

# The cost of equipping languages with infinite data

Gabriele Puppis

joint works with

Thomas Colcombet  
Clemens Ley  
Michał Skrzypczak

$$L \subseteq \mathbb{A}^*$$

finite alphabet  
(classical languages)      infinite alphabet  
**(data languages)**

Data languages are assumed invariant under bijective renamings (i.e. permutations) in  $\mathbb{A}$



*focus on equalities between letters / values*

E.g.  $L = \{ w \in \mathbb{A}^* : \text{at least 2 distinct values in } w \}$

$$= \{ \bullet\bullet, \bullet\bullet\bullet, \bullet\bullet\bullet, \bullet\bullet\bullet\bullet\bullet\bullet, \dots \}$$

$$= \{ oo, ooo, oooooo, \dots \}$$

Finite alphabet

Infinite alphabet

## Monadic Second Order logic

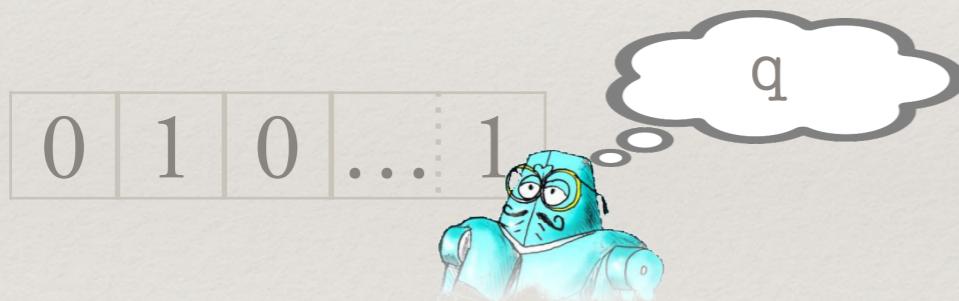
$\exists Z \ first \in Z \wedge last \notin Z \wedge \forall x (x \in Z \leftrightarrow x+1 \notin Z)$

## MSO + binary relation $\approx$

$\exists Z \ first \in Z \wedge last \notin Z \wedge \forall x (x \in Z \leftrightarrow x+1 \notin Z)$   
 $\wedge \forall x, y \in Z \ x \approx y$

$x, y$  have same values

## Finite Automata



## Finite Semigroups

$$s \cdot t = r$$

Finite alphabet

Infinite alphabet

## Monadic Second Order logic

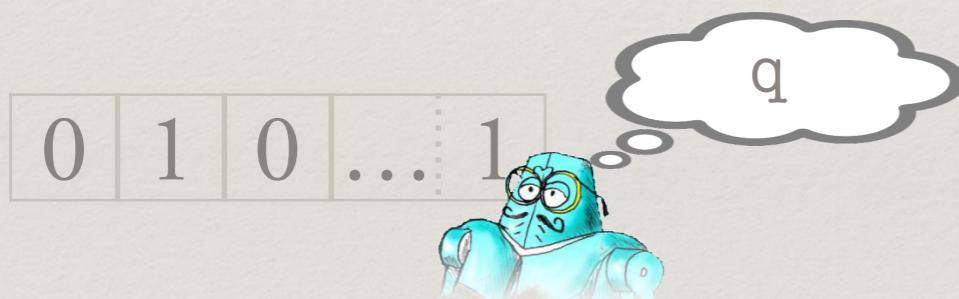
$\exists Z \ first \in Z \wedge last \notin Z \wedge \forall x (x \in Z \leftrightarrow x+1 \notin Z)$

## MSO + binary relation $\approx$

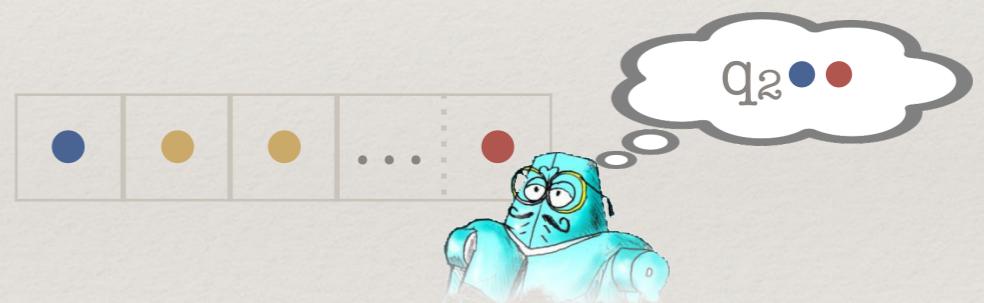
$\exists Z \ first \in Z \wedge last \notin Z \wedge \forall x (x \in Z \leftrightarrow x+1 \notin Z)$   
 $\wedge \forall x, y \in Z \ x \approx y$

