Ehrenfeucht-Fraïssé Games are a key tool for figuring out what is expressible in a given first order language.

### 7.1 Isomorphism and Elementary Equivalence

**Definition 7.1** Let $\Sigma = (R_1^{a_1}, \ldots, R_s^{a_s}; f_1^{r_1}, \ldots, f_t^{r_t})$ and $\mathcal{A}, \mathcal{B} \in \text{STRUC}[\Sigma]$.

Then $\mathcal{A}$ and $\mathcal{B}$ are **isomorphic**, denoted $\mathcal{A} \cong \mathcal{B}$ iff there is a function $\eta : |\mathcal{A}| \to |\mathcal{B}|$ such that

1. for all $R^a \in \Sigma$, $e_1, \ldots, e_a \in |\mathcal{A}|$ \((e_1, \ldots, e_a) \in R^A \iff (\eta(e_1), \ldots, \eta(e_a)) \in R^B\)
2. for all $f^r \in \Sigma$, $e_1, \ldots, e_r \in |\mathcal{A}|$ \((\eta(f^A(e_1, \ldots, e_r)) = f^B(\eta(e_1), \ldots, \eta(e_r))\)

In words, $\mathcal{A} \cong \mathcal{B}$ iff there exists a 1:1 correspondence from the universe of $\mathcal{A}$ to the universe of $\mathcal{B}$ which preserves all of the relevant relations and functions. In this case, the function, $\eta$, is called an **isomorphism**.

Two structures are **isomorphic** iff they are **identical except perhaps for the names of the elements** of their universes.

**Example 7.2** Consider the following structures, $\mathcal{G}, \mathcal{H}, \mathcal{H}' \in \text{STRUC}[\Sigma_{\text{rgb graph}}]$, where $\Sigma_{\text{rgb graph}} = (E^2, R^1, G^1, B^1)$ is the vocabulary of colored graphs with the three colors, $R, G, B$. For example, $R^G = \{g_1, g_4\}$, $G^G = \{g_2, g_3\}$, $B^G = \{g_3, g_6\}$.

Observe that $\mathcal{H} \cong \mathcal{H}'$ where the isomorphism $\eta = \{(h_i, h'_i) \mid 1 \leq i \leq 6\}$. Note that $\mathcal{G} \not\cong \mathcal{H}$. For them to be isomorphic, there would have to be an isomorphism $\sigma : \{g_1, \ldots, g_6\} \simto \{h_1, \ldots, h_6\}$ such that for all $g_1, g_2 \in |\mathcal{G}|$, $(g_1, g_2) \in E^G \Rightarrow (\sigma(g_1), \sigma(g_2)) \in E^H$ and there is no such $\sigma$. To be an isomorphism, $\sigma$ would also have to satisfy, $\forall g \in |\mathcal{G}| \left( g \in R^G \iff \sigma(g) \in R^H \wedge g \in G^G \iff \sigma(g) \in G^H \right) \wedge g \in B^G \iff \sigma(g) \in B^H$.

**Definition 7.3** [Elementary equivalence] Let $\mathcal{A}, \mathcal{B} \in \text{STRUC}[\Sigma]$. $\mathcal{A}$ and $\mathcal{B}$ are **elementary equivalent**, denoted $\mathcal{A} \equiv \mathcal{B}$, iff for all sentences $\varphi \in \mathcal{L}(\Sigma)$, $\mathcal{A} \models \varphi \iff \mathcal{B} \models \varphi$. 

1
Recall that a sentence is a formula with no free variables. Thus \( A \equiv B \) iff they agree on all first-order properties. The reason we insist on sentences, is that we typically don’t care how \( A \) and \( B \) assign default values to the variables.

**Theorem 7.4 (Isomorphism implies elementary equivalence)** Let \( A, B \in \text{STRUCT}[\Sigma] \).

If \( A \cong B \), then \( A \equiv B \).

The proof of Thm. 7.4 is to assume that \( A \cong B \), and to prove by induction on \( \varphi \in \mathcal{L}(\Sigma) \) that \( (A \models \varphi) \iff (B \models \varphi) \).

