# Note on Finite Model Theory and Game Comonads

Elena Dimitriadis, Tyler Hanks, Zhixuan Yang, Richie Yeung

July, 2023

This note is a part of the authors' participation in the Adjoint School 2023, mentored by Dan Marsden and Nihil Shah.

# Part I: Finite Model Theory and Games

# Introduction

Finite model theory (Libkin 2004) studies finite models of logics. Its main motivation comes from computer science: a *finite relational structure*, i.e. a finite set A with a finite set of relations on A, is essentially a database in the sense of good old SQL tables, and a logic formula  $\varphi$  with n free variables is understood as a *query* to the database that selects all n-tuples of A that satisfy the formula  $\varphi$ .

Finite model theory is naturally related to *complexity theory*, as we may ask questions like what's the time complexity to query a finite relational structure with a formula from some logic, and also the converse question—what kind of logic is needed to describe the algorithmic problems in a complexity class. For example, given a finite graph G, we may ask if it is possible to write a first-order logic formula  $\varphi(u, v)$  using a relation symbol E(x, y) saying there is an edge from x to y, such that  $\varphi(u, v)$  is satisfied precisely by vertices  $u, v \in G$  that are connected by a path.

From a logical point of view, what makes finite model theory interesting is that some prominent techniques in model theory fail for finite models. In particular, compactness fails for finite model theory: a theory T may not have a *finite* model even if all its finite sub-theories  $S \subseteq T$  have finite models—consider e.g.  $T = \{\varphi_n \mid n \in \mathbb{N}\}$  where  $\varphi_n$  asserts there are at least n distinct elements.

Fortunately there are model-theoretic techniques remaining valid in the finite context. One of them is *model-comparison games*, which characterise *logical equivalence* of models, i.e. when two models satisfy exactly the same formulas of a logic.

Such logical equivalences are very useful for showing *(in)expressivity* of logics. For example, if we are able to show that two models  $\mathcal{A}$  and  $\mathcal{B}$  of a theory satisfy exactly the same formulas from first-order logic, but there is a semantic property P (in the meta-theory) which  $\mathcal{A}$  satisfies but  $\mathcal{B}$  does not. Then we know the property P necessarily cannot be expressed in first-order logic. This proof technique works for both finite and infinite models. In fact, using this technique one can show that connectivity of finite graphs cannot be expressed as a first-order logic formula with only the relation symbol E(x, y) for edges.

Of course the logic equivalences for different logics need to be characterised by different games: first-order logic is characterised by *Ehrenfeucht–Fraïssé games*; the *k*-variable fragment of first-order logic is characterised by *pebble games*; modal logic is characterised by *bisimulation games*, and so on.

Despite being different, these games are structurally so similar that they almost begged to be unified. Their unification is eventually done by Abramsky and Shah (2018, 2021) in the categorical framework of *game comonads*. This two-part blog post aims to give a brief introduction to finite model theory and

game comonads. This post will cover only a tiny part of this active research programme, but we will aim to exposit in a self-contained way the basics with some intuition that is hidden in the papers.

# **Basics of First-Order Logic**

Let's begin with a quick recap on classical first-order logic (FOL). A purely relational vocabulary  $\sigma$  is a set of relation symbols  $\{P_1, \ldots, P_i, \ldots\}$  where each relation symbol  $P_i$  has an associated arity  $n_i \in \mathbb{N}$ . In this post we do not consider vocabularies with function symbols or constants since they can be alternatively modelled as relations with axioms asserting their functionality, which slightly makes our life easier.

The formulas  $\varphi$  of FOL in a (purely relational) vocabulary  $\sigma$  is inductively generated by the following grammar, where the meta-variable x ranges over a countably infinite set of variables:

 $\varphi ::= x_1 = x_2 \mid P_i(x_1, \dots, x_{n_i}) \mid \top \mid \varphi_1 \land \varphi_2 \mid \bot \mid \varphi_1 \lor \varphi_2 \mid \neg \varphi \mid \exists x.\varphi \mid \forall x.\varphi$ 

A set T of closed formulas (i.e. formulas with no free variables) is called a *theory*.

We will only consider the classical semantics of FOL in the category of sets in this post. A  $\sigma$ -relational structure, or simply a  $\sigma$ -structure,  $\mathcal{A} = \langle A, \langle P_i^A \rangle_{P_i \in \sigma} \rangle$  consists of a set A and an  $n_i$ -ary relation  $P_i^A \subseteq A^{n_i}$  on the set A for each relation symbol  $P_i \in \sigma$  of arity  $n_i$ .

A homomorphism of  $\sigma$ -structures from  $\mathcal{A}$  to  $\mathcal{B}$  is a function  $h: \mathcal{A} \to \mathcal{B}$  on the underlying sets such that for each relation symbol  $P_i$  in  $\sigma$ , we have  $P_i^{\mathcal{A}}(a_1, \ldots, a_{n_i})$  implies  $P_i^{\mathcal{B}}(h(a_1), \ldots, h(a_{n_i}))$  for all  $a_1, \ldots, a_{n_i} \in \mathcal{A}$ . Each vocabulary  $\sigma$  thus yields a category  $\mathcal{R}(\sigma)$  of  $\sigma$ -structures and homomorphisms.

