately after a set is chosen, the opposing player may **contest** that choice instead of letting the round continue normally. When a contestation occurs, the remaining steps of the round are skipped and replaced as follows:

- **Contesting a witness set.** After  $X_j$  is chosen in step 2 or  $X'_j$  is chosen in step 3, the opposing player may contest that the set breaks equivalence by choosing  $\mathbf{w} \in Y$  and a  $\mathbf{w}' \in W^{\mathbf{n}(j)} \setminus Y$  that respects  $\mathbf{x}_j$ -repetitions, where  $Y$  is the contested set and  $W$  is its domain. If the contesting player is the attacker, then the players swap roles. The next position is  $(\mathfrak{W}, \mathfrak{W}, h_W \frac{\mathbf{w}}{\mathbf{x}_j}, h_W \frac{\mathbf{w}'}{\mathbf{x}_j})$ , where  $h_W = h$  if  $W = M$  and  $h_W = h'$  if  $W = N$ .

- **Contesting a spillover set.** After  $P_j$  is chosen in step 2, the opposing player may contest that it contains a type realized in  $M$  by choosing  $\mathbf{w}' \in P_j$  and a  $\mathbf{w} \in M^{\mathbf{n}(j)}$  that respects  $\mathbf{x}_j$ -repetitions. The next position is  $(\mathfrak{M}, \mathfrak{N}, h_{\frac{\mathbf{w}}{\mathbf{x}_j}}, h'_{\frac{\mathbf{w}'}{\mathbf{x}_j}})$ .

In both cases, the game then proceeds to step 5: the partial isomorphism check is performed on the new position, and if it passes, a new round begins from that position.

**Theorem A.4.** *Theorem 3.5 holds also when  $\mathcal{Q}$  is a finite set of quantifiers of arbitrary, finite width and type.*

*Proof.* The proof follows the same structure as before, with the attacker and defender now choosing  $k$  witness sets,  $k$  spillover sets, and  $k$  variable tuples  $\mathbf{x}_1, \dots, \mathbf{x}_k$  corresponding to the width  $k$  of the quantifier. The base case (Lemma 3.2) is thus identical, and below, we give a condensed version of the induction case.

For generalizing Lemma 3.3, assume contrapositively that  $(\mathfrak{A}, f)$  and  $(\mathfrak{B}, g)$  are  $(d+1)$ -equivalent, and suppose the attacker chooses a quantifier  $Q \in \mathcal{Q}$  of width  $k$  and type  $\mathbf{n}$ , variable tuples  $\mathbf{x}_1, \dots, \mathbf{x}_k$ , witness sets  $X_1, \dots, X_k$  (where each  $X_j \subseteq A^{\mathbf{n}(j)}$  respects  $\mathbf{x}_j$ -repetitions), and spillover sets  $P_1, \dots, P_k$  (where each  $P_j \subseteq B^{\mathbf{n}(j)}$  respects  $\mathbf{x}_j$ -repetitions). By the same contestation arguments as before, each  $X_j$  is closed under  $\equiv_{\text{FO}(\mathcal{Q})}^d$  relative to  $(\mathfrak{A}, f)$  and  $\mathbf{x}_j$ , and hence definable by a formula  $\theta_j(\mathbf{x}_j) \in \text{FO}(\mathcal{Q})^d$ . Similarly, the closure under  $\equiv_{\text{FO}(\mathcal{Q})}^d$  relative to  $(\mathfrak{B}, g)$  and  $\mathbf{x}_j$  of each  $P_j$  is definable by a formula  $\psi_j(\mathbf{x}_j) \in \text{FO}(\mathcal{Q})^d$  with  $\|\psi_j\|_{\mathbf{x}_j}^{\mathfrak{A}, f} = \emptyset$ . Since  $(A, X_1, \dots, X_k) \in Q$  and  $\|\theta_j \vee \psi_j\|_{\mathbf{x}_j}^{\mathfrak{A}, f} = X_j$  for each  $j$ , we have  $\mathfrak{A}, f \models Q \mathbf{x}_1, \dots, \mathbf{x}_k (\theta_1 \vee \psi_1, \dots, \theta_k \vee \psi_k)$ . By the  $(d+1)$ -equivalence assumption,  $(\mathfrak{B}, g)$  also satisfies this sentence, so the defender can choose  $X'_j := \|\theta_j \vee \psi_j\|_{\mathbf{x}_j}^{\mathfrak{B}, g}$  for each  $j$ , which satisfies  $(B, X'_1, \dots, X'_k) \in Q$  and  $P_j \subseteq X'_j$ . The attacker then chooses some component  $j$  and one of the two options. In either case, the argument from Lemma 3.3 applies to the chosen component: if the attacker picks  $\mathbf{w}' \in B^{\mathbf{n}(j)} \setminus X'_j$  and  $\mathbf{w} \in X_j$ , then  $\mathfrak{A}, f_{\frac{\mathbf{w}}{\mathbf{x}_j}} \models \theta_j$  and  $\mathfrak{B}, g_{\frac{\mathbf{w}'}{\mathbf{x}_j}} \not\models \theta_j \vee \psi_j$ , so they are  $d$ -separable; if the attacker picks  $\mathbf{w}' \in X'_j$ , then  $\mathbf{w}'$  falls into one of the sub-cases (in  $\|\psi_j\|_{\mathbf{x}_j}^{\mathfrak{B}, g}$  or in  $\|\theta_j\|_{\mathbf{x}_j}^{\mathfrak{B}, g}$ ) and the defender responds exactly as in Lemma 3.3.

