Chapter 6

Pebble Games

We introduce combinatorial games that are useful for determining what can and what cannot be expressed in various logics. Pebble games offer a semantics for first-order logic that is equivalent, but more directly applicable, than the standard definitions.

6.1 Definition of the Games

We begin by defining the pebble games, due originally to Ehrenfeucht and Fraïssé, that will be the subject of this chapter.

Definition 6.1 The game $\mathcal{G}_m^k$ has two players: the Spoiler ($\mathcal{S}$) and the Duplicator ($\mathcal{D}$). It is played on a pair of structures $\mathcal{A}$ and $\mathcal{B}$ of the same vocabulary, τ. $\mathcal{G}_m^k$ is played for m rounds, using k pairs of pebbles. As we will see, $\mathcal{S}$ is trying to point out a difference between the two structures. $\mathcal{D}$ is trying to match his moves so that the differences between them are hidden. In order to keep the two players straight, we will refer to $\mathcal{S}$ with the pronoun, "he", and $\mathcal{D}$ with the pronoun, "she."

At each move, $\mathcal{S}$ places one of the pebbles on an element of the universe of one of the two structures, i.e., he places the i^{th} pebble on an element of $|\mathcal{A}|$, or $|\mathcal{B}|$. $\mathcal{D}$ then responds by placing the other i^{th} pebble an element of the other structures.

The position of the game immediately after move r will be denoted (α_r, β_r). Such a *k-configuration* of $\mathcal{A}, \mathcal{B}$ is a pair of partial functions

$$\begin{array}{rcl} \alpha & : & (\mathrm{const}(\tau) \cup \{x_1, x_2, \ldots, x_k\}) \rightarrow |\mathcal{A}| \\ \beta & : & (\mathrm{const}(\tau) \cup \{x_1, x_2, \ldots, x_k\}) \rightarrow |\mathcal{B}| \end{array} \tag{6.2}$$

where we always require that $\mathrm{dom}(\alpha) = \mathrm{dom}(\beta)$, and for all $c \in \mathrm{const}(\tau)$, $\alpha(c) = c^{\mathcal{A}}$, and $\beta(c) = c^{\mathcal{B}}$.

The meaning of $\alpha_r(x_i) = a$ and $\beta_r(x_i) = b$ is that just after move r, the i^{th} pebbles are sitting on $a \in |\mathcal{A}|$ and $b \in |\mathcal{B}|$. Some variable x_i is not in the domain of α_r iff just after move r, the i^{th} pebbles are off the board. The valid positions of the game $\mathcal{G}^k_m$ on $\mathcal{A}, \mathcal{B}$ consist of any possible k-configuration on $\mathcal{A}$ and $\mathcal{B}$. Write $\mathcal{G}^k_m(\mathcal{A}, \alpha_0, \mathcal{B}, \beta_0)$ to denote the k-pebble, m move game on $\mathcal{A}, \mathcal{B}$, with initial configuration (α_0, β_0). $\mathcal{G}^k_m(\mathcal{A}, \mathcal{B})$ is the game in which all the pebbles start off the board, i.e., $\mathrm{dom}(\alpha_0) = \mathrm{dom}(\beta_0) = \mathrm{const}(\tau)$. The reason we include the constants in the domain of every configuration is to make the statement of conditions simpler in what follows. As we will see, in pebble games constants behave exactly liked pebbles that are fixed at the beginning of the game.

At each move r, $1 \le r \le m$, $\mathcal{S}$ picks up a pair of pebbles, and places one of them on an element of one of the two structures. $\mathcal{D}$ must then answer by placing the other pebble in the pair on an element of the other structure. Thus, for some $i \in \{1, 2, \ldots, k\}$, the i^{th} pair of pebbles is placed on $a \in |\mathcal{A}|$ and $b \in |\mathcal{B}|$. Define the next configuration $(\alpha_r, \beta_r) = (\alpha_{r-1}, a/x_i), (\beta_{r-1}, b/x_i)$,

$$\alpha_r(x_j) = \begin{cases} \alpha_{r-1}(x_j) & \text{if } i \ne j \\ a & \text{if } i = j \end{cases}, \qquad \beta_r(x_j) = \begin{cases} \beta_{r-1}(x_j) & \text{if } i \ne j \\ b & \text{if } i = j \end{cases}$$