### 7.2 Minimum Distinguishing Formulas: Quantifier Depth and Number of Variables

**Definition 7.5 [Quantifier Rank]** The quantifier rank, sometimes called quantifier depth, denoted \( qr \), is the maximum depth of nesting of quantifiers in a formula. We define this inductively:

- **Base case**: \( qr(R(t_1, \ldots, t_n)) = 0 \).
- **Inductive case one**: \( qr(\neg \alpha) = qr(\alpha) \).
- **Inductive case two**: \( qr(\alpha \lor \beta) = qr(\alpha \land \beta) = \max(qr(\alpha), qr(\beta)) \).
- **Inductive case three**: \( qr(\exists \alpha) = qr(\forall x(\alpha)) = 1 + qr(\alpha) \).

**Example 7.6** Let \( \alpha = \exists x(R(x)) \land \forall x \exists y(B(y) \rightarrow E(x, y) \land R(y)) \). Then \( qr(\alpha) = 2 \).

We can also calculate the number of distinct variables needed to express a property.

Let \( \text{dist}_{\leq 1}(x, y) \equiv (x = y \lor E(x, y)) \) meaning that the distance from \( x \) to \( y \) is at most 1. The most natural way to write \( \text{dist}_{\leq 4}(x, y) \) would be to use 3 extra variables:

\[
\varphi_4(x, y) \equiv \exists z_1 z_2 z_3 (\text{dist}_{\leq 1}(x, z_1) \land \text{dist}_{\leq 1}(z_1, z_2) \land \text{dist}_{\leq 1}(z_2, z_3) \land \text{dist}_{\leq 1}(z_3, y))
\]

Note that \( \varphi_4(x, y) \in \mathcal{L}^5(\Sigma) \), meaning that it uses at most 5 distinct variables. Can we write \( \text{dist}_{\leq 4}(x, y) \) using fewer than 5 variables?

Let’s use recursion:

\[
\begin{align*}
\text{dist}_{\leq 4}(x, y) & \equiv \exists (\text{dist}_{\leq 2}(x, z) \land \text{dist}_{\leq 2}(z, y)) \\
\text{dist}_{\leq 2}(x, y) & \equiv \exists (\text{dist}_{\leq 1}(x, z) \land \text{dist}_{\leq 1}(z, y))
\end{align*}
\]

Using the above recursion and dropping the rectification restriction, it is possible to express \( \text{dist}_{\leq n}(x, y) \) as a formula in \( \mathcal{L}^3 \). For example,

\[
\text{dist}_{\leq 4}(x, y) \equiv \exists y (\text{dist}_{\leq 1}(x, y) \land \text{dist}_{\leq 1}(y, z)) \land \exists x (\text{dist}_{\leq 1}(z, x) \land \text{dist}_{\leq 1}(x, y))
\]
Using this scheme in general, and noting that with one more depth of nesting of quantifiers we double the distance, we can show the following by induction on $\lceil \log n \rceil$:

**Proposition 7.7** For all $n \geq 1$ we can write $\text{dist} \leq n \in L^3_{\lceil \log n \rceil}$, i.e., using 3 variables and quantifier rank $\lceil \log n \rceil$.

### 7.3 Ehrenfeucht-Fraïssé Games

**Definition 7.8** [Ehrenfeucht-Fraïssé Games] An Ehrenfeucht-Fraïssé Game, $G^k_m(A, B)$, is a two-player game, with two players, Samson and Delilah, on two structures $A, B \in \text{STRUC}[\Sigma]$, where $\Sigma$ is a finite relational vocabulary – finitely many relation and constant symbols and no function symbols of arity greater than 0.

$G^k_m(A, B)$ has $m$ moves and $k$ pairs of pebbles: $(x_1, x_1), \ldots, (x_k, x_k)$.

At each step, Samson places a one of the pebbles, $x_i$ on one element of the universe of one of the two structures. Delilah responds by placing the other $x_i$ pebble on an element of the other structure. Thus after this move, the default assignments of $x_i^A$ and $x_i^B$ have been changed.

Let $x_i^A$ be the element of $|A|$ just chosen and $x_i^B$ is the element of $|B|$ just chosen.

