
Computability and Definability, Part 3

J.A. Makowsky

Department of Computer Science
Technion - Israel Institute of Technology
Haifa, Israel

janos@cs.technion.ac.il
<http://cs.technion.ac.il/~janos>

Course given the Technion, Haifa, Israel
Winter Semester, 2021/2

<http://www.cs.technion.ac.il/~janos/COURSES/236331-21>

- Location: Taub 8
- Time: Thursday 9:30-12:30 (including Tirgul)
- Office: Taub 628
- Office hours: Thursday 14:00 or by appointment

ZOOM

- makowsky@technion.ac.il is inviting you to a scheduled Zoom meeting.
- Topic: CS 236331
- Time: This is a recurring meeting: Thursday 9:30-12:30
- Join Zoom Meeting at:

<https://technion.zoom.us/j/91510301938>

Course requirements

- Prerequisites are NOT enforced.
Recommended prerequisites: Some familiarity with the concepts of Logic and sets, Automata and formal languages or Computability
- Homework
- Final project

Session 5 (November 25, 2021)

Overview

Part 3: Ehrenfeucht Fraïssé Games and Proving Non-definability

- Ehrenfeucht Fraïssé Games
- Winning strategies for EF-games
- Dense linear orders
- Finite automata and the BET-Theorem
Disjoint unions and concatenation, Finite automata, The BET-Theorem
- Non-definability in MSOL
Pumping Lemmas, Non-definability,

[Back to overview](#)

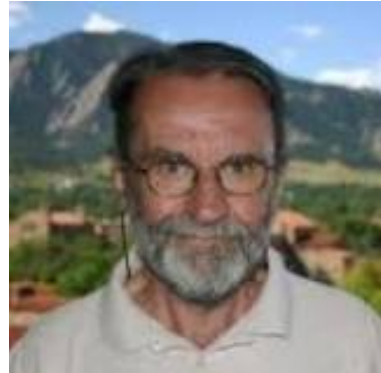
Part 3.1

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

Ehrenfeucht-Fraïssé Games and Hintikka formulas



R. Fraïssé
1920-2008



A. Ehrenfeucht
1932-



J. Hintikka
1929-2015

Their work is from the 1950ties

Tools to Show Non-Definability

- Compactness of First Order Logic
- Ehrenfeucht-Fraïssé Games
- Translation Schemes and transductions
- Feferman-Vaught Theorem for sums
- 0 – 1 Laws

Proving non-definability

The class of τ -structures of finite even cardinality, $EVEN(\tau)$, is *not definable* in First Order Logic, (not even in Monadic Second Order Logic):

- For FOL : use compactness. Every formula true in all finite even structures has an infinite model.
- For FOL (restricted to finite structures): use Pebble Games (Ehrenfeucht-Fraïssé Games)
- For $MSOL$: use Pebble Games adapted to $MSOL$.

Similarly, $DisPath(n)$ is not FOL -definable even for $n = 1$.

Compactness of FOL

Recall:

Σ is **satisfiable** if there is a τ -structure $\mathcal{A}$ such that $\mathcal{A} \models \Sigma$.

Theorem:[Gödel-Mal'cev]

Let Σ be an infinite set of $FOL(\tau)$ -sentences.

Σ is satisfiable iff every finite subset $\Sigma_0 \subseteq \Sigma$ is satisfiable.

*This theorem was stated and proved in **Logic for CS** for Propositional Logic.*

*This theorem was stated, but probably not proved in **Logic for CS** for First Order Logic.*

The proof for FOL is very similar to the one for Propositional Logic.

Using Compactness

Let ϕ_n be the sentence which says that the universe contains at least n elements.

Let Σ_{even} consist of

$$\{(\phi_{2n+1} \rightarrow \phi_{2n+2}) : n \in \mathbb{N}\}$$

All **finite** models of Σ_{even} are of even cardinality.

Assume there is ψ_{even} such that

$$\mathcal{A} \models \psi_{\text{even}} \text{ iff } |A| = 2n$$

Define

$$\Sigma_1 = \{\psi_{\text{even}}\} \cup \{\phi_n : n \in \mathbb{N}\}$$

Every finite subset $\Sigma_0 \subseteq \Sigma_1$ is satisfiable (by a finite model of even cardinality).

But Σ_1 has no model, **contradicting compactness**.

MSOL is not compact

Let $\tau_{a,b} = \tau_{graph} \cup \{a, b\}$ be the vocabulary of graphs with two constants.

In $MSOL(\tau_{a,b})$ we have a formula ϕ_{conn} which says that the graph is connected.

Let $\psi_n(a, b)$ say that the shortest path between a, b is of length n .