Just after move r, the configuration α_r, β_r determines a relation $\beta_r \circ \alpha_r^{-1} \subseteq |\mathcal{A}| \times |\mathcal{B}|$. We say that $\mathcal{D}$ *wins round r of the game* iff the map $\alpha_r(x_i) \mapsto \beta_r(x_i)$, for $x_i \in \mathrm{dom}(\alpha_r)$ determines an isomorphism of the induced substructures[1].

$$\beta_r \circ \alpha_r^{-1} : \langle \mathrm{rng}(\alpha) \rangle^{\mathcal{A}} \;\cong\; \langle \mathrm{rng}(\beta) \rangle^{\mathcal{B}} \tag{6.3}$$

This means in particular that $\beta_r \circ \alpha_r^{-1}$ must be a 1:1 function and so $\alpha_r(x_i) = \alpha_r(x_j)$ iff $\beta_r(x_i) = \beta_r(x_j)$. Furthermore, all constants and relations of the structures must be preserved. So for example if the vocabulary τ includes the constant symbol c and the binary relation symbol E, then we must have that the pair $\langle c^{\mathcal{A}}, \alpha_r(x_i) \rangle \in E^{\mathcal{A}}$ iff $\langle c^{\mathcal{B}}, \beta_r(x_i) \rangle \in E^{\mathcal{B}}$. This can be more easily written as follows:

$$(\mathcal{A}, \alpha_r) \models E(c, x_i) \quad \Leftrightarrow \quad (\mathcal{B}, \beta_r) \models E(c, x_i) \ .$$

[1]The induced substructure $\langle \mathrm{rng}(\alpha) \rangle^{\mathcal{A}}$ has universe the closure of $\mathrm{rng}(\alpha)$ under all the functions of $\mathcal{A}$. When τ has no function symbols, this simply means that we add all of the constants to the universe. For $\tau = \langle R_1^{a_1}, \ldots, R_r^{a_r}, c_1, \ldots, c_s \rangle$, $|\langle S \rangle| = S \cup \{c_1^{\mathcal{A}}, \ldots, c_s^{\mathcal{A}}\}$. The meaning of induced substructure is that the relations of $\langle \mathrm{rng}(\alpha) \rangle^{\mathcal{A}}$ are the restriction of the relations of $\mathcal{A}$ to the universe of $\langle \mathrm{rng}(\alpha) \rangle^{\mathcal{A}}$.

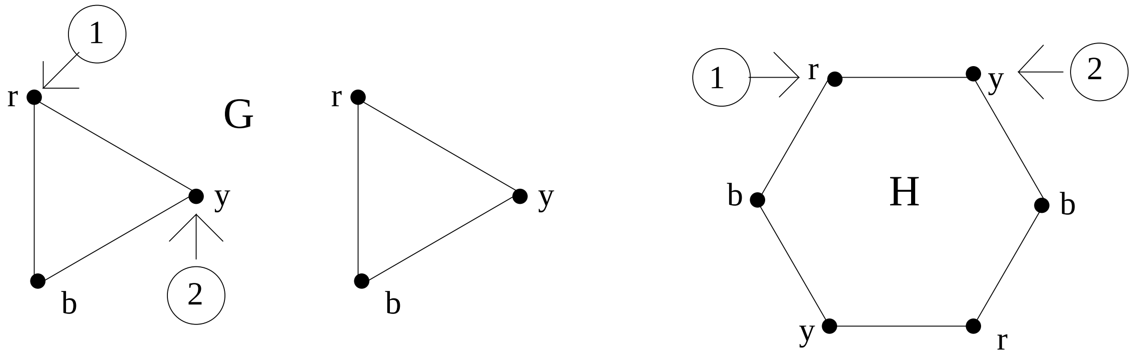

Figure 6.5: A two-pebble game

$\mathcal{D}$ *wins the game* iff she wins every single round. $\mathcal{D}$ must preserve an isomorphism at all times. If a difference between the two structures is ever exposed, then $\mathcal{S}$ wins.

