Theorem 1.6.27 (Ehrenfeucht-Fraïssé-Theorem). *Let Voc be a finite relational vocabulary and $\mathcal{A}$ and $\mathcal{B}$ two structures for Voc. Then, the following two statements are equivalent:*

(a) $\mathcal{A} \cong_k \mathcal{B}$

(b) *$\mathcal{A}$ and $\mathcal{B}$ agree on all* FOL[k]*-sentences, i.e.,* $\mathcal{A} \models \psi$ *iff* $\mathcal{B} \models \psi$ *for all* FOL[k]*-sentence* ψ

Obviously, statement (b) is equivalent to $\mathrm{FOL}[k](\mathcal{A}) = \mathrm{FOL}[k](\mathcal{B})$.

Proof. We use the recursive characterization of $\cong_k$ stated in Lemma 1.6.19 on page 96 and prove the claim by induction on k. The basis of induction ($k = 0$) is obvious as

- $\mathcal{A} \cong_0 \mathcal{B}$ iff $\mathcal{A}$ and $\mathcal{B}$ agree on all atomic sentences (by part (a) of Lemma 1.6.19)
- FOL[0]-sentences are boolean combinations of atomic sentences.

For the step of induction we assume that $k \geqslant 1$ and suppose that the equivalence of statements (a) and (b) holds for $k-1$ (induction hypothesis). Let $A = \mathit{Dom}^{\mathcal{A}}$, $B = \mathit{Dom}^{\mathcal{B}}$.

(a) $\Longrightarrow$ (b): Let $\mathcal{A} \cong_k \mathcal{B}$ and let ϕ be a FOL[k]-sentence. Then, ϕ can be written as a boolean combination of formulas

$$\exists x.\psi(x) \quad \text{where } \psi(x) \in \mathrm{FOL}[k-1](x).$$

Recall that $\mathrm{FOL}[k-1](x)$ denotes the set of FOL-formulas ψ over the given vocabulary with $\mathit{qr}(\phi) < k$ and $\mathit{Free}(\psi) \subseteq \{x\}$. Thus, it suffices to prove that $\mathcal{A}$ and $\mathcal{B}$ agree on such formulas $\exists x.\psi(x)$. Let us assume that $\mathcal{A} \models \exists x.\psi(x)$. Then, there is an element $a \in A$ with

$$(\mathcal{A}, [x := a]) \models \psi(x).$$

We now extend the given vocabulary by a new constant symbol c and define $\psi(c) \overset{\text{def}}{=} \psi[x/c]$ to be the formula that results from ψ by replacing any free occurrence of x in ψ with c. Then, $\psi(c)$ is a FOL[k−1]-sentence over the extended vocabulary. Let $(\mathcal{A}, a)$ denote the structure for the extended vocabulary that extends $\mathcal{A}$ by interpreting c with a. Clearly, we then have

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

Since we assume $\mathcal{A} \cong_k \mathcal{B}$, we may apply part (b) of Lemma 1.6.19 and obtain the existence of an element $b \in B$ such that

$$(\mathcal{A}, a) \cong_{k-1} (\mathcal{B}, b).$$

Since $(\mathcal{A}, a) \models \psi(c)$ and $\psi(c)$ is a FOL[k−1]-sentence the induction hypothesis yields

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

But then $(\mathcal{B}, [x := b]) \models \psi(x)$, which yields $\mathcal{B} \models \exists x.\psi(x)$.

(b) $\Longrightarrow$ (a): We assume that $\mathcal{A}$ and $\mathcal{B}$ yield the same truth value for all FOL[k]-sentences. To show $\mathcal{A} \cong_k \mathcal{B}$ we verify condition (1) in part (b) of Lemma 1.6.19. (Condition (2) follows by symmetry.) That is, we pick some element $a \in A$ and seek for an element $b \in B$ with $(\mathcal{A}, a) \cong_{k-1} (\mathcal{B}, b)$. Again, we consider the vocabulary that extends *Voc* by a new constant symbol c and regard $(\mathcal{A}, a)$ as a structure for the extended vocabulary. We now consider the rank-$(k-1)$-type $\mathrm{FOL}[k-1](\mathcal{A}, a)$ of this structure, and take a $\mathrm{FOL}[k-1]$-formula θ over the extended alphabet that defines $\mathrm{FOL}[k-1](\mathcal{A}, a)$. I.e.,