Let  $\varphi$  be a FOL formula (in the vocabulary  $\sigma$ ) with n free variables. The semantics of  $\varphi$  in a  $\sigma$ -structure  $\mathcal{A}$  is an n-ary relation  $\llbracket \varphi \rrbracket \subseteq A^n$  on the set A. We will write  $\mathcal{A} \models \varphi(\vec{a})$  when some  $\vec{a} \in A^n$  is contained in  $\llbracket \varphi \rrbracket$  and also write  $\mathcal{A} \models \varphi$  when  $\varphi$  is a closed formula and  $\langle \rangle$  is in  $\llbracket \varphi \rrbracket \subseteq A^0$ . The relation  $\llbracket \varphi \rrbracket$  is inductively defined on  $\varphi$  as follows (we implicity treats  $\vec{a} \in A^n$  as a function from the set of the n free variables to the set A):

$\mathcal{A} \models (x_i = x_j)(\vec{a})$	$\iff$	$\vec{a}(x_i) = \vec{a}(x_j)$
$\mathcal{A} \models P_i(x_1, \dots, x_{n_i})(\vec{a})$	$\iff$	$\langle \vec{a}(x_1), \dots, \vec{a}(x_{n_i}) \rangle \in P_i^A$
$\mathcal{A} \models \top$	$\iff$	true
$\mathcal{A} \models (\varphi_1 \land \varphi_2)(\vec{a})$	$\iff$	$\mathcal{A} \models \varphi_1(\vec{a}) \text{ and } \mathcal{A} \models \varphi_2(\vec{a})$
$\mathcal{A} \models \bot$	$\iff$	false
$\mathcal{A} \models (\varphi_1 \lor \varphi_2)(\vec{a})$	$\iff$	$\mathcal{A} \models \varphi_1(\vec{a}) \text{ or } \mathcal{A} \models \varphi_2(\vec{a})$
$\mathcal{A} \models \neg \varphi(\vec{a})$	$\iff$	$\mathcal{A} \models \varphi(\vec{a})$ does not hold
$\mathcal{A} \models (\exists y.\varphi)(\vec{a})$	$\iff$	$\mathcal{A} \models \varphi(\vec{a}[y \mapsto a']) \text{ for some } a' \in \mathcal{A}$
$\mathcal{A} \models (\forall y.\varphi)(\vec{a})$	$\iff$	$\mathcal{A} \models \varphi(\vec{a}[y \mapsto a']) \text{ for all } a' \in A$

where  $\vec{a}[y \mapsto a']$  is the function mapping y to a' and anything else x to  $\vec{a}(x)$ .

### Logical Equivalences and Ehrenfeucht–Fraïssé Games

As motivated earlier, we are interested in characterising when two models  $\mathcal{A}$  and  $\mathcal{B}$  of a relational vocabulary  $\sigma$  satisfy exactly the same formulas, more precisely, when  $\mathcal{A} \models \phi \iff \mathcal{B} \models \phi$  for all closed FOL formulas  $\phi$  in the vocabulary  $\sigma$ . When it is the case,  $\mathcal{A}$  and  $\mathcal{B}$  are sometimes called *elementarily* equivalent.

#### An Example of Logical Equivalence

Let's build up our intuition with a concrete example. Let  $\sigma$  be the vocabulary with just one binary relation symbol  $\leq$ , and let  $\mathcal{A}$  be the model  $\{0, 1, \ldots, 1000\}$  with  $\leq$  being the usual linear order of natural numbers and similarly  $\mathcal{B}$  be the model  $\{0, 1, \ldots, 1001\}$  with the same order. These two models are clearly different, and indeed they can be differentiated by a first-order logic formula in the vocabulary  $\sigma$ —the formula

$$\varphi = \exists x_0 \exists x_1 \cdots \exists x_{1001}, \neg (x_0 = x_1) \land \dots \neg (x_i = x_j) \dots \neg (x_{1000} = x_{1001})$$

saying that there exist 1002 different elements is satisfied by  $\mathcal{B}$  but not by  $\mathcal{A}$ .

However, the formula  $\varphi$  is a pretty big formula—it has 1002 quantifiers and 501501 clauses, so it is possible that small enough formulas cannot differentiate  $\mathcal{A}$  and  $\mathcal{B}$  since they are pretty similar (they are both linear orders). Let's consider some formulas with just a few quantifiers:

- The vocabulary  $\sigma = \{\leq\}$  doesn't have a constant, so there are no closed terms and thus no closed formulas other than  $\perp$  and  $\top$ . Thus  $\mathcal{A}$  and  $\mathcal{B}$  agree on all formulas with 0 quantifiers.
- Consider formulas of the form  $\exists x.\varphi(x)$  where  $\varphi$  doesn't have any quantifiers. We argue that  $\mathcal{A} \models \exists x.\varphi(x)$  iff  $\mathcal{B} \models \exists x.\varphi(x)$  as follows: supposing  $\mathcal{A} \models \exists x.\varphi(x)$  holds, this means that there is some  $a \in \mathcal{A}$  such that  $\mathcal{A} \models \varphi(a)$ , then we may choose any  $b \in \mathcal{B}$ , say  $b = 0 \in \mathcal{B}$ , making  $\mathcal{B} \models \varphi(b)$  because the formula  $\varphi$  is inductively built from the variable x, relations  $\leq$ , =, and propositional connectives; both  $a \leq a$  and  $b \leq b$  are true, so we can inductively show that  $\varphi(a)$  and  $\varphi(b)$  agree for all  $\varphi$ . Conversely, if  $\mathcal{B} \models \exists x.\varphi(x)$  is witnessed by some b, we can always pick an arbitrary  $a \in \mathcal{A}$  witnessing  $\mathcal{A} \models \exists x.\varphi(x)$ .