Since $\mathcal{G}_m^k(\mathcal{A}, \alpha_0, \mathcal{B}, \beta_0)$ is a finite game of perfect information, one of the two players must have a winning strategy. We use the notation $(\mathcal{A}, \alpha_0) \sim_m^k (\mathcal{B}, \beta_0)$ to mean that $\mathcal{D}$ has a winning strategy for $\mathcal{G}_m^k(\mathcal{A}, \alpha_0, \mathcal{B}, \beta_0)$. We will write $(\mathcal{A}, \alpha_0) \sim^k (\mathcal{B}, \beta_0)$ to mean that for all m, $(\mathcal{A}, \alpha_0) \sim_m^k (\mathcal{B}, \beta_0)$. □

$\mathcal{D}$ wins the pebble game iff after every round she preserves the condition that $\beta_r \circ \alpha_r^{-1}$ is an isomorphism of the induced substructures. $\mathcal{S}$ is trying to point out a difference between the two structures, and $\mathcal{D}$ is trying to keep them looking the same. An isomorphism preserves all the symbols of τ. It is important to decide whether or not to include the numeric predicates $\leq$ and BIT. If these relations are available in the language in question and thus as part of the definition of isomorphism, then the game becomes much easier for $\mathcal{S}$ and much harder for $\mathcal{D}$. For this reason, *in this chapter, we will assume unless otherwise noted that ordering and BIT are not present.*

Exercise 6.4 Prove that $\sim_m^k$ is an equivalence relation. (This justifies the notation and it is not hard to show directly from the definition of $\mathcal{G}_m^k$. Please do not use Theorem 6.8!) □

As an example, consider the two-pebble game on the colored graphs G and H shown in Figure 6.5. Here the vocabulary $\tau = \langle E^2, R^1, Y^1, B^1 \rangle$ consists of a binary edge relation and three unary relations which may be thought of as colorings of the vertices.

Assume that the initial configuration has all pebbles off the board, so $\alpha_0 = \beta_0 = \emptyset$. Suppose that $\mathcal{S}$'s first move is to place pebble one on a red vertex in G. $\mathcal{D}$ may answer by putting pebble one on either of the red vertices in H. Now suppose $\mathcal{S}$ puts pebble two on an adjacent yellow vertex in H. $\mathcal{D}$ has a response because in G, $\alpha_1(x_1)$ also has an adjacent yellow vertex. On the third move, suppose that $\mathcal{S}$ puts pebble one on the blue vertex in H that is not adjacent to $\beta_2(x_2)$. $\mathcal{D}$ may answer with the blue vertex in G not adjacent to $\alpha_2(x_2)$. The reader should by now be able to prove by induction that,

Proposition 6.6 *Let G and H be the graphs shown in Figure 6.5. For all m, $G \sim^2_m H$, i.e., $G \sim^2 H$.*

On the other hand $\mathcal{S}$ has an easy win for the game $\mathcal{G}^3_3(G, H)$. He can simply choose three points in the same triangle in G on three consecutive moves. To win this three-pebble game, $\mathcal{S}$ is playing the sentence Δ which says that a triangle exists. Note that $G \models \Delta$ while $H \models \neg\Delta$.

$$\Delta \quad \equiv \quad (\exists x_1)(\exists x_2)(\exists x_3)(E(x_1, x_2) \wedge E(x_2, x_3) \wedge E(x_3, x_1))$$

Note that the sentence Δ has three variables, corresponding to the number of pebble pairs in the game. Define the *quantifier rank* of a formula φ, $(\mathrm{qr}(\varphi))$, to be the depth of nesting of quantifiers in φ. Note that the sentence Δ has quantifier rank 3, corresponding to the number of moves in the game.

Definition 6.7 Define the language $\mathcal{L}^k$ to be the restriction of the language $\mathcal{L}$ in which only variables $x_1, \ldots, x_k$ occur. Define the language $\mathcal{L}^k_m$ to be the restriction of $\mathcal{L}^k$ to formulas of quantifier rank at most m. Let $\mathcal{A}$ and $\mathcal{B}$ be two structures of some vocabulary τ. We will say that $\mathcal{A}$ and $\mathcal{B}$ are $\mathcal{L}$ *equivalent* $(\mathcal{A} \equiv_{\mathcal{L}} \mathcal{B})$ iff they agree on all formulas from $\mathcal{L}$,

$$\begin{array}{rll} \mathcal{A} \equiv \mathcal{B} & \text{iff} & \text{for all } \varphi \in \mathcal{L}(\tau),\ (\mathcal{A} \models \varphi \ \Leftrightarrow\ \mathcal{B} \models \varphi) \\ \mathcal{A} \equiv^k_m \mathcal{B} & \text{iff} & \text{for all } \varphi \in \mathcal{L}^k_m(\tau),\ (\mathcal{A} \models \varphi \ \Leftrightarrow\ \mathcal{B} \models \varphi) \end{array}$$