$$\mathrm{FOL}[k-1](\mathcal{A}, a) = \left\{ \psi : \begin{array}{l} \psi \text{ is a FOL}[k-1]\text{-sentence over the} \\ \text{extended vocabulary s.t. } \theta \Vdash \psi \end{array} \right\}$$

Such a sentence θ exists by Lemma 1.6.26 and we have $(\mathcal{A}, a) \models \theta$. Replacing any occurrence of the new constant symbol c with some fresh variable x in the $\mathrm{FOL}[k-1]$-sentence θ yields a $\mathrm{FOL}[k-1](x)$-formula $\theta[c/x] = \psi(x)$ over the original vocabulary such that $\theta = \psi(c)$. As $(\mathcal{A}, a) \models \theta$ we have:

$$(\mathcal{A}, [x := a]) \models \psi(x),$$

and therefore $\mathcal{A} \models \exists x.\psi(x)$. Since $\mathcal{A}$ and $\mathcal{B}$ agree on all FOL[k]-sentences and $\exists x.\psi(x)$ is a FOL[k]-sentence over the original vocabulary, we get

$$\mathcal{B} \models \exists x.\psi(x).$$

Therefore, there is some $b \in B$ where $(\mathcal{B}, [x := b]) \models \psi(x)$. But then

$$(\mathcal{B}, b) \models \psi(c) = \theta.$$

Since θ defines the rank-$(k-1)$ type of structure $(\mathcal{A}, a)$ for the extended vocabulary, we get:

$$\mathrm{FOL}[k-1](\mathcal{B}, b) = \mathrm{FOL}[k-1](\mathcal{A}, a).$$

Therefore, $(\mathcal{A}, a)$ and $(\mathcal{B}, b)$ agree on all $\mathrm{FOL}[k-1]$-sentences. The induction hypothesis yields $(\mathcal{A}, a) \cong_{k-1} (\mathcal{B}, b)$. □

As the Ehrenfeucht-Fraïssé Theorem states a one-to-one correspondence between the $\cong_k$-equivalence classes and the rank-k-types, we obtain:

Corollary 1.6.28 (k-round game equivalence has a finite index). *For each finite relational vocabulary and natural number* k*, the equivalence relation* $\cong_k$ *is of finite index, i.e., the number of* $\cong_k$*-equivalence classes is finite.*

Example 1.6.29 (Decidability of FOL over the empty vocabulary). Using the Ehrenfeucht-Fraïssé Theorem we can establish the decidability of the satisfiability problem for FOL-formulas with equality over the empty vocabulary. Obviously, given a FOL-formula $\phi(\overline{x})$ with equality over the empty vocabulary that is subject for a satisfiability checker, we first may switch from $\phi(\overline{x})$ to the sentence $\exists \overline{x}.\phi(\overline{x})$ which is satisfiable if and only if so is $\phi(\overline{x})$. Thus, it is no restriction to suppose that the input formula ϕ is closed. Let

$$k \stackrel{\text{def}}{=} qr(\phi).$$

We saw in Example 1.6.16 on page 94 that all structures (i.e., sets) with k or more elements are $\cong_k$-equivalent. By the Ehrenfeucht-Fraïssé Theorem, either ϕ holds for all sets with k or more elements or for none of them. Thus:

$$\phi \text{ is satisfiable} \quad \text{iff} \quad \phi \text{ has a model with at most } k \text{ elements}$$

But then, an algorithm to check satisfiability is obtained by checking whether one of the sets $\{1\}, \{1,2\}, \ldots, \{1,2,\ldots,k\}$ is a model for ϕ. ■

A simple consequence of the Ehrenfeucht-Fraïssé Theorem is that any two structures that are equivalent with respect to $\cong_k$ for all natural numbers k cannot be distinguished by FOL-sentences.