This is in $FOL(\tau_{a,b})$.

Now every finite subset of

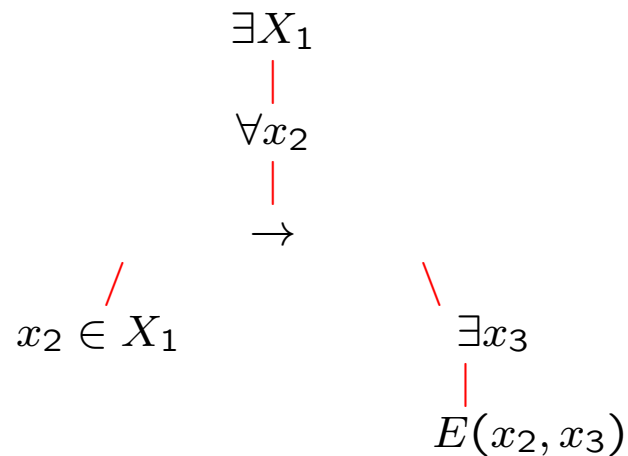
$$\Sigma = \{\phi_{conn} \cup \{\psi_n(a, b) : n \in \mathbb{N}\}$$

is satisfiable, but Σ is not.

Quantifier rank of a formula, I

We write a formula ϕ as a tree:

$$\exists X_1 \forall x_2 (x_2 \in X_1 \rightarrow \exists x_3 E(x_2, x_3))$$



The quantifier rank is biggest number of quantifiers one can find along a path in this tree.

Here it is 3.

Quantifier rank of a formula, II

- For formulas in **prenex normal form** the quantifier rank **equals** the **number of quantifiers**.
- If we **reuse variables**, the quantifier rank can be **smaller** than the number of quantifiers used in prenex normal form.

$$\forall x_1 (\exists x_2 E(x_1, x_2) \wedge \exists x_2 \neg E(x_1, x_2))$$

Quantifier rank **2**

$$\forall x_1 \exists x_2 \exists x_3 (E(x_1, x_2) \wedge \neg E(x_1, x_3))$$

Quantifier rank **3**

Ehrenfeucht-Fraïssé Games, I

Given two τ -structures $\mathcal{A}_0$ and $\mathcal{A}_1$ and their powersets $P(A_0)$ and $P(A_1)$.

Two players I (spoiler), II (duplicator)

k numbered pebbles for each structure

Two kind of moves: Set- and point-moves

Play for n moves

i -th move:

I chooses $\alpha \in \{0, 1\}$ and put pebble on an element in $P(A_\alpha)$ (Set-move) or in A_α (point move).

II puts corresponding pebble on set or point.

Ehrenfeucht-Fraïssé Games, II

After n moves we have from $\mathcal{A}_0$

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

and from $\mathcal{A}_1$

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

These two sequences are *locally isomorphic* if for all j, k

$$a_k^0 \in A_j^0 \text{ iff } a_k^1 \in A_j^1$$

and for each m -ary $R \in \tau$ and $j_1, j_2, \dots, j_m$

$$R^{\mathcal{A}_0}(a_{j_1}^0, a_{j_2}^0, \dots, a_{j_m}^0) \text{ iff } R^{\mathcal{A}_1}(a_{j_1}^1, a_{j_2}^1, \dots, a_{j_m}^1)$$

Lemma: Two sequences in $\mathcal{A}_0$ and $\mathcal{A}_1$ respectively

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

and

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

are locally isomorphic iff for all quantifierfree formulas B we have

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

iff

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

Proof:

Use induction over the construction of B .

Ehrenfeucht-Fraïssé Games, III

Winning the game:

II wins if the correspondence on the pebbles induces a local isomorphism (including the sets).

Theorem: (Ehrenfeucht-Fraïssé, 1953/61)

II has a winning strategy for the k -pebble n -moves game on $\mathcal{A}_0$ and $\mathcal{A}_1$ iff they satisfy the same $MSOL(\tau)$ -sentences with k variables and quantifier depth n .

If no set-moves are played this holds for $FOL(\tau)$.

We write $\mathcal{A}_0 \sim_{k,n}^{MSOL} \mathcal{A}_1$ iff

II has a winning strategy in the game with set moves and

$\mathcal{A}_0 \sim_{k,n}^{FOL} \mathcal{A}_1$ in the game without set moves.