□

We can now state and prove the fundamental theorem of Pebble Games:

Theorem 6.8 *Let $\mathcal{A}$ and $\mathcal{B}$ be structures of the same vocabulary and let α_0, β_0 be a k-configuration of $\mathcal{A}, \mathcal{B}$. Then the following are equivalent:*

1. $(\mathcal{A}, \alpha_0) \sim^k_m (\mathcal{B}, \beta_0)$

2. $(\mathcal{A}, \alpha_0) \equiv^k_m (\mathcal{B}, \beta_0)$

Proof We prove the equivalence of (1) and (2) by induction on m. For $m = 0$, Player II wins the zero move game iff the relation $\beta_0 \circ \alpha_0^{-1}$ is an isomorphism of the induced substructures. This is true just if for every quantifier free formula $\gamma \in \mathcal{L}(\tau)$,

$$(\mathcal{A}, \alpha_0) \models \gamma \quad \Leftrightarrow \quad (\mathcal{B}, \beta_0) \models \gamma \ .$$

Note that α may have as free variables only those variables occurring in $\mathrm{dom}(\alpha_0)$ $= \mathrm{dom}(\beta_0)$. Thus (1) and (2) are equivalent for $m = 0$.

Assume the theorem is true for m, and suppose that $\mathcal{A}$ and $\mathcal{B}$ disagree on the formula $\varphi \in \mathcal{L}^k_{m+1}$. Note that if φ is $\alpha \wedge \beta$ then $\mathcal{A}$ and $\mathcal{B}$ must disagree on one of α and β. Similarly, if φ is $\neg\alpha$, then they disagree on α, so we may assume that φ is $(\exists x_i)\psi$. Suppose that $(\mathcal{A}, \alpha_0) \models \varphi$ and $(\mathcal{B}, \beta_0) \models \neg\varphi$. $\mathcal{S}$'s first move in $\mathcal{G}^k_{m+1}(\mathcal{A}, \alpha_0, \mathcal{B}, \beta_0)$ is to place the i^{th} pebble on a witness for ψ in $\mathcal{A}$. Wherever $\mathcal{D}$ responds, it will not be a witness for ψ because there is none in $\mathcal{B}$. Thus after the first move we have that $(\mathcal{A}, \alpha_1)$ and $(\mathcal{B}, \beta_1)$ disagree on the quantifier depth m formula ψ. By the inductive hypothesis, $\mathcal{S}$ has a winning strategy for the remaining m move game. Thus we have shown that (1) implies (2).

Conversely, suppose that $(\mathcal{A}, \alpha_0) \equiv^k_{m+1} (\mathcal{B}, \beta_0)$. Now let $\mathcal{S}$ make his first move in the game $\mathcal{G}^k_{m+1}(\mathcal{A}, \alpha_0, \mathcal{B}, \beta_0)$. Suppose that he places the i^{th} pebble on an element of $\mathcal{A}$, thus defining α_1. Note that there are only finitely many inequivalent formulas in $\mathcal{L}^k_m$, cf. Exercise 6.9. Let Φ be the conjunction of all these formulas satisfied by $(\mathcal{A}, \alpha_1)$. Thus, we know that

$$(\mathcal{A}, \alpha_0) \models (\exists x_i)\Phi$$

And so, by assumption,

$$(\mathcal{B}, \beta_0) \models (\exists x_i)\Phi \ .$$

$\mathcal{D}$'s answer is to place the other i^{th} pebble on a witness in $\mathcal{B}$ of Φ. Thus we have that $(\mathcal{A}, \alpha_1)$ and $(\mathcal{B}, \beta_1)$ both satisfy Φ, a complete description of every formula from $\mathcal{L}^k_m$ that $(\mathcal{A}, \alpha_1)$ satisfies. Thus $(\mathcal{A}, \alpha_1) \equiv^k_m (\mathcal{B}, \beta_1)$. It follows by induction that $\mathcal{D}$ has a winning strategy for the remaining m moves of the game. $\square$

In the following exercise, you are asked to prove the lemma that there are only finitely many inequivalent formulas of a given quantifier rank. This was needed for the proof that (2) $\Rightarrow$ (1) in Theorem 6.8. As the exercise shows, this lemma holds for infinite structures as well, as long as the vocabulary is finite and has no function symbols.