Definition 1.6.30 (Game equivalence $\cong$). Let $\mathcal{A}$ and $\mathcal{B}$ structures over the same finite relational vocabulary *Voc*. Then, we define:

$$\mathcal{A} \cong \mathcal{B} \quad \text{iff} \quad \mathcal{A} \cong_k \mathcal{B} \quad \text{for all } k \in \mathbb{N}$$ ■

The Ehrenfeucht-Fraïssé Theorem yields that game equivalence $\cong$ agrees with FOL-equivalence:

Corollary 1.6.31 (FOL-equivalence of structures). *Let $\mathcal{A}$ and $\mathcal{B}$ structures over the same finite relational vocabulary Voc. Then, the following statements are equivalent:*

(a) $\mathcal{A} \cong \mathcal{B}$

(b) $\mathcal{A}$ *and* $\mathcal{B}$ *fulfill the same FOL-sentences over Voc.*

(c) $Th(\mathcal{A}) = Th(\mathcal{B})$

Proof. Recall that $Th(\mathcal{A})$ denotes the set of all formulas ψ over *Voc* that are true in $\mathcal{A}$, i.e., $Th(\mathcal{A}) = \{\psi : \mathcal{A} \models \psi\}$.

The equivalence of statements (b) and (c) is obvious. The implication (c) $\Longrightarrow$ (b) is trivial by the definition of $Th(\cdot)$. The implication (b) $\Longrightarrow$ (c) is obvious too, as formula $\psi(\overline{x})$ is true in $\mathcal{A}$ if and only if the sentence $\forall \overline{x}.\psi(\overline{x})$ is true in $\mathcal{A}$. See Lemma 1.4.3 on page 46.

"(a) $\Longrightarrow$ (b)": Let ϕ be a FOL-sentence and k its quantifier rank. Then, ϕ is a FOL[k]-sentence. As $\mathcal{A} \cong \mathcal{B}$ we have $\mathcal{A} \cong_k \mathcal{B}$. But then the Ehrenfeucht-Fraïssé Theorem yields that ϕ either holds for both structures $\mathcal{A}$ and $\mathcal{B}$ or for none of them.

"(b) $\Longrightarrow$ (a)": If $\mathcal{A}$ and $\mathcal{B}$ fulfill the same FOL-sentences then, for each $k \in \mathbb{N}$, they fulfill the same FOL[k]-sentences. The Ehrenfeucht-Fraïssé Theorem yields that $\mathcal{A} \cong_k \mathcal{B}$ for each $k \in \mathbb{N}$. But then $\mathcal{A} \cong \mathcal{B}$. □

Corollary 1.6.32 (FOL-equivalence of infinite sets). *If A and B are infinite sets then A and B fulfill the same FOL-sentences over the empty vocabulary.*

Proof. In Example 1.6.16 on page 94, we saw that any two sets with k or more elements are $\cong_k$-equivalent. Thus, given two infinite sets A and B, we have $A \cong_k B$ for all $k \in \mathbb{N}$, and therefore $A \cong B$. Corollary 1.6.31 yields the claim. □

Another example of non-isomorphic structures that satisfy the same FOL-sentences are obtained by considering dense linear orders without endpoints (see Definition 1.4.14 on page 53).

Lemma 1.6.33 (FOL-equivalence of dense linear orders without endpoints). *Every two dense linear orders $\mathcal{A} = (A, \sqsubseteq_A)$ and $\mathcal{B} = (B, \sqsubseteq_B)$ without endpoints satisfy the same FOL-sentences. (The underlying vocabulary consists of a single binary predicate symbol).*

Proof. By Corollary 1.6.31, it suffices to show that $\mathcal{A} \cong \mathcal{B}$. For this, we show that for each pair $\langle \overline{a}, \overline{b} \rangle$ where $\overline{a} = (a_1, \ldots, a_k) \in A^k$ and $\overline{b} = (b_1, \ldots, b_k) \in B^k$ that defines a partial isomorphism and each element $a_{k+1} \in A$ there exists an element $b_{k+1} \in B$ such that $\langle (\overline{a}, a_{k+1}), (\overline{b}, b_{k+1}) \rangle$ defines a partial isomorphism.