Winning strategies, I

A winning **strategy** is a function which takes a position of length n

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

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

from $\mathcal{A}_0$ and $\mathcal{A}_1$ respectively

together with a move of player I, say

$X_{n+1}^i \in \{a_{n+1}^i, A_{n+1}^i\}$ as input and returns

$X_{n+1}^{1-i} \in \{a_{n+1}^{1-i}, A_{n+1}^{1-i}\}$ as output such that

$$A_1^0, a_2^0, A_3^0, \dots, a_{n-1}^0, A_n^0, X_{n+1}^0$$

$$A_1^1, a_2^1, A_3^1, \dots, a_{n-1}^1, A_n^1, X_{n+1}^1$$

is a winning position

(if it exists, else it is undefined).

Winning strategies, II

Proposition:

$$\mathcal{A}_0 \sim_{k,n}^{FOL} \mathcal{A}_1 \text{ and } \mathcal{A}_0 \sim_{k,n}^{MSOL} \mathcal{A}_1$$

are *equivalence relations* between τ -structures.
I.e., they are **symmetric**, **reflexive** and **transitive**.

Proof:

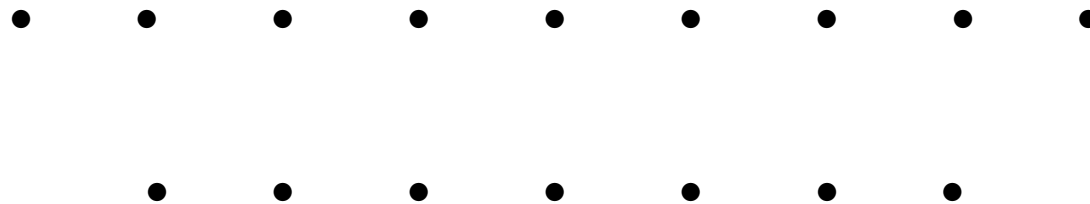
Reflexivity: Copy literally

Symmetry: The structures play exchangeable roles (but not the players)

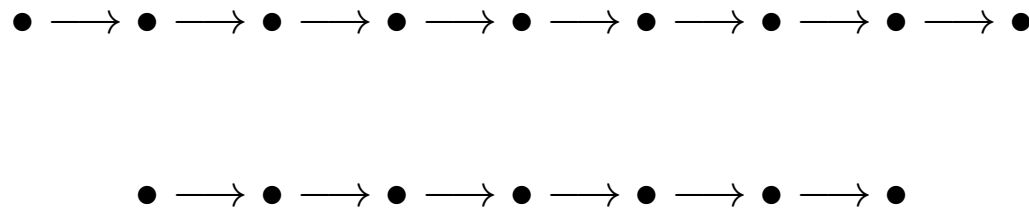
Transitivity: Play on the intermediate board

Winning EF-Games, I

$$\tau = \emptyset$$



$$\tau = \{R_2\}, \text{ linear orders}$$



Winning EF-Games, II

Theorem:

Let $\tau = \emptyset$.

For two sets $\mathcal{A}_0$ and $\mathcal{A}_1$
of size m_0 and m_1 respectively,
we have $\mathcal{A}_0 \sim_{k,n}^{FOL} \mathcal{A}_1$
(in the game without set moves)

iff

$m_0 = m_1$ or
 $k \leq m_0$ and $k \leq m_1$

Winning EF-Games, III

Theorem:

Let $\tau = \{R_2\}$.

For two cycle graphs $\mathcal{G}_0$ and $\mathcal{G}_1$

of size v_0 and v_1 respectively,

we have $\mathcal{G}_0 \sim_{k,n}^{FOL} \mathcal{G}_1$

(in the game without set moves)

provided

$v_0 = v_1$ or

$2^k \leq v_0$ and $2^k \leq v_1$

Does the converse hold ?

Ehrenfeucht-Fraïssé Games, IV

Theorem: (Feferman, Vaught, 1956)

If $\mathcal{A}_0 \sim_{k,n}^{MSOL} \mathcal{B}_0$ and $\mathcal{A}_1 \sim_{k,n}^{MSOL} \mathcal{B}_1$
then $\mathcal{A}_0 \sqcup \mathcal{A}_1 \sim_{k,n}^{MSOL} \mathcal{B}_0 \sqcup \mathcal{B}_1$

Theorem: (Feferman, Vaught, 1956)

If $\mathcal{A}_0 \sim_{k,n}^{FOL} \mathcal{B}_0$ and $\mathcal{A}_1 \sim_{k,n}^{FOL} \mathcal{B}_1$
then $\mathcal{A}_0 \times \mathcal{A}_1 \sim_{k,n}^{FOL} \mathcal{B}_0 \times \mathcal{B}_1$

The same holds for "gluing" operations.

Winning EF-Games, IV

Theorem:

Let $\tau = \{R_2\}$.

Let $\mathcal{G}_0$ consist of one cycle of size 2^k
and $\mathcal{G}_1$ consist of two cycles of size 2^k .