Moreover, since the semantics of propositional connectives are defined compositionally, we can inductively show that  $\mathcal{A}$  and  $\mathcal{B}$  agree on all closed formulas built out of exactly one  $\exists$  and =,  $\neg$ ,  $\land$ , and  $\lor$ . Since we are considering classical logic, universal quantification  $\forall x.\varphi(x)$  can be reduced to  $\neg \exists x.\varphi(x)$ , so  $\mathcal{A}$  and  $\mathcal{B}$  agree on it so we conclude that  $\mathcal{A}$  and  $\mathcal{B}$  agree on all FOL formulas with exactly one quanfier.

• This example gets interesting when we consider two nested quantifiers. Supposing  $\psi = \exists x. \exists y. \varphi$ where  $\varphi$  is quanfier-free, if  $\mathcal{A} \models \psi$ , there exist a and  $a' \in \mathcal{A}$  such that  $\mathcal{A} \models \varphi(\langle a, a' \rangle)$ . Then we can also choose any two elements  $b, b' \in \mathcal{B}$  such that, importantly, (i)  $b \leq b'$  iff  $a \leq a'$ , and (ii) b = b' iff a = a'. This ensures  $\mathcal{B} \models \varphi(\langle b, b' \rangle)$  since the atomic formulas in  $\varphi$  are built from  $\leq$ , =, x and y, on which  $\mathcal{A}$  with the variable assignment  $\langle x \mapsto a, y \mapsto a' \rangle$  and  $\mathcal{B}$  with  $\langle x \mapsto b, y \mapsto b' \rangle$ agree. Conversely, if  $\mathcal{B} \models \psi$  witnessed by  $b, b' \in \mathcal{B}$  we can also choose a matching pair  $a, a' \in \mathcal{A}$ making  $\mathcal{A} \models \psi$ .

Now suppose  $\psi = \exists x. \forall y. \varphi$ . Whenever  $\mathcal{A} \models \psi$ , there is some a such that  $\mathcal{A} \models \forall y. \varphi \langle x \mapsto a \rangle$ . In this case we can also choose an element  $b \in \mathcal{B}$  that mimics  $a \in \mathcal{A}$ : if a is the bottom element 0 in the structure  $\mathcal{A}$ , we let b = 0 as well; if a is the top 1000 in  $\mathcal{A}$ , we let b be the top 1001 in  $\mathcal{B}$ ; otherwise we can choose an arbitrary 0 < b < 1001. We then claim  $\mathcal{B} \models \forall y. \varphi \langle x \mapsto b \rangle$  as well, because if there is some b' making  $\mathcal{B} \models \varphi \langle x \mapsto b, y \mapsto b' \rangle$  not hold, we can also find an a'that is to a in  $\mathcal{A}$  as b' is to b in  $\mathcal{B}$ : precisely, if b' = b, we let a' = a; if b' < b, we let a' be any element in  $\mathcal{A}$  such that a' < a, and similarly for the case b' > b (such a choice is always possible because earlier a and b are chosen to be at the same relative position). This choice of a' entails  $\mathcal{A} \models \varphi \langle x \mapsto a, y \mapsto a' \rangle$  not true, leading to a contradiction. Therefore  $\mathcal{B} \models \psi$ .

By a symmetric argument,  $\mathcal{B} \models \exists x. \forall y. \varphi$  implies  $\mathcal{A} \models \exists x. \forall y. \varphi$  as well. Moreover, by the compositionality of the semantics of propositional connectives, the two paragraphs above imply that  $\mathcal{A}$  and  $\mathcal{B}$  agree on all FOL formulas with quantifiers are nested at most once.

Hopefully working through the example above reveals the intuition behind logical equivalence for FOL: two  $\sigma$ -structures  $\mathcal{A}$  and  $\mathcal{B}$  agree on a FOL formula  $\varphi$  whenever a quantifier in  $\varphi$  picks an element x in A or B, there is always an element y in the other structure that "simulates the behaviour" of x in the model.

#### The Ehrenfeucht–Fraïssé Game

The intuition is precisely formulated as *Ehrenfeucht–Fraissé* (*EF*) games. Given two  $\sigma$ -structures  $\mathcal{A}$  and  $\mathcal{B}$  for any vocabulary  $\sigma$  and a natural number k, the k-round EF game for  $\mathcal{A}$  and  $\mathcal{B}$  is a turn-based game between two players, called *the spoiler* and *the duplicator*. Roughly speaking, the goal of the spoiler is to point out the difference between  $\mathcal{A}$  and  $\mathcal{B}$  while the goal of the duplicator is to advocate that  $\mathcal{A}$  and  $\mathcal{B}$  are the same. The rules are very simple:

- 1. Movement: At each round, the spoiler picks an element from one of the structures and the duplicator must respond with an element from the other structure. For example, if the spoiler picks an element from the structure  $\mathcal{A}$ , then the duplicator must pick an element  $b \in \mathcal{B}$ .
- 2. Winning Condition: After k rounds, the game state consists of  $\vec{a} = (a_1, \ldots, a_k)$  and  $\vec{b} = (b_1, \ldots, b_i)$  representing the elements chosen from each structure at each round. The duplicator wins this play if the mapping  $a_i \mapsto b_i$  defines a partial isomorphism between  $\mathcal{A}$  and  $\mathcal{B}$ , i.e., if the substructures of  $\mathcal{A}$  and  $\mathcal{B}$  generated by  $\vec{a}$  and  $\vec{b}$  are isomorphic. Otherwise, the spoiler has succeeded in showing the structures are different and wins.