In fact, since $\mathcal{B}$ is dense without endpoints we can choose $b_{k+1} \in B$ that relates to $b_1, \ldots, b_k$ in the same way as a_{k+1} relates to $a_1, \ldots, a_k$.

- If $a_{k+1} \in \{a_1, \ldots, a_k\}$, say $a_{k+1} = a_i$, then we define $b_{k+1} = b_i$.
- If $a_{k+1} \sqsubset_A a_i$ for $1 \leqslant i \leqslant k$ then we pick $b_{k+1} \in B$ such that $b_{k+1} \sqsubset_B b_i$ for $1 \leqslant i \leqslant k$. This is possible as $\mathcal{B}$ has no minimal element.
- Similarly, if $a_i \sqsubset_A a_{k+1}$ for $1 \leqslant i \leqslant k$ then we pick $b_{k+1} \in B$ such that $b_i \sqsubset_B b_{k+1}$ for $1 \leqslant i \leqslant k$. This is possible as $\mathcal{B}$ has no maximal element.
- If $a_i \sqsubset_A a_{k+1} \sqsubset_A a_j$ and there is no $h \in \{1, \ldots, k\}$ such that $a_i \sqsubset_A a_h \sqsubset_A a_j$ then the density of $\mathcal{B}$ permits to choose an element $b_{k+1} \in B$ such that $b_i \sqsubset_B b_{k+1} \sqsubset_B b_j$.

By symmetry, this yields that there is a strategy for the duplicator that is winning for each number of rounds. In particular, we get $\mathcal{A} \cong_k \mathcal{B}$ for all $k \in \mathbb{N}$. □

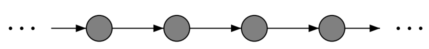
As a consequence of Lemma 1.6.33, we obtain that the FO-theory of the ordered rationals (cf. Section 1.4.3, page 53) agrees with the FO-theory of the ordered reals:

$$\mathit{Th}(\mathbb{R}, \leqslant) = \mathit{Th}(\mathbb{Q}, \leqslant)$$

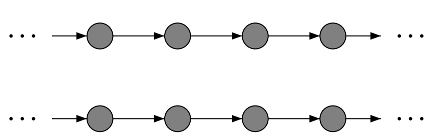
Thus, the quantifier elimination algorithm that we have presented for the ordered rationals can also be applied to check the truth of sentences over the ordered reals.

Example 1.6.34 (Two infinite graphs). In the former examples for structures that are $\cong$-equivalent (infinite sets and dense linear orders) the duplicator has a strategy that is winning for each number of rounds. However, there are also structures where k-round winning strategies exist for each $k \in \mathbb{N}$, but these strategies depend on k. For an example, we regard two infinite graphs. The underlying vocabulary is $\mathit{Voc}_{\mathrm{graph}}$.

- Graph $\mathcal{A}$ has the node-set $A = \mathbb{Z}$ and an edge from n to $n+1$ for each $n \in \mathbb{Z}$.
- Graph $\mathcal{B}$ consists of two disjoint copies of $\mathcal{A}$. Say the node-set of $\mathcal{B}$ is $B = \mathbb{Z} \times \{1,2\}$ and the edge relation is $E^{\mathcal{B}} = \{ \langle (n,1), (n+1,1) \rangle, \langle (n,2), (n+1,2) \rangle : n \in \mathbb{Z} \}$.