Exercise 6.9 Theorem 6.8 requires the lemma that there are only finitely many inequivalent formulas of quantifier rank r.

1. As usual, let τ be a finite, relational vocabulary. Prove that there are only finitely many inequivalent first-order formulas in $\mathcal{L}_r(\tau)$. [Hint: induction on r.]

2. Let τ be a finite vocabulary which may include function symbols. Let $\Gamma \subset \mathcal{L}(\tau)$ be a first-order theory. Suppose that for every model $\mathcal{A}$ of Γ — including infinite models — and for any finite set $S \subseteq |\mathcal{A}|$, we have that $\langle S \rangle^{\mathcal{A}}$ — the induced substructure of $\mathcal{A}$ generated by S — is finite. Prove that the theory Γ admits only finitely many inequivalent formulas of quantifier rank r. In this case, φ and ψ are equivalent iff $\Gamma \vdash \varphi \leftrightarrow \psi$.

3. Give a counterexample to Theorem 6.8 when $\tau = \langle R_1^1, R_2^1, \ldots \rangle$ consists of infinitely many unary relation symbols.

4. Give a counterexample to Theorem 6.8 when $\tau = \langle R^1, f^1 \rangle$ consists of one unary relation symbol, and one unary function symbol.

$\Box$

Theorem 6.8 lets us determine precisely how many variables and how much quantifier rank is needed to express a given query. Here are two examples.

Proposition 6.10 *Let* CLIQUE(k) *be the set of undirected graphs that contain a clique, i.e., a complete subgraph, of size k.* CLIQUE(k) *is expressible with k variables, but not $k-1$ variables in the language without ordering,*

$$\text{CLIQUE}(k) \quad \in \quad \mathcal{L}^k(\tau_g - \{\leq\}) \; - \; \mathcal{L}^{k_1}(\tau_g - \{\leq\})$$

Proof It is easy to write CLIQUE(k) in $\mathcal{L}^k$:

$$(\exists x_1 x_2 \ldots x_k)(\text{distinct}(x_1, \ldots, x_k) \wedge E(x_1, x_2) \wedge \ldots \wedge E(x_1, x_k) \wedge \ldots E(x_{k-1}, x_k))$$

To prove that k variables are necessary, we show that $K_k \sim^{k-1} K_{k-1}$, where K_r is the complete graph on r vertices. $\mathcal{D}$ has a simple winning strategy for the game $\mathcal{G}_{k-1}(K_k, K_{k-1})$: When $\mathcal{S}$ places the i^{th} pebble on an unpebbled vertex in one of the two graphs, $\mathcal{D}$ places the other i^{th} pebble on any unpebbled vertex in the other graph. Since there are only $k-1$ pebble pairs, such an unpebbled vertex is always available. Note that this is a winning strategy since edges exists between all points in each graph. Thus any 1:1 correspondence is an isomorphism. □

The following exercise shows that we never have to consider a move of a game in which $\mathcal{S}$ pebbles an element that is already pebbled by another pebble or constant.

Exercise 6.11 Prove that in any game, $\mathcal{G}_m^k(\mathcal{A}, \alpha_0, \mathcal{B}, \beta_0)$, if $\mathcal{S}$ has a winning strategy, then he still has a winning strategy if he is never allowed to place a pebble on a constant or an element that already has another pebble sitting on it. □

Proposition 6.12 *Let the formula* $\mathrm{PATH}_k(x, y) \in \mathcal{L}(\tau_g)$ *mean that there is a path of length at most* 2^k *from* x *to* y*. Then quantifier-rank* k *is necessary and sufficient to express* PATH_k*, with or without ordering. Furthermore,* PATH_k *is expressible in this quantifier rank using only three variables. In symbols,*

$$\mathrm{PATH}_k \quad \in \quad \mathcal{L}_k^3(\tau_g - \{\leq\}) \; - \; \mathcal{L}_{k-1}(\tau_g)$$