If the duplicator can guarantee a win after k rounds no matter how the spoiler plays, we say the duplicator has a k-round winning strategy.

The quantifier rank  $qr(\varphi)$  of a FOL formula  $\varphi$  is the depth of nesting of the quantifiers in  $\varphi$ :

$$\begin{array}{rcl} qr(\varphi) &=& 0 & \text{for atomic } \varphi \\ qr(o(\varphi_1, \dots, \varphi_n)) &=& \max(qr(\varphi_1), \dots, qr(\varphi_n)) & \text{for propositional connectives } o \\ qr(Qx.\varphi) &=& 1 + qr(\varphi) & \text{for quantifiers } Q = \forall, \exists \end{array}$$

**Theorem** (Ehrenfeucht–Fraïssé). If the duplicator has a winning strategy for the k-round EF game for  $\mathcal{A}$  and  $\mathcal{B}$ ,  $\mathcal{A}$  and  $\mathcal{B}$  agree on all closed FOL formulas of quantifier-rank k. When the vocabulary  $\sigma$  is finite, the converse is also true.

Proof sketch: Assuming a winning strategy for the duplicator, and let  $\psi$  be any formula of quantifier rank k. Without loss of generality, we can assume  $\psi = Q_1 x_1 \dots Q_k x_k \varphi$  where  $Q_i \in \{\forall, \exists\}$  are quantifiers and  $\varphi$  is quantifier-free. We need to show  $\mathcal{A} \models \psi \iff \mathcal{B} \models \psi$ . As we demonstrated in the example above, we consider how  $- \models \psi$  is defined inductively: if a quantifier  $Q_i = \exists$  and one of  $\mathcal{A}$  and  $\mathcal{B}$  satisfies the formula, we use the winning strategy for the duplicator to pick a matching element in the other structure; if a quantifier  $Q_i = \forall$  and one of  $\mathcal{A}$  and  $\mathcal{B}$  satisfies the formula, every counter-witness in this structure by the winning strategy of the duplicator, thus a contradiction.

The converse direction is also interesting and is not demonstrated in the example in the last section. We only give a rough sketch here and refer interested readers to Libkin (2004, Section 3) for details.

Assuming that  $\mathcal{A}$  and  $\mathcal{B}$  agree on all FOL formulas of rank k, the duplicator's strategy is that whenever the spoiler picks an element  $a_i \in \mathcal{A}$  (or symmetrically  $b_i \in \mathcal{B}$ ), the duplicator first constructs a FOL formula  $\phi_i$  that "maximally describes the current board on  $\mathcal{A}$ ". Roughly speaking,  $\phi_i$  is the conjunction of all formulas in the following set

$$\{\psi \mid qr(\psi) = i - 1 \text{ and } \mathcal{A} \models \psi \langle a_1, \dots, a_i \rangle \}$$

A priori, this set may have infinitely many formulas, but since there are only finitely many atomic formulas when  $\sigma$  is finite,  $\phi_i$  can be reduced to a finite formula using the rules of propositional connectives. With the formula  $\phi_i$  in hand, the duplicator can use the fact that

$$\mathcal{A} \models \exists x_i.\phi_i \langle x_1 \mapsto a_1, \dots, x_{i-1} \mapsto a_{i-1} \rangle,$$

witnessed by  $x_i \mapsto a_i$ . Now using the assumption that  $\mathcal{A}$  and  $\mathcal{B}$  agree on FOL up to rank k and the fact that  $\exists x_i.\phi_i$  is of rank i,  $\mathcal{B}$  has an element witnessing the truth of this formula as well, which is going to be the duplicator's response.

EF games are very useful for showing inexpressivity results of FOL. Suppose we are interested in a property P (in the meta-theory) on a class M of  $\sigma$ -structures. If for every natural number k, we can find two models  $\mathcal{A}_k, \mathcal{B}_k \in M$  such that

- 1. the duplicator has a winning strategy for the k-round EF game on  $\mathcal{A}_k$  and  $\mathcal{B}_k$ , but
- 2. only one of  $\mathcal{A}_k$  and  $\mathcal{B}_k$  satisfy the property P,

then by the EF theorem, the property P cannot be expressed by a FOL formula, whatever the quantifier it has.

For example, using this technique, we can show that the evenness of finite linear orders is not expressible there isn't a FOL formula  $\varphi$  in the vocabulary  $\{\leq\}$  such that for every finite linear order  $\mathcal{A} = \langle A, \leq^A \rangle$ ,  $\mathcal{A} \models \varphi$  exactly when A has an even number of elements. (Hint: We pick  $\mathcal{A}_k$  and  $\mathcal{B}_k$  to be the linear order of  $2^k$  and  $2^k + 1$  elements respectively and play the EF games.)