structure $\mathcal{A}$



structure $\mathcal{B}$

Let us first observe that there is no strategy for the duplicator that is winning for each number of rounds. To see this, we consider the following strategy for the spoiler. In the first two rounds the spoiler selects structure $\mathcal{B}$ and the nodes $b_1 = \langle 0,1 \rangle$ and $b_2 = \langle 0,2 \rangle$. Let $a_1 = n$ and $a_2 = m$ be the answers given by the duplicator in rounds 1 and 2, respectively. We may assume that $n \neq m$ as otherwise the 2-round-outcome $\langle (a_1,b_1),(a_2,b_2) \rangle$ is not winning and the spoiler would win already after the second round. W.l.o.g. $n < m$. Then, in rounds $3,4,\ldots,m-n+2$, the spoiler chooses structure $\mathcal{A}$ and the elements $a_3 = n+1$, $a_4 = n+2$, $a_5 = n+3$, ..., $a_{m-n+2} = m$. The outcome $\langle (a_1,a_2,\ldots,a_{m-n+2}),\ (b_1,b_2,\ldots,b_{m-n+2}) \rangle$ after round $m-n+2$ does not define a partial isomorphism, as $a_1\, a_3 \ldots a_{m-n+2}$ constitutes a path from node $a_1 = n$ to node $a_2 = a_{m-n+2} = m$ in $\mathcal{A}$, while node $b_2 = b_{m-n+2}$ is not reachable from node b_1 in $\mathcal{B}$.

We now show that $\mathcal{A} \cong \mathcal{B}$. For this we pick some $k \in \mathbb{N}$ and show that there is a winning strategy for $\mathcal{A} \cong_k \mathcal{B}$. The cases $k = 0$ or $k = 1$ are trivial since Voc_{graph} does not have constant symbols and no node in $\mathcal{A}$ or $\mathcal{B}$ has a self-loop. Let now $k \geqslant 2$. We define a distance function d for the nodes in $\mathcal{A}$ and $\mathcal{B}$, respectively, as follows. Given two nodes v and w in $\mathcal{A}$ or $\mathcal{B}$ we define $d(v,w)$ as the length of the unique path from v to w, provided that w is reachable from v. Otherwise $d(v,w) = \infty$. Thus, if $v = n$ and $w = m$ are nodes in $\mathcal{A}$ then

$$d(n,m) \;=\; \begin{cases} m-n & : \text{ if } n \leqslant m \\ \infty & : \text{ if } n > m \end{cases}$$

For nodes v, w in $\mathcal{B}$, the values of the distance function are as follows. If $v = (n,\mathfrak{c})$, $w = (m,\mathfrak{c})$ where $n,m \in \mathbb{Z}$, $n \leqslant m$ and $\mathfrak{c} \in \{1,2\}$ then $d(n,m) = m-n$. In all other cases, $d(v,w) = \infty$. Note that d is not symmetric.

The idea of the k-round winning strategy $\mathcal{S}_k$ is to ensure that the outcome after the i-th round yields an isomorphism of the 2^{k-i}-neighbourhood of each selected element. Formally, we deal with the functions

$$d_K(v,w) \;\stackrel{\text{def}}{=}\; \begin{cases} d(v,w) & : \text{ if } d(v,w) \leqslant 2^K \\ \infty & : \text{ otherwise} \end{cases}$$

for $K = 0,1,\ldots,k$ and aim to design $\mathcal{S}_k$ in such a way that all $\mathcal{S}_k$-outcomes $\langle (a_1,\ldots,a_i)\,(b_1,\ldots,b_i) \rangle$ after the i-th round (for $0 \leqslant i \leqslant k$) enjoy the property:

$$d_{k-i}(a_h,a_j) \;=\; d_{k-i}(b_h,b_j) \quad \text{for all } h,j \in \{1,\ldots,i\} \tag{*}$$

Note that with $k = i$, condition (*) yields:

$$d_0(a_h,a_j) \;=\; d_0(b_h,b_j) \quad \text{for all } 1 \leqslant h,j \leqslant k.$$

Thus, condition (*) for $k = i$ yields that $\langle (a_1,\ldots,a_k),(b_1,\ldots,b_k) \rangle$ is winning as we have:

$$d_0(v,w) \;=\; \begin{cases} 0 & \text{iff } \; v = w \\ 1 & \text{iff } \; (v,w) \text{ is an edge} \\ \infty & \text{iff } \; v \neq w \text{ and } (v,w) \text{ is not an edge} \end{cases}$$