$x, y$  have same values

## Finite Automata



## Register Automata



## Finite Semigroups

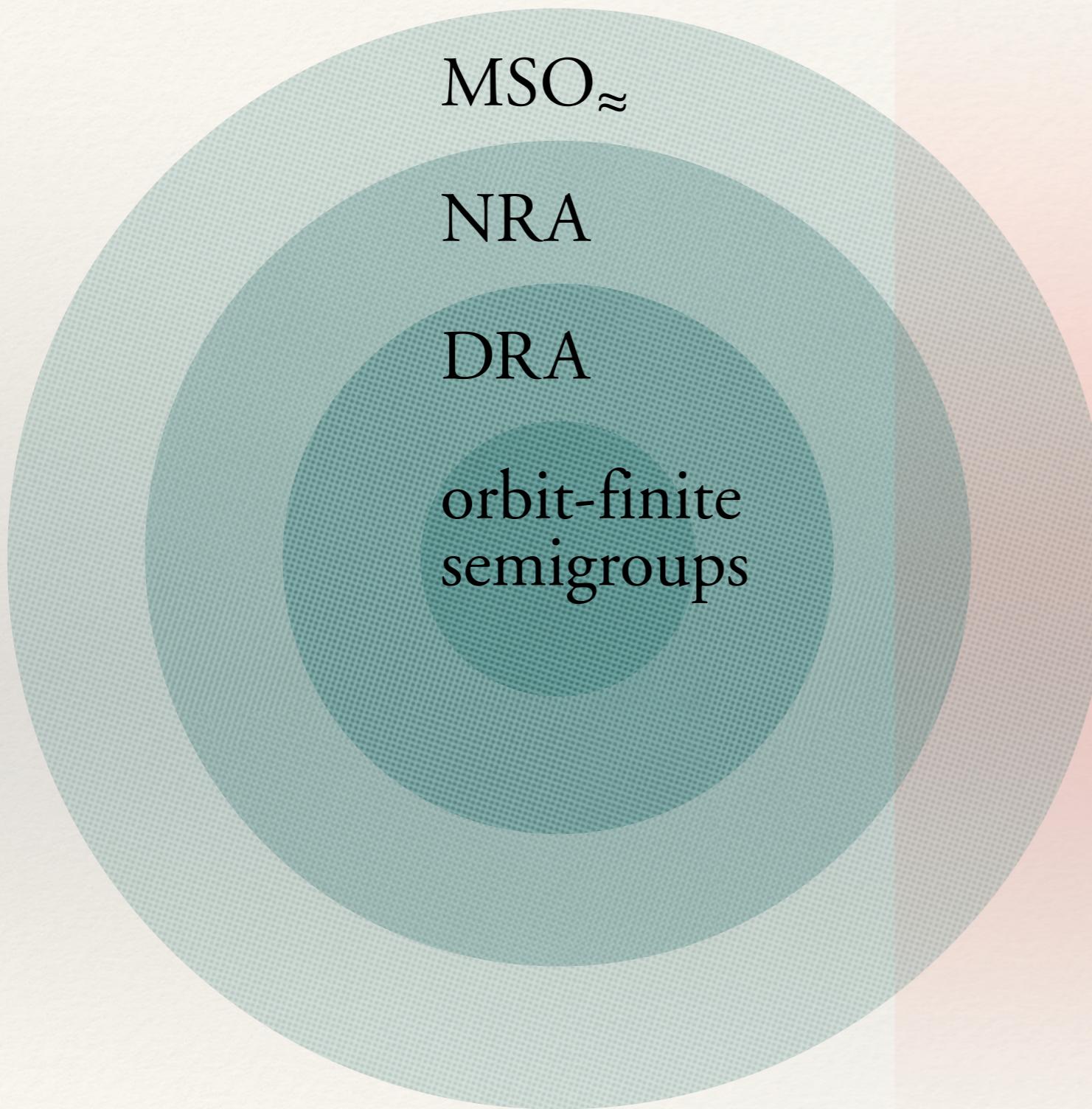
$$s \cdot t = r$$

## Semigroups + group actions

$$s[\bullet\bullet] \cdot t[\bullet] = r[\bullet\bullet]$$
$$s[\bullet\bullet] \cdot t[\bullet] = r[\bullet\bullet]$$



