Non-Definability in First Order Logic and Monadic Second Order Logic

Ehrenfeucht-Fraïssé Theorem
Hintikka Formulas

Ehrenfeucht-Fraïssé Theorem, I

Theorem:(Easy part)
Assume there is a $MSOL(\tau)$-sentence ϕ with k variables and quantifier depth n in Prenex Normal Form such that $\mathcal{A}_0 \models \phi$ and $\mathcal{A}_1 \models \neg\phi$.

Then I has a winning strategy for the k-pebble n-moves game on $\mathcal{A}_0$ and $\mathcal{A}_1$.

Ehrenfeucht-Fraïssé Theorem, II

We first assume that there infinitely many pebbles.

We write ϕ and $\neg\phi$ in Prenex Normal Form:

$$\phi = \exists X_1 \exists x_2 \forall X_3 \exists x_4 \ldots \exists x_{n-1} \exists X_n$$
$$B(X_1, x_2, \ldots, x_{n-1}, X_n)$$
$$\neg\phi = \forall X_1 \forall x_2 \exists X_3 \forall x_4 \ldots \forall x_{n-1} \forall X_n$$
$$\neg B(X_1, x_2, \ldots, x_{n-1}, X_n)$$

where B is without quantifiers.

We can read from the quantifier prefix a winning strategy.

Ehrenfeucht-Fraïssé Theorem, III

Assume $\mathcal{A}_0 \models \phi$ and $\mathcal{A}_1 \models \neg\phi$.
Player I follows the existential quantifiers.

Player I picks in $\mathcal{A}_0$ a set A_1 such that

$$\mathcal{A}_0, A_1^0 \models \exists x_2 \forall X_3 \exists x_4 \ldots \exists x_{n-1} \exists X_n$$
$$B(A_1^0, x_2, \ldots, x_{n-1}, X_n)$$

Whatever player II picks as A_1^1

$$\mathcal{A}_1, A_1^1 \models \forall x_2 \exists X_3 \forall x_4 \ldots \forall x_{n-1} \forall X_n$$
$$\neg B(A_1^1, x_2, \ldots, x_{n-1}, X_n)$$

Next player I picks an element a_2^0 in $\mathcal{A}_0$ such that

$$\mathcal{A}_0, A_1^0, a_2^0 \models \forall X_3 \exists x_4 \ldots \exists x_{n-1} \exists X_n$$
$$B(A_1^0, a_2^0, \ldots, x_{n-1}, X_n)$$

Whatever player II picks as a_2^1

$$\mathcal{A}_1, A_1^1, a_2^1 \models \exists X_3 \forall x_4 \ldots \forall x_{n-1} \forall X_n$$
$$\neg B(A_1^1, a_2^1, \ldots, x_{n-1}, X_n)$$

Now player I picks in $\mathcal{A}_1$ a set A_3^1 such that

$$\mathcal{A}_1, A_1^1, a_2^1, A_3^1 \models \forall x_4 \ldots \forall x_{n-1} \forall X_n$$
$$\neg B(A_1^1, a_2^1, A_3^1, \ldots, x_{n-1}, X_n)$$

and so on..........

Ehrenfeucht-Fraïssé Theorem, IV

Finally the outcome is from $\mathcal{A}_0$

$$A_1^0, a_2^0, A_3^0, \ldots, a_{n-1}^0, A_n^0$$

and from $\mathcal{A}_1$

$$A_1^1, a_2^1, A_3^1, \ldots, a_{n-1}^1, A_n^1$$

such that

$$\mathcal{A}_0 \models B(A_1^0, a_2^0, A_3^0, \ldots, a_{n-1}^0, A_n^0)$$

and

$$\mathcal{A}_1 \models \neg B(A_1^1, a_2^1, A_3^1, \ldots, a_{n-1}^1, A_n^1)$$

which shows that player I wins, as this cannot be a local isomorphism
(Lemma on local isomorphisms and quantifierfree formulas)

How many non-equivalent formulas? FOL atomic case

Assume we have (first order) variables

$$x_1, x_2, \ldots, x_v$$

This gives $\binom{v}{2} + \binom{v}{1} = O(v^2)$ many instances of $x_i = x_j$ with $i \leq j$.

For a r-ary relation symbol R we get r^v many instances of $R(x_{j_1}, x_{j_2}, \ldots, x_{j_r})$.

If we allow $c_1, c_2, \ldots, c_{v'}$ constants the numbers become $O((v+v')^2)$ and $r^{v+v'}$ respectively.

Proposition:
For a fixed finite relational vocabulary τ with constants and v first order variables, there are a finite number of atomic formulas $\alpha_{\tau,v}^{FOL}$.