Then we have $\mathcal{G}_0 \sim_{k,n}^{FOL} \mathcal{G}_1$
(in the game without set moves)

Corollary:

Connectivity is not *FOL*-definable in the language of graphs.

Winning EF-Games, V

First we play the game for *MSOL* for $\tau = \emptyset$.

A_0 is a set of 2^n elements

A_1 is a set of $2^n - 1$ elements

How many moves does player I need to win?

C_n is the undirected graph with n vertices which is connected and 2-regular.

$\mathcal{A}_0$ is the graph C_{2^n}

$\mathcal{A}_1$ is the graph $C_{2^{n-1}}$ elements

How many moves does player I need to win?

Winning EF-Games, VI

The rôle of the pebbles.

How long can we play (without set moves) with **two** pebbles?



These are **directed** graphs.

How long can we play with **three** pebbles?

Back to overview

Test your understanding, I

Let τ_{unary}^k be the vocabulary with k unary relation symbols $P_1, \dots, P_k$.

- For $k = 1$, characterize the τ_{unary}^k -structures $\mathfrak{A}, \mathfrak{B}$ such that duplicator has a winning strategy for n -moves picking only elements.
- For $k = 1$, characterize the τ_{unary}^k -structures $\mathfrak{A}, \mathfrak{B}$ such that duplicator has a winning strategy for n -moves picking elements and subsets.
- Same for $k = 2, 3, \dots$

Test your understanding, II

We look at two graphs of the same order $g \geq 2$, the first with directed edges, the second with symmetric (bi-directed) edges.

$$\mathfrak{A} : \bullet \longrightarrow \bullet \longrightarrow \bullet \longrightarrow \bullet \longrightarrow \bullet \longrightarrow \bullet \longrightarrow \bullet \longrightarrow \bullet \longrightarrow \bullet$$

$$\mathfrak{B} : \bullet \longleftrightarrow \bullet \longleftrightarrow \bullet \longleftrightarrow \bullet \longleftrightarrow \bullet \longleftrightarrow \bullet \longleftrightarrow \bullet \longleftrightarrow \bullet$$

We play on $\mathfrak{A}, \mathfrak{B}$ picking only elements.

Show that II (spoiler) wins in games of length $n \geq 2$.

Part 3.2

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 $\mathcal{A}_1$.

Test your understanding, III

Let $\tau_\emptyset$ the empty vocabulary.

Use the easy part of the Theorem to show:

- There is no first order formula $\phi \in \text{FOL}(\emptyset)$ in Prenex Normal Form with n quantifiers which says **there are at least $n + 1$ elements**.
- There is a monotone increasing function $f : \mathbb{N} \rightarrow \mathbb{N}$ for which the following holds:
There is no formula $\phi \in \text{MSOL}(\emptyset)$ in Prenex Normal Form with n (first order and second order) quantifiers which says **there are at least $f(n)$ elements**.
- What can you say about $f(n)$?

Test your understanding, IV

Let $\tau_{graph} = \{E\}$ with one binary relation.

Let $G = C_n$ and $H = C_n \sqcup C_k$, where $\sqcup$ is the disjoint union graphs for τ_{graph} .

Assume $m, k \geq 3$.

- Show that in the game with set moves I (spoiler) has a winning strategy independently of $n, k \geq 3$.
- How many moves does the spoiler need?
- Analyze the same but for $\tau_{hypergraph}$, G and H as hypergraphs.

Ehrenfeucht-Fraïssé Theorem, II

We first assume that there infinitely many pebbles.

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

$$\begin{aligned}\phi &= \exists X_1 \exists x_2 \forall X_3 \exists x_4 \dots \exists x_{n-1} \exists X_n \\ &\quad B(X_1, x_2, \dots, x_{n-1}, X_n) \\ \neg\phi &= \forall X_1 \forall x_2 \exists X_3 \forall x_4 \dots \forall x_{n-1} \forall X_n \\ &\quad \neg B(X_1, x_2, \dots, x_{n-1}, X_n)\end{aligned}$$

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

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

Ehrenfeucht-Fraïssé Theorem, IV

Whatever player II (duplicator) picks as A_1^1

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

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

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

Whatever player II (duplicator) picks as a_2^1

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

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

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

and so on.....

Ehrenfeucht-Fraïssé Theorem, V

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

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

and from $\mathcal{A}_1$

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

such that

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

and

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

which shows that player I (spoiler) wins, as this cannot be a local isomorphism

(We need a Lemma on local isomorphisms and quantifierfree formulas)

How many non-equivalent formulas?