**Delilah wins** $(A \sim^k_m B)$ if after every step, $j \leq m$ the function, $\eta_j$ that maps $c^A \mapsto c^B$ for constant symbols $c \in \Sigma$ and $x_i^A \mapsto x_i^B$ for all pebbles $x_i$ that have been placed so far, is an isomorphism of the induced substructures of $A$ and $B$. Note: any constant symbols, $c \in \Sigma$, are considered permanently chosen points, so $\eta_j$ is defined on the constants as well, i.e., $\eta_j(c^A) = c^B$. **Samson wins** if at any step, $\eta_j$ is not an isomorphism. □

As an example, let’s play the game $G^2_6(G, H)$ for the graphs $G, H$ shown again below. We have two pairs of pebbles and up to six moves. Suppose that in the first move, Samson decides to place $x_1$ on vertex $g_2 \in |G|$. Delilah must reply with a green vertex, e.g., $h_5$, from $|H|$. Thus, after this first move, $x_1^G = g_2$, $x_1^H = h_5$ and $x_2^G$ and $x_2^H$ are not yet defined. Note that Delilah has not lost because the map $\eta_1 = \{(g_2, h_5)\}$ is an isomorphism of the induced substructures, $G_1 = (\{g_2\}, \emptyset, \emptyset, \{g_2\}, \emptyset)$, $H_1 = (\{h_5\}, \emptyset, \emptyset, \{h_5\}, \emptyset)$, each consisting of a single green vertex and no edges.

---

1We will talk about substructures soon. The idea is that the induced substructures consist of universes $\{c_1^A, \ldots, c_t^A, x_1^A, \ldots, x_k^A\} \subseteq |A|$ and $\{c_1^B, \ldots, c_t^B, x_1^B, \ldots, x_k^B\} \subseteq |B|$ and they inherit their interpretations of all of the symbols of $\Sigma$ from their parent structures.
In the second move, Samson might choose a vertex in $H$, e.g., $h_6$. In order not to lose, Delilah must preserve the isomorphism. Note that now $H \models G(x_1) \land B(x_2) \land E(x_1, x_2)$. Delilah must place $x_2$ on a vertex in $G$ which is blue and has an edge to $g_2$. Fortunately for her, there is such a vertex, namely $g_6$. Thus Delilah places $x_2$ on $g_6$ and the function $\eta_2 = \{(g_2, h_5), (g_6, h_6)\}$ is an isomorphism of the induced substructures, $G_2 = \{(g_2, g_6), ((g_2, g_6), (g_6, g_2)), \emptyset, \emptyset, \{g_2\}, \{g_6\}\}$, $H_2 = \{(h_5, h_6), ((h_5, h_6), (h_6, h_5)), \emptyset, \{h_5\}, \{h_6\}\}$.

Thus, Delilah has not lost.

**Notation:** We write $A \sim^k_m B$ to mean that Delilah has a winning strategy in the game $G^k_m(A, B)$.

It is not hard to see that wherever Samson places his pebble, Delilah always has a good answer, i.e., on a vertex of the same color which agrees with the other structure on whether it does nor does not have an edge to the other chosen point. It follows that

**Proposition 7.9** For all $n \in \mathbb{N}$, $G \sim^2_n H$.

However, it is easy to see that $G \not\sim^3_3 H$, i.e., Samson can win the three pebble game. His winning strategy would be to put his three pebbles on $g_1, g_2, g_6$. Samson would be playing the sentence, “has-triangle”, which is true in $G$ but not in $H$.

\[
\text{has-triangle} \equiv \exists x_1 x_2 x_3 \ (E(x_1, x_2) \land E(x_2, x_3) \land E(x_3, x_1))
\]
Next time we will prove the following theorem which shows that Ehrenfeucht-Fraïssé games exactly characterize the number of variables, via the number of pebble pairs and the quantifier rank, via the numbers of moves, needed to distinguish two structures:

**Theorem 7.10 (Fundamental Thm of Ehrenfeucht-Fraïssé Games)**  *For all finite relational vocabularies, Σ, let A, B ∈ Σ. Then the following two conditions are equivalent:*

- A ≡_m^k B
- A ∼_m^k B

As a corollary of Prop. [7.9], we see that two variables cannot distinguish the graphs G and H.

**Corollary 7.11**  \( G \equiv^2 H \).

Here, \( A \equiv^k B \) is shorthand for \( \forall m \ A \equiv_m^k B \).