How many non-equivalent formulas? $MSOL$ atomic case

Assume we have first and second order variables

$$x_1, x_2, \ldots, x_{v_1}, U_1, U_2, \ldots, U_{v_2}$$

This gives
$O(v_1^2)$ many instances of $x_i = x_j$ with $i \leq j$
and $v_1 \cdot v_2$ many instances of $x_i \in U_j$.

For a r-ary relation symbol R we get r^v many instances of $R(x_{j_1}, x_{j_2}, \ldots, x_{j_r})$.
If we allow $c_1, c_2, \ldots, c_{v_3}$ constants the numbers become $\binom{v_1+v_3}{2}$, $(v_1+v_3)v_2$ and $r^{v_1+v_3}$ respectively.

Proposition:
For a fixed finite relational vocabulary τ with constants and v first order variables, there are a finite number of atomic formulas $\alpha_{\tau,v}^{MSOL}$.

How many non-equivalent formulas? Quantifierfree case

For quantifierfree formulas we only count formulas in CNF.

There are $2^{\alpha_{\tau,v}^{FOL}}$, resp. $2^{\alpha_{\tau,v}^{MSOL}}$ many disjunctions

$$\bigvee_{j=1}^{2^{\alpha_{\tau,v}^{FOL}}} (\neg)^{\nu(j)} A_j$$

where A_j ranges over atomic formulas.

Hence we have (at most) $2^{2^{\alpha_{\tau,v}^{FOL}}}$ many formulas in CNF.

Proposition:
For a fixed finite relational vocabulary τ with constants and v first order variables, there are a finite number of atomic formulas $\beta_{\tau,v}^{FOL}$ and $\beta_{\tau,v}^{MSOL}$, respectively.

How many non-equivalent formulas? Quantifiers I: PNF

Counting quantified formulas is a bit more tricky. We can assume that the formulas are in

Prenex Normal Form

But then variables are NOT reused.

So for each CNF formula with v variables there are $3^v \cdot v!$ many quantifier prefixes ($\exists, \forall$, not quantified).

This gives at most

$$3^v \cdot v! \cdot \beta\tau, v^{FOL}$$

many prenex normal form formulas.

How many formulas are there ? Quantifiers II: quantifier rank

Theorem:
For each τ and $v = v_1 + v_2$ many variables

$$x_1, x_2, \ldots, x_{v_1}, U_1, U_2, \ldots, U_{v_2}$$

there are only $\gamma^{MSOL}_{\tau,v,q}$ many formulas of quantifier rank q.

Proof: We estimate this number by induction over q for $MSOL$.

For $q = 0$ we have at most γ many formulas with $\gamma_0 = \beta\tau, v^{MSOL}$.

Treating them as atomic formulas we have $2v$ many ways of adding one quantifier, and hence at most

$$\gamma^{MSOL}_{\tau,v,q+1} = \gamma_{q+1} = 2^{2^{2v \cdot \eta q}}$$

many formulas of rank $q + 1$.

How many formulas are there ? Quantifiers II: quantifier rank

How many non-equivalent formulas are there really?

Exact estimates to the best of our knowledge unknown.

Hintikka formulas, I

τ is a finite, relational vocabulary.

We denote by $Fm^{MSOL}_{k,q}(\tau)$ the set of $MSOL(\tau)$ formulas such that the variables are among

$$x_1, \ldots, x_k, U_1, \ldots, U_k$$

and each formula has quantifier rank atmost q.

Similarly with $Fm^{FOL}_{k,q}(\tau)$.

Definition:
ϕ and ψ are (finitely) equivalent if the have the same (finite) models.
Free variables are uninterpreted constants

Note: There are, up to logical equivalence infinitely many formulas in three variables (use repeated quantification).

The boolean algebra $Fm_{k,q}(\tau)$, I

Proposition:
There are, up to (finite) equivalence, only finitely many formulas in $Fm_{k,q}(\tau)$.

If ϕ and ψ have only infinite models, they are finitely equivalent (**false**).
There are fewer formulas for finite equivalence.
The number of equivalence classes is growing very fast.

Proposition:
$Fm_{k,q}(\tau)$ is closed under conjunction $\wedge$, disjunction $\vee$ and negation $\neg$,
i.e. it forms a finite *boolean algebra*.

The boolean algebra $Fm_{k,q}(\tau)$, II

The formula $\exists x(x \neq x)$ is the *minimal element*.

The formula $\exists x(x = x)$ is the *maximal element*.