FOL atomic case

Assume we have (first order) variables

$$x_1, x_2, \dots, x_v$$

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 $R(x_{j_1}, x_{j_2}, \dots, x_{j_r})$.

If we allow $c_1, c_2, \dots, c_{v'}$ 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

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

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 $R(x_{j_1}, x_{j_2}, \dots, x_{j_r})$.
If we allow $c_1, c_2, \dots, c_{v_3}$ 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

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

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 non-equivalent quantifier'free 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

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

there are only $\gamma_{\tau, v, q}^{MSOL}$ 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

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

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 $Fm_{k,q}^{MSOL}(\tau)$ the set of $MSOL(\tau)$ formulas such that the variables are among

$$x_1, \dots, x_k, U_1, \dots, U_k$$

and each formula has quantifier rank at most q .

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

Definition:

ϕ and ψ are (finitely) equivalent if they 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 $\mathcal{B}_{k,q}^f(\tau)$ 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 $\mathcal{B}_{k,q}^f(\tau)$ is denoted by $\mathcal{H}_{k,q}(\tau)$ and $\mathcal{H}_{k,q}^f(\tau)$.

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

Hintikkika formulas, III

Proposition:

- (i) 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.

- (ii) 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
 $\mathcal{A} \models h_{k,q}(\mathcal{A})$.
- (iii) 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:

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

We have shown already $(i) \Rightarrow (ii)$.

$(ii) \Rightarrow (iii)$ is trivial.

$(iii) \Rightarrow (i)$ follows from the properties of Hintikka formulas.

We are left with $(iii) \Rightarrow (i)$.

Constructing the Hintikka sentence, I

Assume we have more pebbles than moves.

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

We define a formula $\phi(v_1, \dots, v_s)_{\bar{a}}^m$

such that

$$\mathcal{A}, \bar{a} \models \phi(v_1, \dots, v_s)_{\bar{a}}^m$$

and whenever

$$\mathcal{B}, \bar{b} \models \phi(v_1, \dots, v_s)_{\bar{a}}^m$$

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

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

Constructing the Hintikka sentence, II

$$\begin{aligned}
 \phi(v_1, \dots, v_k)_{\bar{a}}^0 := & \\
 & \left(\bigwedge \{R(v_{j_1}, \dots, v_{j_s}) : R \in \tau, \mathcal{A}, \bar{a} \models R(v_{j_1}, \dots, v_{j_s})\} \right) \\
 & \quad \wedge \\
 & \left(\bigwedge \{\neg R(v_{j_1}, \dots, v_{j_s}) : R \in \tau, \mathcal{A}, \bar{a} \models \neg R(v_{j_1}, \dots, v_{j_s})\} \right) \\
 & \quad \wedge \\
 & \left(\bigwedge \{v_{j_1} = v_{j_2} : j_1, j_2 \leq s \text{ and } \mathcal{A}, \bar{a} \models v_{j_1} = v_{j_2}\} \right) \\
 & \quad \wedge \\
 & \left(\bigwedge \{v_{j_1} \neq v_{j_2} : j_1, j_2 \leq s \text{ and } \mathcal{A}, \bar{a} \models v_{j_1} \neq v_{j_2}\} \right)
 \end{aligned}$$

The formula is finite, provided τ is.

Homework (compulsory)

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

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

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:

- $\mathcal{A}, \bar{a} \models \phi(v_1, \dots, v_s)_{\bar{a}}^m$
- whenever $\mathcal{B}, \bar{b} \models \phi(v_1, \dots, v_s)_{\bar{a}}^m$
then player II (duplicator) 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 if there are fewer pebbles than moves?
- What happens if play infinitely long?

We shall return to these questions later.

Back to overview

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 (duplicator) 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 (duplicator) has a winning strategy 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

$$a_{k_1}^0 \leq a_{k_2}^0 \leq \dots \leq a_{k_m}^0 \text{ and } a_{k_1}^1 \leq a_{k_2}^1 \leq \dots \leq a_{k_m}^1$$

and player I (spoiler) chooses, w.l.o.g., $a_{m+1}^0 = b$.

There are three cases

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

In each case II (duplicator) can reply correspondingly.

In the last case we use density.

In the first case we use the absence of first/last elements.

Dense linear orders, VI

Exercise:

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

[Back to overview](#)

Part 3.3

Disjoint unions and concatenation

Finite Automata

Regular Languages

The Büchi-Elgot-Trakhtenbrot Theorem

Disjoint unions of structures, I

There are several ways of looking at **disjoint unions** of structures.