We now turn to the definition of such a k-round winning strategy $\mathcal{S}_k$. Suppose $0 \leqslant i < k$ and that $\langle (a_1,\ldots,a_i),(b_1,\ldots,b_i) \rangle$ is a game configuration after round i such that (*) holds. For the remaining argument, structures $\mathcal{A}$ and $\mathcal{B}$ are treated in a symmetric way. So, we may suppose that the spoiler picks graph $\mathcal{A}$ and node $a = a_{i+1} \in A$. We now explain how to define an appropriate element $b = \mathcal{S}_k((a_1, \ldots, a_i, a), (b_1, \ldots, b_i))$ in structure $\mathcal{B}$.

Case 1. $d(a_j, a) \leqslant 2^{k-i-1}$ for some $j \in \{1, \ldots, i\}$. Let

$$H \stackrel{\text{def}}{=} \{ h \in \{1, \ldots, i\} : d(a_h, a) \leqslant 2^{k-i-1} \text{ or } d(a, a_h) \leqslant 2^{k-i-1} \}.$$

Then for each index $h \in H$, node a_h is in the 2^{k-i}-neighbourhood of a_j, i.e., we have:

$$d(a_j, a_h) \leqslant 2^{k-i} \quad \text{or} \quad d(a_h, a_j) \leqslant 2^{k-i} \quad \text{for all } h \in H$$

Note that for each index $h \in H$:

- if $d(a, a_h) \leqslant 2^{k-i-1}$ then
$$d(a_j, a_h) \leqslant d(a_j, a) + d(a, a_h) \leqslant 2^{k-i-1} + 2^{k-i-1} = 2^{k-i}$$
- if $d(a_h, a) \leqslant 2^{k-i-1}$ and $h \leqslant j$ then we get $d(a_h, a_j) \leqslant d(a_h, a) \leqslant 2^{k-i-1}$, while for $j < h$ we get $d(a_j, a_h) \leqslant d(a_j, a) \leqslant 2^{k-i-1}$.

That is, all elements a_h for $h \in H$ are in the 2^{k-i}-neighbourhood of a_j. By assumption (*):

$$d(a_j, a_h) = d(b_j, b_h) \quad \text{for all indices } h \in H$$

On the basis of this observation, the duplicator can choose an element $b = b_{i+1}$ in the 2^{k-i}-neighbourhood of b_j such that

(i) $d(a_h, a) = d(b_h, b)$ and $d(a, a_h) = d(b, b_h)$
for all indices $h \in \{1, \ldots, i\}$ where $d(a_h, a) \leqslant 2^{k-i}$ or $d(a, a_h) \leqslant 2^{k-i}$.

This choice of b yields that:

(ii) $d(b_h, b) > 2^{k-i-1}$ and $d(b, b_h) > 2^{k-i-1}$
for all indices $h \in \{1, \ldots, i\}$ where $d(a_h, a) > 2^{k-i}$ and $d(a, a_h) > 2^{k-i}$

Let us check why this holds. Suppose by contradiction that $d(b, b_h) \leqslant 2^{k-i-1}$ for some index $h \in \{1, \ldots, i\}$ where $d(a_h, a) > 2^{k-i}$ and $d(a, a_h) > 2^{k-i}$. Then:

$$d(b_j, b_h) \leqslant d(b_j, b) + d(b, b_h) \leqslant 2^{k-i-1} + 2^{k-i-1} = 2^{k-i}$$

Condition (*) implies that $d(a_j, a_h) = d(b_j, b_h) \leqslant 2^{k-i}$. But then $d(a, a_h) = d(b, b_h) \leqslant 2^{k-i}$ by the choice of b (see (i)). Contradiction. With an analogue argument, the case $d(b_h, b) \leqslant 2^{k-i-1}$ can be ruled out whenever a_h is not in the 2^{k-i}-neighbourhood of a.