Proof For the upper bound, we express PATH_k inductively as follows:

$$\begin{aligned} \mathrm{PATH}_0(x, y) \quad &\equiv \quad x = y \; \vee \; E(x, y) \\ \mathrm{PATH}_{k+1}(x, y) \quad &\equiv \quad (\exists z)(\mathrm{PATH}_k(x, z) \; \wedge \; \mathrm{PATH}_k(z, y)) \end{aligned}$$

Thus, PATH_k is expressible using three variables and quantifier rank k. Only three variables are needed because the right hand side of the inductive definition of $\mathrm{PATH}_{k+1}(x, y)$ may be written in a way that reuses variables:

$$\begin{aligned} \mathrm{PATH}_k(x, z) \quad &\equiv \quad (\exists y)(\mathrm{PATH}_{k-1}(x, y) \; \wedge \; \mathrm{PATH}_{k-1}(y, z)) \\ \mathrm{PATH}_k(z, y) \quad &\equiv \quad (\exists x)(\mathrm{PATH}_{k-1}(z, x) \; \wedge \; \mathrm{PATH}_{k-1}(x, y)) \end{aligned}$$

For the lower bound, let $L_n \in \mathrm{STRUC}[\tau_g]$ be a directed line segment of length $n - 1$, so $\|L_n\| = n$. Let the ordering on L_n be from left to right, see Figure 6.13.

Letting $n = 2^{k+1} + 1$, we have that,

$$L_n \models \mathrm{PATH}_{k+1}(0, max); \qquad L_{n+1} \models \neg\mathrm{PATH}_{k+1}(0, max)$$

We now show that $L_n \sim_k L_{n+1}$ and the lower bound follows.

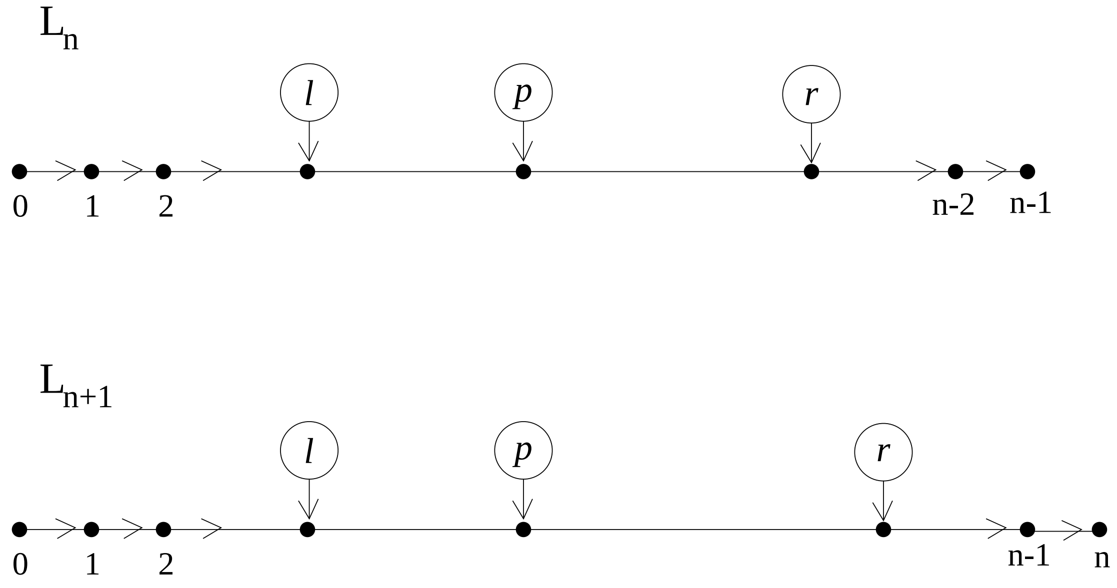

Figure 6.13: With $n = 2^{k+1} + 1,\ L_n \sim_k L_{n+1}$

The idea behind $\mathcal{D}$'s winning strategy is that in quantifier rank s — or equivalently, with s moves remaining in the game — distances greater than 2^s are indistinguishable from infinite distances. With this idea in mind, let us define the notion $i =_d j$ to mean that i and j are equal or that i and j are both greater than d. $\mathcal{D}$'s winning strategy in $\mathcal{G}_k(L_n, L_{n+1})$ is to maintain the following invariant:

$$\begin{array}{c} \text{After the } m^{th} \text{ move of } \mathcal{G}_k(L_n, L_{n+1}), \text{ for } p, q \in \operatorname{dom}(\alpha_m), \\ \operatorname{DIST}(\alpha_m(p), \alpha_m(q)) \quad =_{2^{k-m}} \quad \operatorname{DIST}(\beta_m(p), \beta_m(q)) \end{array} \tag{6.14}$$