The most general version generally used might be:

$\mathcal{A}_0$ a τ_0 -structure, $\mathcal{A}_1$ a τ_1 -structure,
 $\sigma = \tau_0 \sqcup \tau_1 \sqcup \{P_0, P_1\}$

$\mathcal{B} = \mathcal{A}_0 \sqcup \mathcal{A}_1$ is the σ -structure with

$B = A_0 \sqcup A_1$, $P_i(\mathcal{B} = A_i$ and
 for $R \in \tau_i$, $R(\mathcal{B}) = R(\mathcal{A}_i)$

Remark: For $\tau_0 = \tau_1 = \tau$ one puts often $R(\mathcal{B}) = R(\mathcal{A}_0) \sqcup R(\mathcal{A}_1)$.

Sometimes the predicates P_1 are omitted.

Only with the definition above are the parts $\mathcal{A}_i$ definable from the disjoint union.

Disjoint unions of structures, II

Theorem:(Feferman, Vaught, Ehrenfeucht)

(i) If $\mathcal{A}_0 \sim_{q,v}^{MSOL} \mathcal{B}_0$ and $\mathcal{A}_1 \sim_{q,v}^{MSOL} \mathcal{B}_1$ so $\mathcal{A}_0 \sqcup \mathcal{A}_1 \sim_{q,v}^{MSOL} \mathcal{B}_0 \sqcup \mathcal{B}_1$.

(ii) If $h_{q,v}(\mathcal{A}_0) = h_{q,v}(\mathcal{B}_0)$ and $h_{q,v}(\mathcal{A}_1) = h_{q,v}(\mathcal{B}_1)$
so also

$$h_{q,v}(\mathcal{A}_0 \sqcup \mathcal{A}_1) = h_{q,v}(\mathcal{B}_0 \sqcup \mathcal{B}_1).$$

In other words:

The (q, v) -Hintikka sentence of a disjoint union is uniquely determined by the (q, v) -Hintikka sentence of its parts.

Concatenation, I

The **concatenation of two words** over an alphabet Σ is a special case of a disjoint union of *ordered structures*, where the second part follows the first.

We denote, for a word $w \in \Sigma^*$ the corresponding structure by $\mathcal{A}_w$.

We denote by $\mathcal{A}_v \bullet \mathcal{A}_w$ the structure corresponding to the word vw .

Concatenation, II

Theorem:(Büchi, Ehrenfeucht)

(i) If $\mathcal{A}_0 \sim_{q,v}^{MSOL} \mathcal{B}_0$ and $\mathcal{A}_1 \sim_{q,v}^{MSOL} \mathcal{B}_1$ so $\mathcal{A}_0 \bullet \mathcal{A}_1 \sim_{q,v}^{MSOL} \mathcal{B}_0 \bullet \mathcal{B}_1$.

(ii) If $h_{q,v}(\mathcal{A}_0) = h_{q,v}(\mathcal{B}_0)$ and $h_{q,v}(\mathcal{A}_1) = h_{q,v}(\mathcal{B}_1)$ so

$$h_{q,v}(\mathcal{A}_0 \bullet \mathcal{A}_1) = h_{q,v}(\mathcal{B}_0 \bullet \mathcal{B}_1) \quad (+)$$

In other words:

The (q, v) -Hintikka sentence of a concatenation is uniquely determined by the (q, v) -Hintikka sentence of its parts.

Finite Automata, I

We have **deterministic** and **non-deterministic** finite automata (Turing machines without work tape).

We **one-directional** and **two-directional** finite automata.

Let

$X \in \{(det, one), (n - det, one), (det, two), (n - det, two)\}$.

A language (set of words) L is $X - FA$, if it is accepted by some X finite automaton.

Theorem:(Rabin and Scott, 1959)

L is $X - FA$ iff L is $Y - FA$

for each $X, Y \in \{(det, one), (n - det, one), (det, two), (n - det, two)\}$.

The proof was given in the course Automata and Formal Languages

Finite Automata, II

We can also look at

- **multi-tape, k -tape** finite automata with one simultaneous head on the tapes.
- **multi-head, k -head** finite automata.
- **k -pebble** finite automata with pebbles (markers) on the tape.

Theorem:

A language L is k -tape $X - FA$ iff L is 1-tape $X - FA$.

But there are **more** languages which are 2-head $X - FA$ than with one head.
The same even with **only one** pebble.

Regular Languages, I

Let Σ be a finite alphabet.

λ denotes the empty word.

Σ^* is the set of all finite words (including λ).

Σ^+ is the set of all non-empty finite words, (excluding λ).

Regular Σ -expression are

- $\emptyset$, and a for each $a \in \Sigma$;
- if r, s are regular expressions, so are $(r \cup s)$, (rs) and r^+ .

Regular Languages, II

For a regular expression r we define a language $Lang(r)$.