# Part II: Game Comonads

# Introduction

In the Part I of this post, we saw how logical equivalences of first-order logic (FOL) can be characterised by a combinatory game, but there are still a few unsatisfactory aspects of the formulation of EF games in Part I:

- 1. The game was formulated in a slightly informal way, delegating the precise meaning of "turns", "moves", "wins" to our common sense.
- 2. There are variants of the EF game that characterise logical equivalences for other logics, but these closely related games are defined ad hoc rather than as instances of one mathematical framework.
- 3. We have confined ourselves entirely to the classical semantics of FOL in the category of sets, rather than general categorical semantics.

So you, a patron of the n-Category Café, must be thinking that category theory is perfect for addressing these problems! This is exactly what we are gonna talk about today—the framework of *game comonads* that was introduced by Abramsky, Dawar and Wang (2017) and Abramsky and Shah (2018).

(We will not address the third point above in this post though, but hopefully the reader will agree that what we talk about below is a useful first step towards model comparison games for general categorical logic.)

### **One-Way EF Games**

Let's warm up by recalling EF games and considering a simplified version of them.

Recall that a k-round EF game is parameterized by two  $\sigma$ -relational structures  $\mathcal{A}$  and  $\mathcal{B}$  for some relational vocabulary  $\sigma$ . The rule is that in every round  $1 \leq i \leq k$ , the spoiler picks either an element from  $\mathcal{A}$  or an element from  $\mathcal{B}$ , and then the duplicator responds with an element from the other structure. The duplicator wins if after k-rounds, these elements form a partial isomorphism.

In the game the spoiler has the freedom in each round to pick the structure  $\mathcal{A}$  or  $\mathcal{B}$ , so EF games are also sometimes called *back-and-forth* games. We can also consider the *one-way* variant of EF games from  $\mathcal{A}$  to  $\mathcal{B}$ , where the spoiler can only pick elements  $a_i$  from  $\mathcal{A}$  (so the duplicator responds with elements  $b_i$  from  $\mathcal{B}$ ). Additionally, we weaken the winning condition for the duplicator to be  $a_i \mapsto b_i$  forming a partial homomorphism from  $\mathcal{A}$  to  $\mathcal{B}$  (rather than a partial isomorphism).

It is not difficult to modify the Ehrenfeucht–Fraïssé theorem that we saw last time to show that such one-way EF games characterise the fragment of FOL that only uses  $\exists, \land, \lor$ , true, false. This fragment of FOL is known as *existential-positive fragment* of first-order logic or *coherent logic*.

**Theorem** (Existential Ehrenfeucht–Fraïssé). If the duplicator has a winning strategy for the k-round one-way EF game from  $\mathcal{A}$  to  $\mathcal{B}$ , then  $A \models \varphi$  implies  $B \models \varphi$  for all closed formulas  $\varphi$  of quantifier rank k in the existential-positive fragment.

A consequence of the one-way EF games is that now a winning strategy for the duplicator can be thought of as a function from the half-board of  $\mathcal{A}$ -elements  $\langle a_1, \dots, a_i \rangle$  in each round *i* to the duplicator's response  $b_i \in \mathcal{B}$ , instead of a function from the whole board  $\langle \langle a_1, \dots, a_i \rangle, \langle b_1, \dots, b_{i-1} \rangle \rangle$ to their response  $b_i$ . The reason is that the  $\mathcal{B}$ -elements  $\langle b_1, \dots, b_{i-1} \rangle$  were all picked by the duplicator themselves before, so the duplicator knows what they are, and they don't have to be in the input to the duplicator's winning strategy.

Write  $A^{\leq k}$  for the set  $\{\langle a_1, \dots, a_i \rangle \in A^* \mid 1 \leq i \leq k\}$  of non-empty A-sequences of length at most k. Clearly, not every function  $f: A^{\leq k} \to B$  is a valid winning strategy for the duplicator for the one-way EF game, since the duplicator has to make sure their responses  $b_i$  maintain a partial homomorphism to the spoiler's choices  $a_i$ .

More precisely, for all half-boards  $\langle a_1, \dots, a_i \rangle \in A^{\leq k}$ , if some of its elements are related by a relation  $P \in \sigma$  of arity m, i.e.  $P^A(a_{j_1}, \dots, a_{j_m})$  holds for  $1 \leq j_1, \dots, j_m \leq i$ , the duplicator's strategy  $f: A^{\leq k} \to B$  must make

 $P^B(f \langle a_1, \ldots, a_{j_1} \rangle, \ldots, f \langle a_1, \ldots, a_{j_m} \rangle)$  hold.