Statements (i) and (ii) yield that with $b = b_{i+1}$ we get:

$$d_{k-i-1}(a_h, a_\ell) = d_{k-i-1}(b_h, b_\ell) \quad \text{for all } h, \ell \in \{1, \ldots, i+1\}$$

Case 2. $d(a, a_j) \leqslant 2^{k-i-1}$ for some $j \in \{1, \ldots, i\}$. This case can be treated in an analogous way to the first case.

Case 3. There is no index $j \in \{1, \ldots, i\}$ such that $d(a_j, a) \leqslant 2^{k-i-1}$ or $d(a, a_j) \leqslant 2^{k-i-1}$.

Then the duplicator chooses an arbitrary element $b = b_{i+1}$ such that $d(b_j, b)$, $d(b, b_j) > 2^{k-i-1}$ for all $j \in \{1, \ldots, i\}$. This is possible because $\mathcal{B}$ is infinite and there are only finitely many elements which are in the 2^{k-i-1}-neighbourhood of $b_1, \ldots, b_i$.

In all three cases, the outcome $\langle (a_1, \ldots, a_i, a_{i+1}), (b_1, \ldots, b_i, b_{i+1}) \rangle$ of the so designed k-round strategy after the $(i+1)$-st round enjoys the property that $d_{k-i-1}(a_h, a_j) = d_{k-i-1}(b_h, b_j)$ for all $h, j \in \{1, \ldots, i+1\}$. Hence, this strategy for the duplicator is a k-round winning strategy. This completes the proof for $\mathcal{A} \cong_k \mathcal{B}$. ■

The above considerations illustrate the elegance of the Ehrenfeucht-Fraïssé theory to show that two structures agree on all FOL-sentences up to some quantifier rank or even for all FOL-sentences. However, the Ehrenfeucht-Fraïssé Theorem also yields a powerful tool to prove that a certain property on structures is not FO-definable.

Example 1.6.35 (Non-FO-definability of reachability). Consider again the two graphs $\mathcal{A}$ and $\mathcal{B}$ of Example 1.6.34 on page 104. The fact that $\mathcal{A} \cong \mathcal{B}$ can be used for an alternative proof that reachability in directed graphs is not FO-definable. Suppose by contradiction that $\phi(x, y)$ is a FOL-formula over *Voc*$_{\text{graph}}$ such that for each graph $\mathcal{G}$ and nodes a, b in $\mathcal{G}$:

$$(\mathcal{G}, a, b) \models \phi(x, y) \quad \text{iff} \quad b \text{ is reachable from } a \text{ in } \mathcal{G}$$

Let $\psi \stackrel{\text{def}}{=} \exists x \forall y. (\phi(x, y) \vee \phi(y, x))$. Then, $\mathcal{A} \models \psi$, while $\mathcal{B} \not\models \psi$. This is impossible as $\mathcal{A}$ and $\mathcal{B}$ fulfill the same FOL-sentences over *Voc*$_{\text{graph}}$. ■

Theorem 1.6.36 (FO-definability and Ehrenfeucht-Fraïssé games). *Let Voc be a finite relational vocabulary and $\mathcal{C}$ a class of structures for Voc. Then, the following statements are equivalent:*

(a) $\mathcal{C}$ is FO-definable.

(b) There exists $k \in \mathbb{N}$ such that for all structures $\mathcal{A}$ and $\mathcal{B}$ over Voc:

$$\textit{if } \mathcal{A} \in \mathcal{C} \textit{ and } \mathcal{A} \cong_k \mathcal{B} \textit{ then } \mathcal{B} \in \mathcal{C}.$$

Proof. (a) $\Longrightarrow$ (b): Assume $\mathcal{C}$ is defined by the FOL-sentence ϕ over *Voc*. Let $k = qr(\phi)$ be the quantifier rank of ϕ. Then, ϕ is a FOL[k]-sentence. Let $\mathcal{A}$ and $\mathcal{B}$ be structures such that $\mathcal{A} \cong_k \mathcal{B}$. The Ehrenfeucht-Fraïssé Theorem yields that $\mathcal{A}$ and $\mathcal{B}$ have the same truth value for ϕ. But then either $\mathcal{A}, \mathcal{B} \models \phi$ (in which case both $\mathcal{A}, \mathcal{B}$ belong to $\mathcal{C}$) or $\mathcal{A}, \mathcal{B} \not\models \phi$ (in which case neither $\mathcal{A}$ nor $\mathcal{B}$ belongs to $\mathcal{C}$).