For generalizing Lemma 3.4, suppose the formula  $Q \mathbf{x}_1, \dots, \mathbf{x}_k (\varphi_1, \dots, \varphi_k)$  separates  $(\mathfrak{A}, f)$  and  $(\mathfrak{B}, g)$ . The attacker chooses  $X_j := \|\varphi_j\|_{\mathbf{x}_j}^{\mathfrak{A}, f}$  and  $P_j := \{\mathbf{v}' \in \|\varphi_j\|_{\mathbf{x}_j}^{\mathfrak{B}, g} \mid \mathfrak{B}, g_{\frac{\mathbf{v}'}{\mathbf{x}_j}} \not\equiv_{\text{FO}(\mathcal{Q})}^d \mathfrak{A}, f_{\frac{\mathbf{v}}{\mathbf{x}_j}} \text{ for all } \mathbf{v} \in A^{\mathbf{n}(j)}\}$  for each  $j$ . Since we know  $\mathfrak{B}, g \not\models Q \mathbf{x}_1, \dots, \mathbf{x}_k (\varphi_1, \dots, \varphi_k)$ , the defender's witness sets satisfy  $(B, X'_1, \dots, X'_k) \in Q$  but  $X'_j \neq \|\varphi_j\|_{\mathbf{x}_j}^{\mathfrak{B}, g}$  for at least one  $j$ . The attacker chooses such a  $j$ : then either some  $\mathbf{w}' \in X'_j$  does not satisfy  $\varphi_j$  or some  $\mathbf{w}' \in B^{\mathbf{n}(j)} \setminus X'_j$  satisfies  $\varphi_j$ . In either case, the corresponding argument from Lemma 3.4 applies to the chosen component  $j$ .  $\square$

## A.2 Formula-Size Game

The **size** of a formula of the form  $Q\mathbf{x}_1, \dots, \mathbf{x}_k(\varphi_1, \dots, \varphi_k)$  is defined to be  $s(\varphi_1) + \dots + s(\varphi_k) + 1$ .

**Definition A.5.** Let  $\mathcal{A}$  and  $\mathcal{B}$  be classes of  $\tau$ -pairs. The  $\text{EF}_s\{\mathcal{Q}\}(\mathcal{A}, \mathcal{B})$ -game is a two-player game that starts in the position  $(s, \mathcal{A}, \mathcal{B})$ . The  $i$ th round proceeds from the position  $(s_i, \mathcal{A}_i, \mathcal{B}_i)$  as follows:

1. If there is an atomic  $\text{FO}(\mathcal{Q})$ -formula  $\varphi$  that separates  $\mathcal{A}_i$  and  $\mathcal{B}_i$ , then the game ends and Player I wins.
2. Otherwise, if  $s_i = 1$ , then the game ends and Player III wins.
3. If neither of the above conditions holds, then Player I chooses one of the following three options:
  - (a) **Swap classes (negation).** A new round begins from the position  $(s_i - 1, \mathcal{B}_i, \mathcal{A}_i)$ .
  - (b) **Right splitting move (conjunction).** Player I chooses  $u, v \in \mathbb{Z}_+$  such that  $u+v = s_i$ , and chooses (possibly overlapping) sets  $\mathcal{C}, \mathcal{D} \subseteq \mathcal{B}_i$  such that  $\mathcal{C} \cup \mathcal{D} = \mathcal{B}_i$ . Player III then responds by choosing whether the next round starts from the position  $(u, \mathcal{A}_i, \mathcal{C})$  or from the position  $(v, \mathcal{A}_i, \mathcal{D})$ .
  - (c) **Supplementing move (quantifier).** Player I first chooses a quantifier  $Q \in \mathcal{Q}$  of width  $k \in \mathbb{Z}_+$  and type  $\mathbf{n} \in \mathbb{Z}_+^k$ , and  $k$  tuples of variable symbols  $\mathbf{x}_1, \dots, \mathbf{x}_k$ , where  $\mathbf{x}_j \in \text{VAR}^{\mathbf{n}(j)}$  for each  $1 \leq j \leq k$ . Player I then chooses  $u_1, \dots, u_k \in \mathbb{Z}_+$  such that  $u_1 + \dots + u_k = s_i - 1$ , and  $k$  functions  $P_1, \dots, P_k$ , each of which is a function  $P_j : \mathcal{A}_i \cup \mathcal{B}_i \rightarrow \bigcup\{\mathcal{P}(C^{\mathbf{n}(j)}) \mid (\mathcal{C}, h) \in \mathcal{A}_i \cup \mathcal{B}_i\}$  such that  $P_j(\mathcal{C}, h) \subseteq C^{\mathbf{n}(j)}$  respects  $\mathbf{x}_j$ -repetitions,