A formula ϕ is an *atom*, if

- it is not (finitely) equivalent to $\exists x(x \neq x)$,
- but for each ψ either $\phi \wedge \psi$ is equivalent to ϕ or to $\exists x(x \neq x)$.

Hintikka formulas, II

We denote by $\mathcal{B}_{k,q}(\tau)$ and $\mathcal{B}^f{}_{k,q}(\tau)$ the finite boolean algebra of $Fm_{k,q}^{MSOL}(\tau)$ up to equivalence and finite equivalence, resp.
The elements are denoted by $\bar{\phi}$.

The set of atoms in $\mathcal{B}_{k,q}(\tau)$ and $\mathcal{B}^f{}_{k,q}(\tau)$ is denoted by $\mathcal{H}_{k,q}(\tau)$ and $\mathcal{H}^f{}_{k,q}(\tau)$.

The formulas ϕ with $\bar{\phi} \in \mathcal{H}_{k,q}(\tau)$ ($\bar{\phi} \in \mathcal{H}^f{}_{k,q}(\tau)$) are called *Hintikka formulas*.

Hintikkika formulas, III

Proposition:

1. Every sentence $\phi \in Fm_{k,q}(\tau)$ is equivalent to the disjunction of a unique set of (k,q)- Hintikka sentences $\bigvee_i h_{i(\phi)}$, with $\bar{h}_{i(\phi)} \in \mathcal{H}_{k,q}(\tau)$.

   Not computable from k, q, τ and ϕ alone.

2. For every k, q, τ and τ-structure $\mathcal{A}$ there is a unique Hintikka sentence $h_{k,q}(\mathcal{A}) \in Fm_{k,q}(\tau)$ such that $\mathcal{A} \models h_{k,q}(\mathcal{A})$.

3. Furthermore, if $\mathcal{A}$ is finite, $h_{k,q}(\mathcal{A})$ is computable from k, q, τ and $\mathcal{A}$.

   But only highly ineffective algorithms are known.

Hintikka formulas, IV

Theorem:(Ehrenfeucht-Fraïssé)

For two τ-structures $\mathcal{A}_1$ and $\mathcal{A}_2$ the following are equivalent:

1. II has a winning strategy in the game with n moves and k point pebbles and k set pebbles.
2. $\mathcal{A}_1$ and $\mathcal{A}_2$ satisfy the same sentences of $Fm_{k,m}(\tau)$.
3. $\mathcal{A}_1$ and $\mathcal{A}_2$ satisfy the same unique (up to equivalence) (k, m)-Hintikka sentence.

We have shown already (1) $\Rightarrow$ (3).
(2) $\Rightarrow$ (3) is trivial.
(3) $\Rightarrow$ (2) follows from the properties of Hintikka formulas.
We are left with (3) $\Rightarrow$ (1).

Constructing the Hintikka sentence, I

Assume we have more pebbles than moves.

Let $\mathcal{A}$ be a finite τ-structure and $a_1, a_2, \ldots, a_s$ elements $\mathcal{A}$.

We define a formula $\phi(v_1, \ldots, v_s)^m_{\bar{a}}$

such that

$$\mathcal{A}, \bar{a} \models \phi(v_1, \ldots, v_s)^m_{\bar{a}}$$

and whenever

$$\mathcal{B}, \bar{b} \models \phi(v_1, \ldots, v_s)^m_{\bar{a}}$$

then player II has a winning strategy in the game for FOL for m more moves starting with $\mathcal{A}, \bar{a}$ and $\mathcal{B}, \bar{b}$.

$\phi(v_1, \ldots, v_k)^q_{\bar{a}}$ (i.e. $k = s, q = m$) will be a Hintikka formula for $Fm^{FOL}_{k,q}(\tau)$.

Constructing the Hintikka sentence, II

$$\phi(v_1, \ldots, v_k)^0_{\bar{a}} :=$$

$$\left(\bigwedge\{R(v_{j_1}, \ldots, v_{j_s}) : R \in \tau, \mathcal{A}, \bar{a} \models R(v_{j_1}, \ldots, v_{j_s})\}\right)$$
$$\wedge$$
$$\left(\bigwedge\{\neg R(v_{j_1}, \ldots, v_{j_s}) : R \in \tau, \mathcal{A}, \bar{a} \models \neg R(v_{j_1}, \ldots, v_{j_s})\}\right)$$
$$\wedge$$
$$\left(\bigwedge\{v_{j_1} = v_{j_2} : j_1, j_2 \leq s \text{ and } \mathcal{A}, \bar{a} \models v_{j_1} = v_{j_2}\}\right)$$
$$\wedge$$
$$\left(\bigwedge\{v_{j_1} \neq v_{j_2} : j_1, j_2 \leq s \text{ and } \mathcal{A}, \bar{a} \models v_{j_1} \neq v_{j_2}\}\right)$$

The formula is finite, provided τ is.