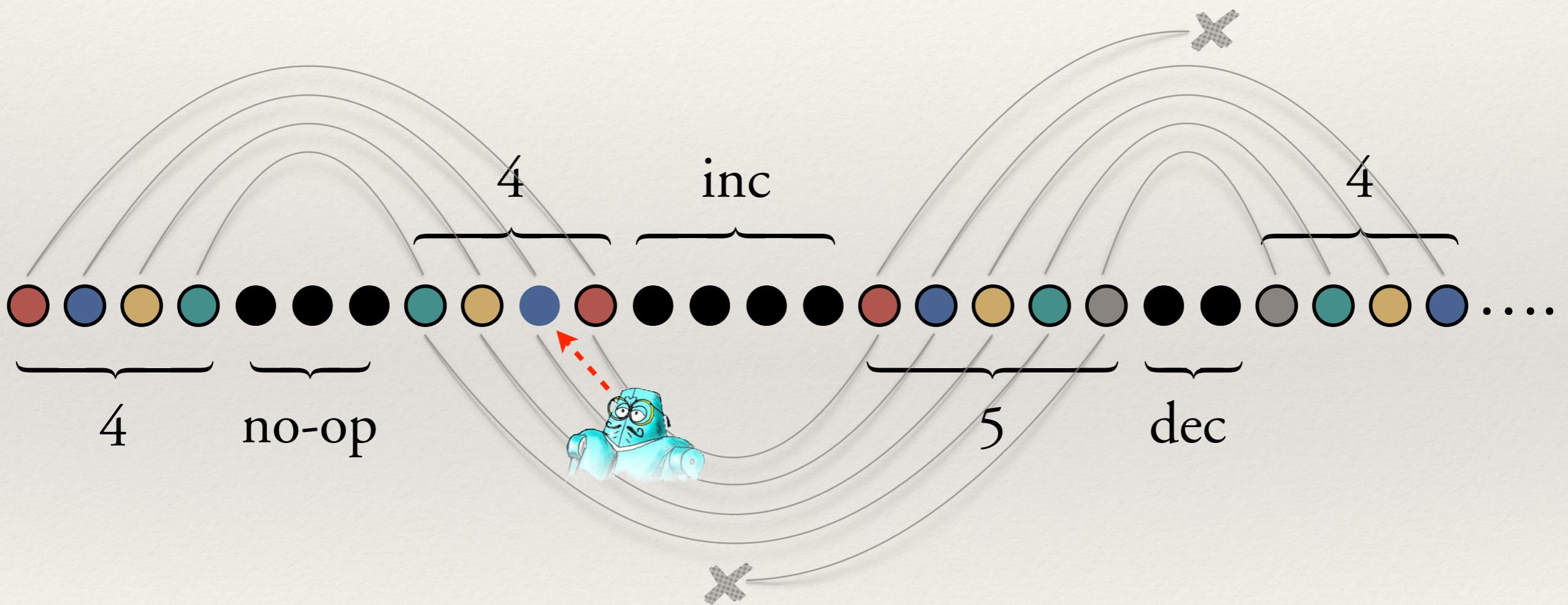
undecidable  
equivalence



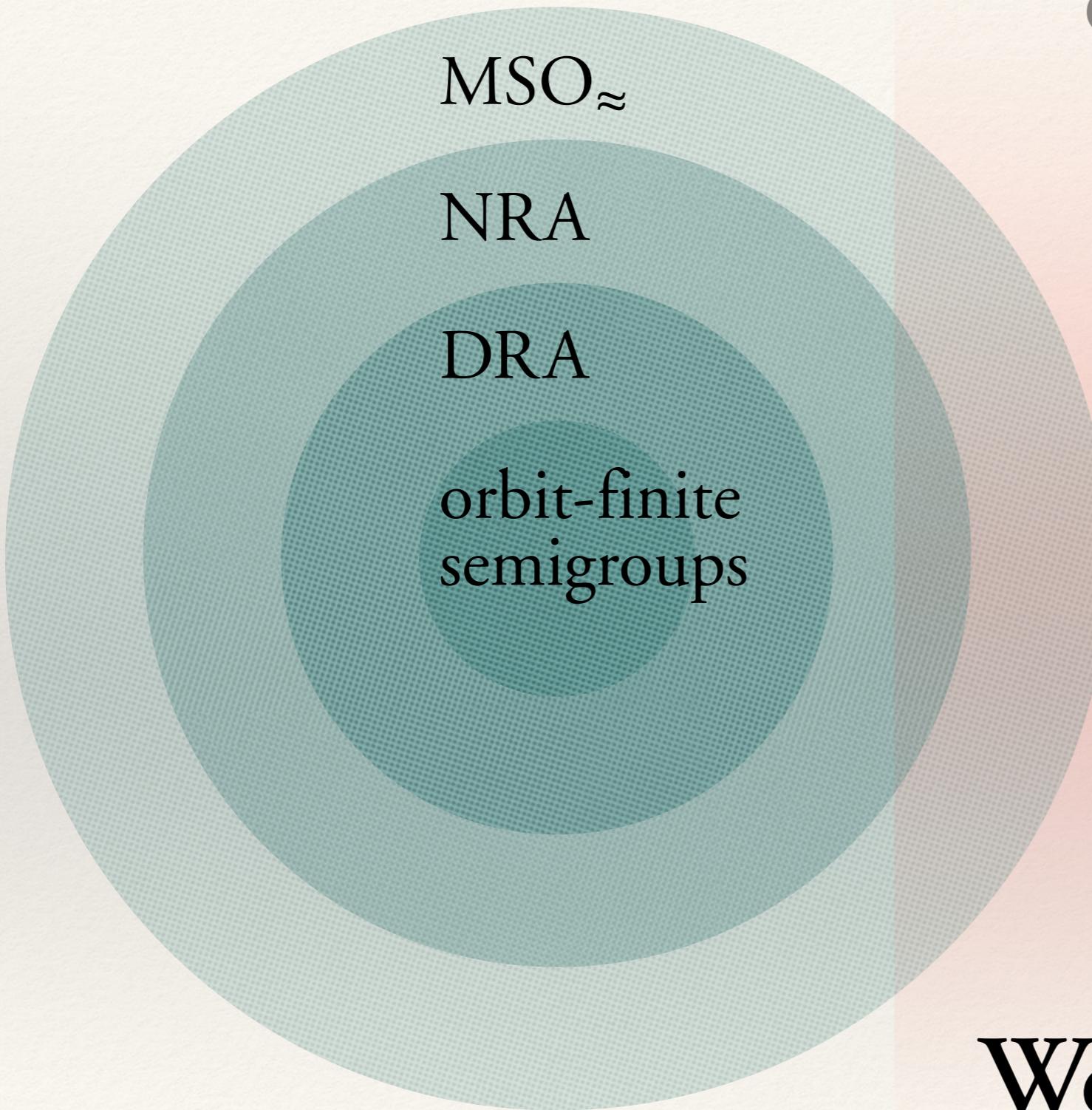
undecidable  
equivalence

## Equivalence of $\text{MSO}_{\approx}$ / NRA is undecidable

Consider encodings of halting runs of counter machines:



Wrong encodings can be detected in  $\text{MSO}_{\approx}$  or by an NRA



Goal: find a formalism

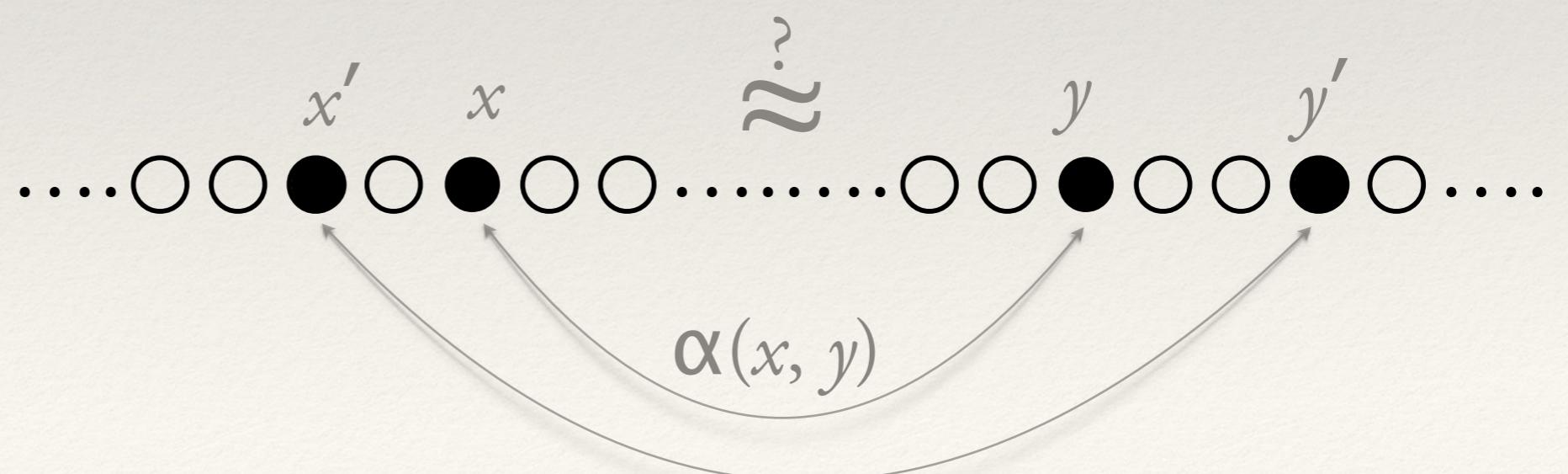
- decidable
- expressive
- robust

Attempt #1:  
Weaken MSO $\approx$  ...

# Weaken MSO $\approx$

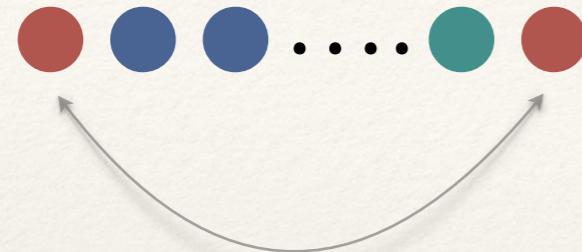


Constrain every use of  $x \approx y$  to appear in conjunction with some guard  $\alpha(x, y)$  defining a **partial bijection** between interpretations of  $x, y$

$$\varphi ::= x < y \parallel x \in Z \parallel \neg\varphi \parallel \varphi \wedge \varphi \parallel \exists x \varphi \parallel \exists Z \varphi \parallel \alpha(x, y) \wedge x \approx y$$


“first value = last value”:

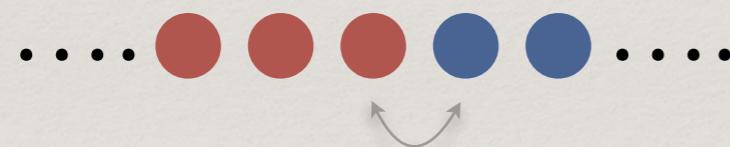
$$\exists x, y \ x = \text{first} \wedge y = \text{last} \wedge x \approx y$$



...but “first value reappears” cannot be defined

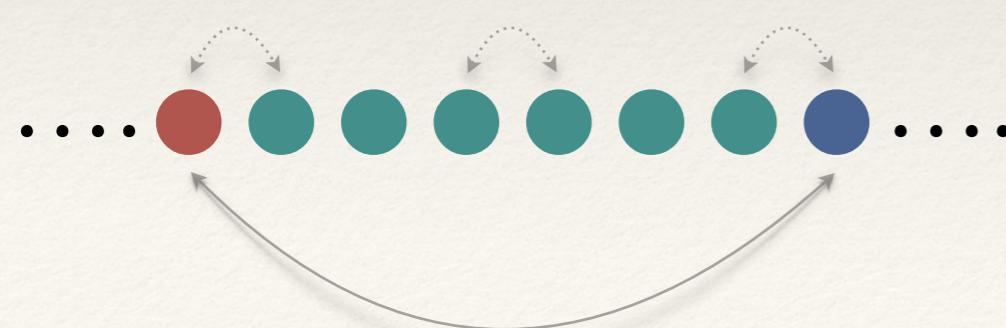
“at least 2 different values”:

$$\exists x, y \ y = x+1 \wedge x \neq y$$



“at least 3 different values”:

$$\begin{aligned} \exists x, y \ & (x \neq x+1) \wedge (y-1 \neq y) \\ & \wedge \forall z \ (x < z < y-1) \rightarrow (z \approx z+1) \\ & \wedge x \neq y \end{aligned}$$