Assume $Lang(r) = R$ and $Lang(s) = S$.

- $Lang(\emptyset) = \emptyset$, $Lang(a) = \{a\}$ for $a \in \Sigma$.
- $Lang(r \cup s) = R \cup S$
- $Lang(rs) = \{uv : u \in R, v \in S\} = RS$
- We define $R^1 = R$ and $R^{n+1} = R^n R$, and $R^+ = \bigcup_{1 \leq n} R^n$.
- $Lang(r^+) = R^+$.

A language L is **regular** iff $L = Lang(r)$ for some Σ -regular expression r .

Regular Languages, III

Complementation:

For r we form the expression $\neg r$ with $Lang(\neg r) = \Sigma^+ - Lang(r)$.

Theorem:

For every regular expression r $lang(\neg r)$ is regular.

An expression is *regular plus-free* if it is defined inductively by

- $\emptyset, \{a\}$
- $(r \cup s), (rs), (\neg r)$

A regular language is **plus-free** if it is of the form $Lang(r)$ for some plus-free expression.

Finite Automata, III

Theorem:

(Kleene, 1953, Rabin and Scott 1959)

The following are equivalent for languages L :

- L is regular
- L is $(det, one) - FA$
- L is $(n - det, two) - FA$

and also for

$(det, two) - FA$ and $(n - det, one) - FA$.

The proof was given in the course Automata and Formal Languages

Finite Automata, IV

Theorem:(Büchi-Elgot-Trakhtenbrot)

A set of words L is regular

iff

the set of its structures K_L is definable in $MSOL$

Theorem:(McNaughton)

A set of words L is plus-free regular

iff

the set of its structures K_L is definable in FOL

Proof of Büchi's Theorem, I

Proof:

If L is regular, it can be defined by a regular expression r .

We use r to construct an *MSOL*-formula which defines L .

We use induction.

For $\vee$, concatenation and complement, we use *FOL* operations. For $+$ we quantify over sets of positions and relativize the formulas of the induction hypothesis.

Note that we did not use (r^*) .

We avoid the empty word λ .

How could we include it?

Proof of Büchi's Theorem, II

Now assume that K_L is defined by $\phi \in Fm_{q,v}^{MSOL}(\tau)$.

We define the the automaton for L .

The states are $\mathcal{H}_{q,v}(\tau)$.

The transitions are given by $(+)$ of the previous theorem (Büchi-Ehrenfeucht) with the second word a singleton.

The accepting states are the (q, v) -Hintikka formulas the disjunction of which is equivalent to ϕ .

This works both for FOL and $MSOL$ with the according modifications.

Hintikka sentences
are used to define
a finite automaton !

Part 3.4

Pumping Lemma

Proving non-definability in *MSOL*

Translation schemes

[Back to overview](#)

Pumping Lemma, I

Theorem: Let A be a finite (deterministic, one-directional) finite automaton with n states and defining the language $L(A)$.

Let $w \in L(A)$ with length $\ell(w) \geq n$.
Then there exists words x, y, z such that

- $w = xyz$ and $y \neq \Lambda$ and
- for each $k \in \mathbb{N}$ $xy^kz \in L(A)$

A pumping lemma for **context free** languages was stated first in 1961 by Bar-Hillel, Perles, Shamir.

Pumping Lemma, II

We want to apply the Pumping Lemma to $MSOL$.

Theorem: Let ϕ be a $MSOL(\tau_{words}(\Sigma))$ -sentence over words in Σ^+ with quantifier rank q and v variables and defining the language $L(\phi)$.

Let $\eta_{v,q,\Sigma} \leq \gamma_{v,q,\Sigma}$ be the number of Hintikka sentences in $Fm_{q,v}^{MSOL}(\tau(\Sigma))$.

Let $w \in L(\phi)$ with length $\ell(w) \geq \eta_{q,v,\Sigma}$. Then there exists words x, y, z such that

- $w = xyz$ and $y \neq \Lambda$ and
- for each $k \in \mathbb{N}$ $xy^kz \in L(\phi)$

Pumping Lemma, III

Examples

The following are not regular

- $\{a^i b^i : i \in \mathbb{N}\}, \{a^i b^i c^i : i \in \mathbb{N}\},$
 $\{a^i b^j : i, j \in \mathbb{N}, i \leq j\},$
- The set of prime numbers as binary words.
This follows easily from an easy theorem on primes:

Theorem: For every $n \in \mathbb{N}$ there are successive primes $p_{i(n)}, p_{i(n)+1}$ such that $p_{i(n)+1} - p_{i(n)} \geq n$.

Hint: Look at the sequence $n! + i, i = 2, 3, \dots, n$.