# The EF Game Comonad

Now we are ready to formulate EF games in a more categorical way using comonads. Recall that the Kleisli presentation a comonad  $(G, \epsilon, (-)^*)$  on a category  $\mathcal{C}$  is given by

- a map  $G: Obj(\mathcal{C}) \to Obj(\mathcal{C}),$
- for all  $A \in \mathcal{C}$ , a *counit*  $\epsilon_A : GA \to A$ , and
- a co-extension operation  $(-)^*$  that takes every morphism  $f:GA\to B$  to another morphism  $f^*:GA\to GB$

such that the following equations hold for all  $f: GA \to B$  and  $g: GB \to C$ :

$$\begin{array}{rcl} g \circ f^*)^* &=& g^* \circ f^* &: GA \to GC \\ id_{GA} &=& \epsilon^*_A &: GA \to GA \\ f &=& \epsilon_B \circ f^* &: GA \to B. \end{array}$$

Given any natural number k, the mapping from sets A to sets  $A^{\leq k}$  of non-empty A-sequences of length at most k can be equipped with a comonad structure  $E_k$  on **Set**:

- $E_k A = A^{\leq k}$  for all sets A;
- $\epsilon_A \langle a_1, \ldots, a_i \rangle = a_i$  extracts the last element of the sequence (which corresponds to the newest choice by the spoiler in the one-way EF game);
- for all  $f: E_k A \to B$ , the co-extension  $f^*: E_k A \to E_k B$  is

 $f^* \langle a_1, \ldots, a_i \rangle = \langle f \langle a_1 \rangle, \ f \langle a_1, a_2 \rangle, \ \ldots, f \langle a_1, \ldots, a_i \rangle \rangle,$ 

which intuitively means that the duplicator can recall their own historical moves on the half-board on B given the spoiler's half-board on A.

A co-Kleisli map  $f: E_k A \to B$  for this comonad is then a function from half-boards of A-elements to responses in B, so f encodes precisely a (not necessarily winning) strategy for the duplicator.

The way to formulate winning strategies is to lift the comonad  $E_k$  on **Set** to a comonad  $\mathbb{E}_k$  on the category  $\mathcal{R}(\sigma)$  of  $\sigma$ -relational structures.

**Definition** (EF Comonad). The comonad  $\mathbb{E}_k$  on  $\mathcal{R}(\sigma)$  is defined as follows:

- The object mapping sends every  $\sigma$ -structure  $\mathcal{A} = \langle A, \{P^A\}_{P \in \sigma} \rangle$  to the  $\sigma$ -structure  $\mathbb{E}_k \mathcal{A} = \langle E_k A, \{P^{E_k A}\}_{P \in \sigma} \rangle$  whose underlying set is just  $E_k A$ , i.e. non-empty A-sequences of length at most k; for every relation symbol  $P \in \sigma$  of arity n, its interpretation  $P^{E_k A} \subseteq (E_k A)^n$  relates sequences  $\langle s_1, \ldots, s_n \rangle$  satisfying the following two conditions:
  - 1. for all  $1 \leq i, j \leq n$ , the sequence  $s_i$  is a prefix of  $s_j$  or  $s_j$  is a prefix of  $s_i$ , and 2.  $\langle \epsilon_A(s_1), \ldots, \epsilon_A(s_n) \rangle \in P^A$ .
- The counit  $\epsilon_{\mathcal{A}} : \mathbb{E}_k \mathcal{A} \to \mathcal{A}$  and the co-extension  $f^* : \mathbb{E}_k \mathcal{A} \to \mathbb{E}_k \mathcal{B}$  are the same as those of the comonad  $E_k : \mathbf{Set} \to \mathbf{Set}$ . It can be checked that they are valid morphisms in the category  $\mathcal{R}(\sigma)$ .

Co-Kleisli morphisms  $\mathbb{E}_k \mathcal{A} \to \mathcal{B}$  are exactly winning strategies for the duplicator in the one-way EF game frome  $\mathcal{A}$  to  $\mathcal{B}$  (modulo one subtle problem that we will talk about later). Let's gain some intuition by looking at a small example.

Let us have two  $\sigma$ -structures,  $\mathcal{A}$  and  $\mathcal{B}$ , with underlying sets  $A = \{a, b, c\}$  and  $B = \{a', b', c', d'\}$  respectively, and let us play a 3-round one-way EF Game. We will try to associate the objects of the comonad as we go:

• The spoiler chooses  $a \in A$ . In response, the duplicator chooses  $a' \in B$ :

$$\epsilon_A \langle a \rangle = a, \quad f \langle a \rangle = a', \quad f^* \langle a \rangle = \langle a' \rangle.$$

• The spoiler chooses  $b \in A$ . In response, the duplicator chooses  $b' \in B$ :

$$\epsilon_A \langle a, b \rangle = b, \quad f \langle a, b \rangle = b', \quad f^* \langle a, b \rangle = \langle a', b' \rangle.$$

• The spoiler chooses  $c \in A$ . In response, the duplicator chooses  $c' \in B$ :

$$\epsilon_A \langle a, b, c \rangle = c, \quad f \langle a, b, c \rangle = c', \quad f^* \langle a, b, c \rangle = \langle a', b', c' \rangle.$$

For intuition on how the EF comonad  $\mathbb{E}_k$  acts on relations, let us look at the binary relation case. The general case has the same intuition, but with more terms.

Condition (1) of the above definition imposes that one sequence be a prefix of the other. In game terms, this amounts to that  $s_1$  and  $s_2$  can be stages of the same game play: either  $s_1$  evolves to  $s_2$  or  $s_2$  evolves to  $s_1$ .

Condition (2) says that two sequences  $s_1$  and  $s_2$  are related iff their last elements are related. Since under Condition (1),  $s_1$  and  $s_2$  are two stages of the same play, so  $s_1$  and  $s_2$  being related mean precisely that two elements in a game are related.