Note that Equation 6.14 implies that $\mathcal{D}$ wins the game because a map that preserves distances of length at most one is an isomorphism. Equation 6.14 holds after the 0^{th} move because

$$\operatorname{DIST}(0^{L_n}, max^{L_n}) \quad = \quad 2^{k+1} \quad =_{2^k} \quad 2^{k+1} + 1 \quad = \quad \operatorname{DIST}(0^{L_{n+1}}, max^{L_{n+1}})$$

Assume inductively that Equation 6.14 holds just after the m^{th} move. Let $\mathcal{S}$ begin the $m + 1^{st}$ move by placing pebble p on some vertex. Let l and r be the closest pebbles to the left and right of p. Our inductive assumption tells us that $\operatorname{DIST}(\alpha_m(l), \alpha_m(r)) =_{2^{k-m}} \operatorname{DIST}(\beta_m(l), \beta_m(r))$. Assume without loss of generality that l is the closer of the two pebbles to p, or that they are equidistant. $\mathcal{D}$'s response is to place the other pebble p on the point to the right of the other l so that $\mathrm{DIST}(\alpha_{m+1}(l), \alpha_{m+1}(p)) = \mathrm{DIST}(\beta_{m+1}(l), \beta_{m+1}(p))$. It follows, that $\mathrm{DIST}(\alpha_{m+1}(p), \alpha_{m+1}(r)) =_{2^{k-(m+1)}} \mathrm{DIST}(\beta_{m+1}(p), \beta_{m+1}(r))$ and so Equation 6.14 holds after move $m+1$. Thus $\mathcal{D}$ wins the game, and we have proved that PATH_k is not expressible with quantifier rank less than k, even for ordered structures. $\square$

6.2 Methodology for First-Order Expressibility

As we now show, pebble games provide a complete methodology for proving that a query is not first-order.

Theorem 6.15 (Methodology Theorem) *Let $\mathcal{C}$ be any class of finite or infinite structures of some finite, relational vocabulary. Let $S \subseteq \mathcal{C}$ be a boolean query on $\mathcal{C}$. To prove that S is* not *first-order describable on $\mathcal{C}$ it is necessary and sufficient to show that for all $r \in \boldsymbol{N}$, there exist structures $\mathcal{A}_r, \mathcal{B}_r \in \mathcal{C}$ such that*

1. *$\mathcal{A}_r \in S$ and $\mathcal{B}_r \notin S$, and*
2. *$\mathcal{A}_r \sim_r \mathcal{B}_r$*

Proof It is easy to see that the methodology is sufficient. The above two conditions imply that $\mathcal{A}_r$ and $\mathcal{B}_r$ agree on all formulas in $\mathcal{L}_r$, but disagree on S. Thus, S is not expressible in $\mathcal{L}_r$ for any r.

Conversely, suppose that S is not first-order expressible over $\mathcal{C}$. Recall from Exercise 6.9 that for any fixed r there are only a bounded number of inequivalent sentences of quantifier rank r. Let $\varphi_1, \varphi_2, \ldots, \varphi_B$ be a list of all inequivalent, complete quantifier-rank r sentences. For every quantifier-rank r sentence ψ, each φ_i must either assert ψ or $\neg\psi$. Observe that each structure from $\mathcal{C}$ satisfies a unique φ_i. Suppose that there are structures $\mathcal{A}_r \in S$ and $\mathcal{B}_r \in \mathcal{C} - S$ that satisfy the same φ_i. Then $\mathcal{A}_r$ and $\mathcal{B}_r$ satisfy the above conditions. If there is no such pair, then the φ_i's are partitioned by S. In this case, let $Y = \{i \mid (\exists \mathcal{A} \in S)(\mathcal{A} \models \varphi_i)\}$ and let,

$$\varphi \quad \equiv \quad \bigvee_{i \in Y} \varphi_i$$

Then φ is a first-order formula of quantifier rank r that expresses S. $\square$