We look at the example of a linear order with $s = 3$ and $m = 2$.

Assume $a_2 < a_1 = a_3$ in $\mathcal{A}$.

Compute the formula.

Constructing the Hintikka sentence, III

$$\phi(v_1, \ldots, v_k)^m_{\bar{a}} :=$$
$$\left(\bigwedge_{a \in A} \exists v_{s+1} \phi(\bar{v}, v_{s+1})^{m-1}_{\bar{a} \cdot a}\right) \wedge$$
$$\left(\forall v_{s+1} \bigvee_{a \in A} \phi(\bar{v}, v_{s+1})^{m-1}_{\bar{a} \cdot a}\right)$$

This is finite by the previous theorem.

We look at the example of a linear order with $s = 3$ and $m = 2$.

Assume $a_2 < a_1 = a_3$ in $\mathcal{A}$.

Compute the formula.

Constructing the Hintikka sentence, IV

We have to verify:

- $\mathcal{A}, \bar{a} \models \phi(v_1, \ldots, v_s)^m_{\bar{a}}$
- whenever $\mathcal{B}, \bar{b} \models \phi(v_1, \ldots, v_s)^m_{\bar{a}}$ then player II has a winning strategy in the game for FOL for m more moves starting with $\mathcal{A}, \bar{a}$ and $\mathcal{B}, \bar{b}$.

Constructing the Hintikka sentence, V

- We can do *"the same"* for $MSOL$ and even for SOL^n or SOL.
- How do we have to modify the construction of there are fewer pebbles than moves?
- What happens if play infintely long?

We shall return to these questions later.

Dense linear orders, I

We look at linear orders such that between any two distinct elements there is a third element. These are called *dense linear orders*.

Exercise:
Express this in FOL.
Show that such an order is always infinite.

There are variations:

- with/without first element.
- with/without last element.

Dense linear orders, II

Examples are

- The real numbers $\mathbb{R}$, which are uncountably infinite.
- The irrational numbers $\mathbb{I} \subseteq \mathbb{R}$, which are also uncountably infinite.
- The rationals $\mathbb{Q}$, which are countably infinite.
- The open intervals $(a, b) \subseteq \mathbb{R}$.
- The open intervals $(a, b) \subseteq \mathbb{Q}$.
- The corresponding closed intervals $[a, b]$ and the intervals $(a, b]$ and $[a, b)$.

Dense linear orders, III

There is a sentence $\phi_{cut} MSOL(\tau_{ord})$ which is true in $\mathbb{Q}$ but not in $\mathbb{R}$.

ϕ_{cut} says:

"The universe is the disjoint union of two open intervals"

Exercise:
Write down this formula.

In $\mathbb{Q}$ we take $(-\infty, \sqrt{2}) \cup (\sqrt{2}, \infty)$.

In $\mathbb{R}$ every Cauchy sequence converges, hence such a decomposition is not possible.

Dense linear orders, IV

Theorem:(Cantor ca. 1870)
Let $\mathcal{A}_0$ and $\mathcal{A}_1$ be two dense linear orders with the same configuration of first and last elements.

Then player II has one (extendible) winning strategy WS in the FOL game for games of arbitrary finite length.

Note that this is **stronger** than the statement:
For every game length n player II has a winningstrategy WS_n.

Corollary:
No $FOL(\tau_{ord})$ sentence ϕ can distinguish
$\mathbb{Q}$ from $\mathbb{R}$, or
$(a, b] \cap \mathbb{Q}$ from $(a, b] \cap \mathbb{R}$ for $a, b \in \mathbb{Q}$,
etc...

Dense linear orders, V

Proof: (No first, no last element)

Assume we have played

$a^0_{k_1} \leq a^0_{k_2} \leq \ldots \leq a^0_{k_m}$ and $a^1_{k_1} \leq a^1_{k_2} \leq \ldots \leq a^1_{k_m}$

and player I chooses, w.l.o.g., $a^0_{m+1} = b$.

There are three cases

- $b < a^0_{k_1}$ or $a^0_{k_m} < b$.
- $b = a^0_{k_j}$ for some $j \leq m$.
- $a^0_{k_{j-1}} < b < a^0_{k_j}$ for some $j \leq m$.

In each case II can reply correspondingly.
In the last case we use density.
In the first case we use the absence of first/last elements.

Exercise:
Complete the proof also for the cases with first/last elements.