A direct proof is in
Michael Harrison, Introduction to Formal Language Theory, Addison-Wesley 1978, chapter 2.2

Unary languages

Proposition:

A unary language L is regular iff $X = \{i : a^i \in L\}$ is ultimately periodic.

$X \in \mathbb{N}$ (in increasing order) is ultimately periodic iff there is p such that for i large enough $x_{i+p} = x_i$.

[Back to overview](#)

Non-definability in $MSOL_1$, I

$MSOL_1$ is the $MSOL$ for structures which are graphs of the form $G = \langle V, E \rangle$ (E a binary relation).

The following are not $MSOL_1$ -definable.

- HALF-CLIQUE: graphs with a clique of size at least $\frac{|V|}{2}$
- HAM: graphs which have a hamiltonian cycle.
- EULER: graphs which have an Eulerian circuit.

Non-definability in $MSOL_1$, II

Proof for HALF-CLIQUE:

Assume $\phi_{half-clique} \in MSOL_1$
defines HALF-CLIQUE.

For each word $w = a^i b^j, i, j \neq 0$ of length n
we define a graph G_w as follows:

$$V = \{1, \dots, n\}$$

$$E = \{(u, v) \subseteq V^2 : \psi(u, v) = P_b(u) \wedge P_b(v) \wedge u \neq v\}$$

Clearly G_w in HALF-CLIQUE iff
 $w = a^i b^j$ with $i \leq j$.

But then let Φ be the formula we obtain from substituting $E(x, y)$ in ϕ by $\psi(x, y)$.

$w \models \Phi$ iff $w = a^i b^j$ with $i \leq j$.

By Büchi's Theorem, this implies that
 $\{a^i b^j : i \leq j\}$ is regular, a **contradiction**.

Non-definability in $MSOL_1$, III

Proof for HAM:

Assume $\phi_{ham} \in MSOL_1$ defines HAM.

For each word $w = a^i b^j, i, j \neq 0$ of length n we define a graph G_w as follows:

$$V = \{1, \dots, n\}$$
$$E = \{(u, v) \subseteq V^2 : \psi(u, v) = P_a(u) \wedge P_b(v)\}$$

Clearly G_w in HAM iff $w = a^i b^j$ with $i = j$.

But then let Φ be the formula we obtain from substituting $E(x, y)$ in ϕ by $\psi(x, y)$.

$w \models \Phi$ iff $w = a^i b^j$ with $i = j$.

By Büchi's Theorem, this implies that $\{a^i b^i : i \in \mathbb{N}\}$ is regular, a **contradiction**.

Non-definability in $MSOL_1$, IV

Proof for EULER:

A graph is eulerian iff it is connected and all vertices have even degree.
Hence, the complete graph K_n is eulerian iff $n = 2m + 1$.

For each word $w = a^i b^j, i, j \neq 0$ of length n
we define a graph G_w as follows:

$$V = \{1, \dots, n\}$$
$$E = \{(u, v) \subseteq V^2 : \psi(u, v) = u \neq v\}$$

Clearly G_w is EULER iff
 $w = a^i b^j$ with $i + j = 2m + 1$.

Similarly as before, this implies that
 $\{a^i b^j : i + j = 2m + 1\}$ is regular.
But it is regular.

THIS PROOF DOES NOT WORK !

Non-definability in $MSOL_1$, **V**

The proofs for HALF-CLIQUE and HAM actually show more:

Theorem:

HAM and HALF-CLIQUE are not $MSOL$ -definable even on **ordered graphs**.

An ordered graph $G = \langle V, E, < \rangle$ is a graph with a **linear order on the vertices**.

But EULER is $MSOL$ definable on ordered graphs, because on linear orders there is a formula $\phi_{\text{even}}(X)$ which says that $|X|$ is even.

Note also that on unary words

$$\{a^i : i = 2m\}$$

is ultimately periodic and hence regular.

Non-definability in $MSOL_1$, V

Exercise:

To prove that EULER is not $MSOL_1$ -definable

Hint:

Use that sets of even cardinality are not $MSOL$ -definable.

Translation schemes, I

In these proofs we used a technique which we will spell out in full generality:

- For a word $w \in L$ we **defined** a graph G_w
- **defined** by an *MSOL*-formula
actually a *FOL*-formula ψ
- Then we assumed that the class of graphs K was definable by ϕ .
- Put $\Phi = \text{subst}_E(\phi, \psi(x, y))$
- Show that $w \in L$ iff $G_w \in K$
- Conclude that L is defined by Φ .

We shall develop a formalism for

Translation schemes

which will play a central rôle in the sequel of the
course.

[Back to overview](#)