(b) $\Longrightarrow$ (a): Let $k \in \mathbb{N}$ such that $\mathcal{C}$ does not distinguish between $\cong_k$-equivalent structures, i.e., for all structures $\mathcal{A}$, $\mathcal{B}$ for *Voc*:

$$\text{if } \mathcal{A} \cong_k \mathcal{B} \text{ then:} \quad \mathcal{A} \in \mathcal{C} \text{ iff } \mathcal{B} \in \mathcal{C}$$

By Lemma 1.6.26 (page 99), there are only finitely many rank-k-types and each of them is defined by a FOL[k]-sentence. Let $\mathfrak{R}_1, \mathfrak{R}_2, \ldots, \mathfrak{R}_s$ be an enumeration of the rank-k-types for *Voc* and let θ_i be a FOL[k]-sentence that defines $\mathfrak{R}_i$, i.e.,

$$\mathfrak{R}_i \ = \ \big\{ \phi : \phi \text{ is a FOL[k]-sentence with } \theta_i \models \phi \ \big\}$$

and $\mathcal{A} \models \theta_i$ iff $\text{FOL}[k](\mathcal{A}) = \mathfrak{R}_i$. We define I as the set of indices $i \in \{1, \ldots, s\}$ such that $\mathfrak{R}_i$ is the rank-k-type of some $\mathcal{C}$-structure, i.e.,

$$I \ \stackrel{\text{def}}{=} \ \big\{ i \in \{1, \ldots, s\} : \text{there exists } \mathcal{A} \in \mathcal{C} \text{ s.t. } \mathfrak{R}_i = \text{FOL}[k](\mathcal{A}) \ \big\}$$

and let ϕ be the disjunction of the FOL[k]-sentences that define the rank-k-types $\mathfrak{R}_i$ for $i \in I$:

$$\phi \ \stackrel{\text{def}}{=} \ \bigvee_{i \in I} \theta_i.$$

Then, ϕ is a FOL[k]-sentence that defines $\mathcal{C}$. To see this we have to show that $\mathcal{B} \models \phi$ iff $\mathcal{B} \in \mathcal{C}$.

- If $\mathcal{B} \models \phi$ then $\mathcal{B} \models \theta_i$ for some $i \in I$. But then

  $$\text{FOL}[k](\mathcal{B}) \ = \ \mathfrak{R}_i \ = \ \text{FOL}[k](\mathcal{A})$$

  for some $\mathcal{A} \in \mathcal{C}$. Then, $\mathcal{A} \cong_k \mathcal{B}$. Assumption (b) yields $\mathcal{B} \in \mathcal{C}$.

- If $\mathcal{B} \in \mathcal{C}$ then $\text{FOL}[k](\mathcal{B}) = \mathfrak{R}_i$ for some $i \in I$ and hence, $\mathcal{B} \models \theta_i$. This yields $\mathcal{B} \models \phi$.

Thus, $\mathcal{C}$ is FO-definable, as stated in (a). □

By negating both conditions (a) and (b) in Theorem 1.6.37 we get the following criterion to establish non-FO-definability results:

Corollary 1.6.37 (Non-FO-definability and Ehrenfeucht-Fraïssé games). *Let* Voc *be a finite relational vocabulary and $\mathcal{C}$ a class of structures for* Voc*. Then: $\mathcal{C}$ is* not *FO-definable if and only if there exist sequences $(\mathcal{A}_k)_{k \geqslant 0}$ and $(\mathcal{B}_k)_{k \geqslant 0}$ of structures over* Voc *such that $\mathcal{A}_k \cong_k \mathcal{B}_k$, $\mathcal{A}_k \in \mathcal{C}$ and $\mathcal{B}_k \notin \mathcal{C}$ for all $k \geqslant 0$.*