Let us go back to our earlier example with our  $\sigma$ -structures  $A = \{a, b, c\}$  and  $B = \{a', b', c', d'\}$ , and suppose there is binary relation symbol  $\leq$  in  $\sigma$  whose interpretation in A and B are the alphabetical order  $\leq$ . Then, we can say that  $\langle a, b \rangle \leq^{E_k A} \langle a, b, c \rangle$ , because on one hand both sequences start in the same way (this game has started with the spoiler playing a and then b), and on the other hand  $b \leq c$  in the alphabetical order.

Now the intuition behind the comonad  $\mathbb{E}_k$  might be clearer: in the  $\sigma$ -relational structure  $\mathbb{E}_k \mathcal{A}$ , some sequences  $s_1, \ldots, s_n$  are related means precisely that there is a game play for which  $s_1, \ldots, s_n$  are the

spoiler's half-boards on  $\mathcal{A}$  in certain rounds, and the spoiler's choices  $\epsilon(s_1), \ldots, \epsilon(s_n)$  are related by some relation in  $\mathcal{A}$ . Therefore, a  $\sigma$ -structure homomorphism  $\mathbb{E}_k \mathcal{A} \to \mathcal{B}$  is a winning strategy for the duplicator.

There is one problem though: we have lifted all relations  $P \in \sigma$  on  $\mathcal{A}$  to  $\mathbb{E}_k \mathcal{A}$ , but there is a special built-in relation in FOL—the equality x = y. The one-way EF game asks that after every round, the chosen  $\mathcal{A}$ -elements and the chosen  $\mathcal{B}$ -elements form a partial isomorphism, so if the spoiler chooses the same element  $a \in \mathcal{A}$  in two rounds, the duplicator must respond with the same elements as well. Otherwise it won't be a partial isomorphism. However, this requirement is not captured by co-Kleisli morphisms  $f : \mathbb{E}_k \mathcal{A} \to \mathcal{B}$ , since e.g.  $\langle a \rangle$  and  $\langle a, b, a \rangle$  are different elements of  $\mathbb{E}_k \mathcal{A}$ , f do not need to map them to the same response.

This motivates us to consider what is called *I-morphisms*, which are coKleisli morphisms  $f : \mathbb{E}_k \mathcal{A} \to \mathcal{B}$ such that  $f(s_1) = f(s_2)$  whenever  $s_1$  is a prefix of  $s_2$  and  $\epsilon(s_1) = \epsilon(s_2)$ . (There is an alternative definition of I-morphisms using relative comonads, but the direct definition suffices for our purposes in this post.)

**Theorem.** There is an winning strategy for the duplicator in the k-round one-way EF game from  $\mathcal{A}$  to  $\mathcal{B}$  if and only if there is an I-morphism  $f : \mathbb{E}_k \mathcal{A} \to \mathcal{B}$ .

Since k-round one-way EF games characterise logical refinement of existential-positive FOL, a direct corollary is that two  $\sigma$ -structure  $\mathcal{A}$  and  $\mathcal{B}$  agree on all closed existential-positive FOL formula  $\varphi$  of rank k if and only if there are two I-morphisms  $\mathbb{E}_k \mathcal{A} \to \mathcal{B}$  and  $\mathbb{E}_k \mathcal{B} \to \mathcal{A}$ .

# **Back-and-Forth Games**

If we ask to have two coKleisli morphisms but nothing else of them, we get the existential-positive fragment, but recall our original goal was to characterise FOL and EF games. What if we ask for the morphisms to have a relationship between them? What if, as we usually do, we ask them to be inverses of each other?

It turns out to characterise the logical equivalence of FOL augmented with *counting quantifiers*  $\exists_{\leq n} x.\varphi$ and  $\exists_{\geq n} x.\varphi$ , whose semantics is that there exist at most n, and respectively at least n, elements  $a \in \mathcal{A}$ making  $\varphi$  true.

**Proposition**. Two finite  $\sigma$ -structures  $\mathcal{A}$  and  $\mathcal{B}$  agree on all closed formulas on FOL with counting quantifiers if and only if there is a pair of *I*-morphisms  $f : \mathbb{E}_k \mathcal{A} \to \mathcal{B}$  and  $g : \mathbb{E}_k \mathcal{B} \to \mathcal{A}$  that are mutual inverses (in the co-Kleisli category of  $\mathbb{E}_k$ ).

So if we don't ask anything from the homomorphisms we get too little; but if we ask for them to be an isomorphism we go overboard and get too much. Can we find a middle ground?

Yes, the key intuition for this is that the duplicator can play the back-and-forth game like a one-way game when the spoiler keeps choosing elements from one structure, but the duplicator must have a plan B in mind in case the spoiler switches to choosing elements from the other structure in the next round.

This motivates what is called a *locally invertible pair*. Given two  $\sigma$ -structure  $\mathcal{A}$  and  $\mathcal{B}$  and a natural number k as usual, a *locally invertible pair*  $\langle F, G \rangle$  consists of a set F of co-Kleisli I-morphisms  $\mathbb{E}_k \mathcal{A} \to \mathcal{B}$  and a set G of co-Kleisli I-morphisms  $\mathbb{E}_k \mathcal{B} \to \mathcal{A}$  such that

1. for all  $f \in F$  and  $s \in \mathbb{E}_k \mathcal{A}$ , there is some  $g_{f,s} \in G$  with  $g_{f,s}^*(f^*(s)) = s$ , and

2. for all  $g \in G$  and  $t \in \mathbb{E}_k \mathcal{B}$ , there is some  $f_{g,t} \in F$  with  $f_{g,t}^*(g^*(t)) = t$ .