$$\begin{aligned} (A, P_1(\mathfrak{A}, f), \dots, P_k(\mathfrak{A}, f)) &\in Q \quad \text{for all } (\mathfrak{A}, f) \in \mathcal{A}_i, \text{ and} \\ (B, P_1(\mathfrak{B}, g), \dots, P_k(\mathfrak{B}, g)) &\notin Q \quad \text{for all } (\mathfrak{B}, g) \in \mathcal{B}_i. \end{aligned}$$

(If such a collection of functions does not exist, Player I cannot choose this move.) This induces, for each  $1 \leq j \leq k$ , a pair of classes

$$\begin{aligned} \mathcal{C}_j^+ &:= \{(\mathcal{C}, h \frac{\mathbf{v}}{\mathbf{x}_j}) \mid (\mathcal{C}, h) \in \mathcal{A}_i \cup \mathcal{B}_i \text{ and } \mathbf{v} \in P_j(\mathcal{C}, h)\} \text{ and} \\ \mathcal{C}_j^- &:= \{(\mathcal{C}, h \frac{\mathbf{v}}{\mathbf{x}_j}) \mid (\mathcal{C}, h) \in \mathcal{A}_i \cup \mathcal{B}_i \\ &\text{and } \mathbf{v} \in C^{\mathbf{n}(j)} \setminus P_j(\mathcal{C}, h) \text{ respects } \mathbf{x}_j\text{-repetitions}\}. \end{aligned}$$

Player III then chooses  $j \in \{1, \dots, k\}$ , and the next round starts from the position  $(u_j, \mathcal{C}_j^+, \mathcal{C}_j^-)$ .

Notice that now, the supplementing move induces a split in the size budget and a choice for Player III, since multiple formulas can be quantified.

**Theorem A.6.** *Theorem 4.2 holds also when  $\mathcal{Q}$  is a finite set of quantifiers of arbitrary, finite width and type.*

*Proof.* The base case and the induction cases for negation and conjunction are proved exactly as in Theorem 4.2, so we prove the induction case for quantification. Suppose that  $s > 1$  and that the statement holds for all  $l < s$ .

We first prove the  $\implies$  direction. Suppose Player  $\text{II}$  has a winning strategy for the  $s$ -game that begins by choosing a quantifier  $Q \in \mathcal{Q}$  of width  $k$  and type  $\mathbf{n}$ ,  $k$  tuples of variable symbols  $\mathbf{x}_1, \dots, \mathbf{x}_k$ , the splits  $u_1, \dots, u_k \in \mathbb{Z}_+$  and the functions  $P_1, \dots, P_k$ . Since the strategy is winning, Player  $\text{II}$  has a winning strategy in the  $\text{EF}_{u_j}\{Q\}(\mathcal{C}_j^+, \mathcal{C}_j^-)$ -game for all  $1 \leq j \leq k$ . Since each  $u_j$  is smaller than  $s$ , by the induction hypothesis, for each pair of classes  $\mathcal{C}_j^+$  and  $\mathcal{C}_j^-$ , there exists a formula  $\varphi_j$  of size  $\leq u_j$  that separates them.

We show that  $P_j(\mathfrak{C}, h) = \|\varphi_j\|_{\mathbf{x}_j}^{\mathfrak{C}, h}$  for all  $(\mathfrak{C}, h) \in \mathcal{A} \cup \mathcal{B}$  and all  $1 \leq j \leq k$ . If  $\mathbf{v} \in P_j(\mathfrak{C}, h)$ , then  $(\mathfrak{C}, h \frac{\mathbf{v}}{\mathbf{x}_j}) \in \mathcal{C}_j^+$  by definition, so  $\mathfrak{C}, h \frac{\mathbf{v}}{\mathbf{x}_j} \models \varphi_j$  by separation. Conversely, if  $\mathbf{v} \notin P_j(\mathfrak{C}, h)$ , then  $(\mathfrak{C}, h \frac{\mathbf{v}}{\mathbf{x}_j}) \in \mathcal{C}_j^-$  by definition, so  $\mathfrak{C}, h \frac{\mathbf{v}}{\mathbf{x}_j} \not\models \varphi_j$  by separation. Together, we thus have  $\mathbf{v} \in P_j(\mathfrak{C}, h) \iff \mathfrak{C}, h \frac{\mathbf{v}}{\mathbf{x}_j} \models \varphi_j$ , establishing the desired equality.