**Theorem.** The duplicator has a winning strategy for the k-round EF game on  $\mathcal{A}$  and  $\mathcal{B}$  if and only if there is a non-empty locally invertible pair  $\langle F, G \rangle$ .

*Proof.* Assuming a non-empty locally invertible pair  $\langle F, G \rangle$ , the duplicator has the following strategy such that the chosen elements  $s_i \in A^i$  and  $t_i \in B^i$  after round *i* satisfy the condition

$$\phi_i = \exists f_i \in F. \exists g_i \in G. (s_i, t_i) = (s_i, f^* s_i) = (g^* t_i, t_i).$$

- 1. In round 1, if the spoiler chooses an element  $a_1 \in A$ , the duplicator can pick an arbitrary  $f \in F$  and respond with  $f \langle a_1 \rangle \in B$ . The condition  $\phi_1$  is witnessed by  $f_1 = f$  and  $g_1 = g_{f,\langle a_1 \rangle}$  given by the definition of locally invertible pairs. If the spoiler chooses an element  $b_1 \in B$ , the argument is the same.
- 2. In round i + 1, if the spoiler chooses an element  $a_{i+1}$ , the duplicator responds with  $b_{i+1} = f_i \langle a_1, \ldots, a_{i+1} \rangle$ . The condition  $\phi_{i+1}$  is then witnessed by  $f_{i+1} = f_i$  and  $g_{i+1} = g_{f_i, \langle a_1, \ldots, a_{i+1} \rangle}$ .

After k-rounds,  $\phi_k$  is true and it implies that  $\langle a_1, \ldots, a_k \rangle$  and  $\langle b_1, \ldots, b_k \rangle$  are a partial isomorphism because  $f_k$  and  $g_k$  are coKleisli morphisms  $\mathbb{E}_k \mathcal{A} \to \mathcal{B}$  and  $\mathbb{E}_k \mathcal{B} \to \mathcal{A}$ .

For the other direction, assuming the duplicator has a winning strategy, let  $\Phi \subseteq A^{\leq k} \times B^{\leq k}$  be the set of all possible game states in each round following the winning strategy. The locally invertible pair is

$$F = \{f : \mathbb{E}_k \mathcal{A} \to \mathcal{B} \mid \forall s \in \mathbb{E}_k \mathcal{A}.(s, f^*s) \in \Psi\},\$$
  
$$G = \{g : \mathbb{E}_k \mathcal{B} \to \mathcal{A} \mid \forall t \in \mathbb{E}_k \mathcal{B}.(g^*t, t) \in \Psi\}.$$

First we argue that for all  $(s,t) \in \Psi$ , (i)  $\exists f_{s,t} \in F.f_{s,t}^*s = t$  and (ii)  $\exists g_{s,t} \in G.g_{s,t}^*t = s$ . To show (i) (and symmetrically for (ii)) we construct  $f_{s,t}$  as follows: for every  $s' \in \mathbb{E}_k \mathcal{A}$ , let m be the length of the longest common prefix of s and s'. We consider the game play where in the first m-rounds the spoiler and the duplicator play in the way as  $(s,t) \in \Phi$ , and afterwards, the spoiler always chooses elements according to s' and the duplicator follows the winning strategy. The last  $\mathcal{B}$ -element picked in this game play is the value of the function  $f_{s,t}$  at s'.

Now we can see  $\langle F, G \rangle$  defined as above is a locally invertible pair, for every  $g \in G$  and  $t \in \mathbb{E}_k \mathcal{B}$ ,  $(g^*t, t)$  is in  $\Phi$ , and thus by (i) there exists  $f_{g^*t,t} \in F$  with  $f_{g^*t,t}^*(g^*t) = t$ . The symmetric condition for F similarly holds.

# Wrap-up

In Part I of the post, we have seen how logical equivalences for first-order logic can be characterised by combinatorial games, and how this can be used for showing inexpressivity results of first-order logic. In Part II of this post, we have seen how such games can be formulated in a concise way using comonads.

As an active research subject, what we didn't say about game comonads in this blog post is a lot:

- 1. Many other logics have model comparison games and have received a comonadic treatment, including modal logics (Abramsky and Shah 2021), the *k*-variable fragment of FOL (Abramsky, Dawar and Wang 2017), guarded logics (Abramsky and Marsden 2021), monadic second-order logic (Jakl, Marsden and Shah 2022), finite-variable logics with generalised quantifiers (Conghaile and Dawar 2021).
- 2. Coalgebras of game comonads usually reveal interesting information about the combinatorial structure of finite structures. For example,  $\mathbb{E}_k$ -coalgebras  $\mathcal{A} \to \mathbb{E}_k \mathcal{A}$  are in bijection with forest covers of height  $\leq k$  for the Gaifman graph of  $\mathcal{A}$ .
- 3. Back-and-forth games can be defined in an axiomatic way (Abramsky and Reggio 2021).

Moreover, we have only considered the classical semantics of FOL in sets, so a natural question is how finite model theory interacts with the various notions of finiteness in constructive mathematics and the general categorical semantics of FOL in hyperdoctrines.