It follows that for all  $(\mathfrak{A}, f) \in \mathcal{A}$ , we have  $P_j(\mathfrak{A}, f) = \|\varphi_j\|_{\mathbf{x}_j}^{\mathfrak{A}, f}$ , and since  $(A, P_1(\mathfrak{A}, f), \dots, P_k(\mathfrak{A}, f)) \in Q$  by the definition of  $P$ , we conclude  $\mathfrak{A}, f \models Q\mathbf{x}_1, \dots, \mathbf{x}_k(\varphi_1, \dots, \varphi_k)$ . Similarly, for all  $(\mathfrak{B}, g) \in \mathcal{B}$ , we have  $P_j(\mathfrak{B}, g) = \|\varphi_j\|_{\mathbf{x}_j}^{\mathfrak{B}, g}$ , and since  $(B, P_1(\mathfrak{B}, g), \dots, P_k(\mathfrak{B}, g)) \notin Q$  by the definition of  $P$ , we conclude  $\mathfrak{B}, g \not\models Q\mathbf{x}_1, \dots, \mathbf{x}_k(\varphi_1, \dots, \varphi_k)$ . Thus  $Q\mathbf{x}_1, \dots, \mathbf{x}_k(\varphi_1, \dots, \varphi_k)$  of size  $u_1 + \dots + u_k + 1 = s$  separates  $\mathcal{A}$  and  $\mathcal{B}$ .

We then prove the  $\impliedby$  direction. Suppose  $Q\mathbf{x}_1, \dots, \mathbf{x}_k(\varphi_1, \dots, \varphi_k)$  of size  $\leq s$  separates  $\mathcal{A}$  and  $\mathcal{B}$ . Player  $\text{II}$  chooses  $P_j(\mathfrak{C}, h) := \|\varphi_j\|_{\mathbf{x}_j}^{\mathfrak{C}, h}$  for all  $(\mathfrak{C}, h) \in \mathcal{A} \cup \mathcal{B}$ . This is a valid move, since for all  $(\mathfrak{A}, f) \in \mathcal{A}$ , we have  $\mathfrak{A}, f \models Q\mathbf{x}_1, \dots, \mathbf{x}_k(\varphi_1, \dots, \varphi_k)$ , so  $(A, P_1(\mathfrak{A}, f), \dots, P_k(\mathfrak{A}, f)) \in Q$ , and for all  $(\mathfrak{B}, g) \in \mathcal{B}$ , we have  $\mathfrak{B}, g \not\models Q\mathbf{x}_1, \dots, \mathbf{x}_k(\varphi_1, \dots, \varphi_k)$ , so  $(B, P_1(\mathfrak{B}, g), \dots, P_k(\mathfrak{B}, g)) \notin Q$ . Player  $\text{II}$  also chooses  $u_j := s(\varphi_j)$  for each  $j$ .

For any  $1 \leq j \leq k$  chosen by Player  $\text{III}$ , the game continues from the position  $(u_j, \mathcal{C}_j^+, \mathcal{C}_j^-)$ . Now, for any  $(\mathfrak{C}, h) \in \mathcal{A} \cup \mathcal{B}$  and  $\mathbf{v} \in C^{\mathbf{n}(j)}$ , if  $\mathbf{v} \in P_j(\mathfrak{C}, h)$ , then  $\mathfrak{C}, h \frac{\mathbf{v}}{\mathbf{x}_j} \models \varphi_j$  and  $(\mathfrak{C}, h \frac{\mathbf{v}}{\mathbf{x}_j}) \in \mathcal{C}_j^+$ , and if  $\mathbf{v} \notin P_j(\mathfrak{C}, h)$ , then  $\mathfrak{C}, h \frac{\mathbf{v}}{\mathbf{x}_j} \not\models \varphi_j$  and  $(\mathfrak{C}, h \frac{\mathbf{v}}{\mathbf{x}_j}) \in \mathcal{C}_j^-$ . Thus  $\mathcal{C}_j^+$  and  $\mathcal{C}_j^-$  are separated by  $\varphi_j$ , and since  $\varphi_j$  is of size  $\leq u_j < s$ , by the induction hypothesis, Player  $\text{II}$  has a winning strategy from this position.  $\square$

Finally, it is clear that Corollary 4.5, Theorem 4.6 and Lemma 4.8 hold (with appropriate modifications) also when the game is played with quantifiers of arbitrary, finite width and type.