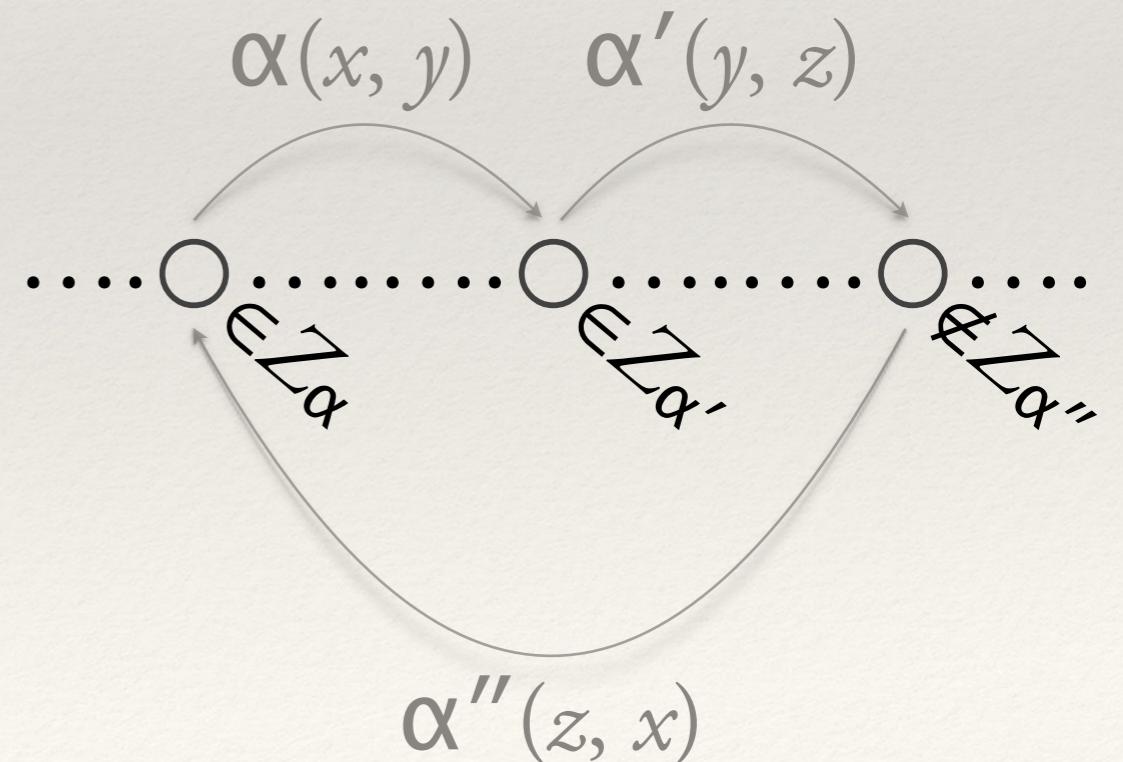


## Satisfiability of Guarded MSO $\approx$ is decidable

 Encode guarded tests by unary predicates and reduce to satisfiability of MSO (without  $\approx$ ):

$$\alpha(x, y) \wedge x \approx y \longrightarrow x \in Z_\alpha$$

+ forbid invalid encodings:



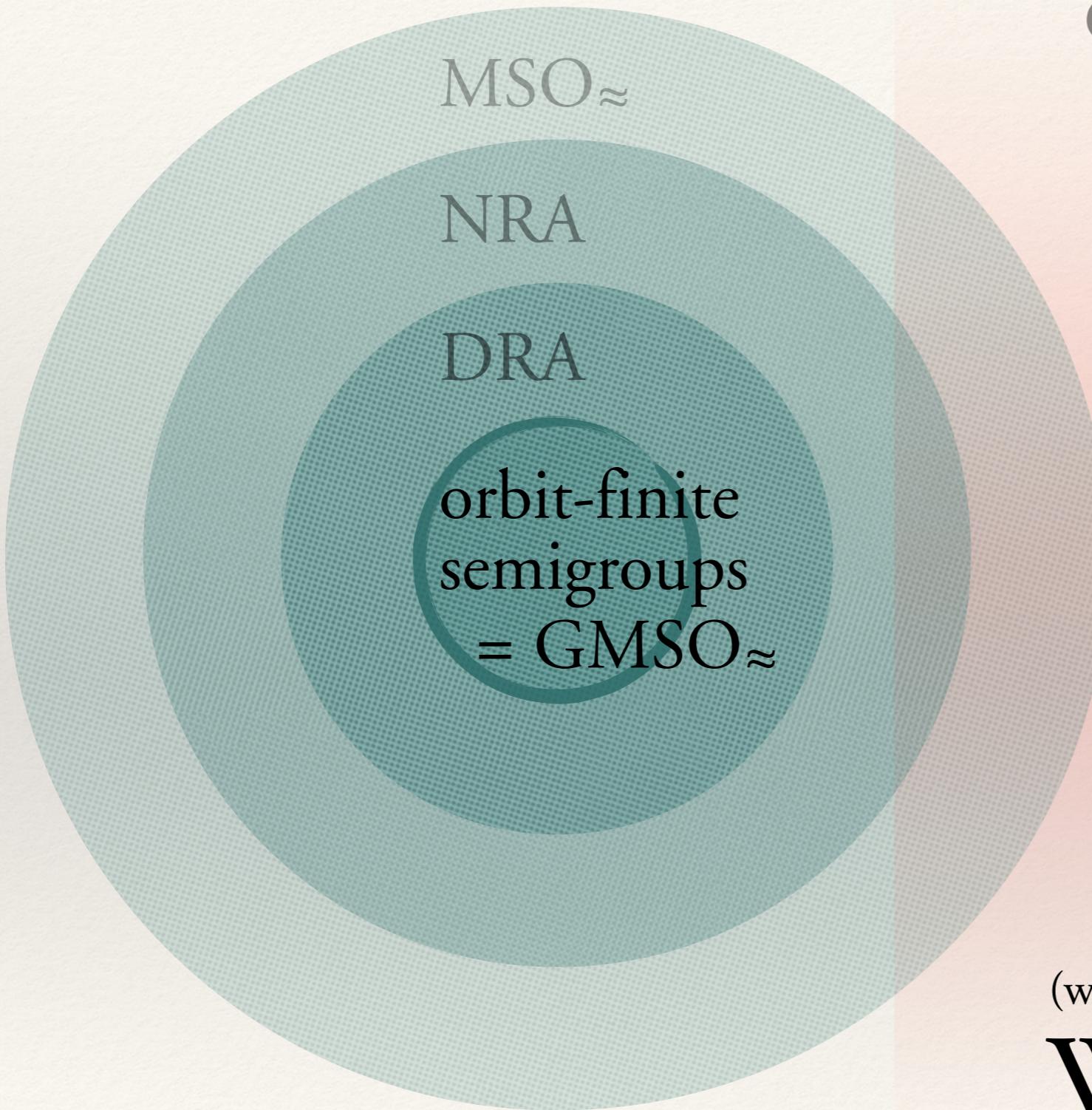
Guarded MSO $\approx$  = Orbit-finite semigroups

Guarded FO $\approx$  = Orbit-finite aperiodic semigroups

 From MSO $\approx$  to Orbit-finite semigroups:  
exploit closure under *union, complement, projection*

From Orbit-finite semigroups to MSO $\approx$ :  
variant of Schützenberger's characterisation of star-free languages

$$\begin{array}{c}
 s[\bullet\bullet\bullet] \cdot t[\bullet] = r[\bullet] \\
 \underbrace{\hspace{10em}}_{\text{Orbit-finite semigroup}} \quad \underbrace{\hspace{10em}}_{\text{MSO } \approx} \\
 \begin{array}{ccccccc} \circ & \dots & \bullet & \dots & \bullet & \dots & \circ \\ \uparrow & & \uparrow & & \uparrow & & \uparrow \\ \varphi_s(x, z_1, z_2, y) & \circ & \varphi_t(x', z', y') & = & \varphi_r(x, z_2, y') \end{array}
 \end{array}$$

Guarded MSO $\approx$ 

- decidable
- expressive
- robust
- algebraic characterisation

**Attempt #2:**  
(work in progress)  
**Weaken NRA ...**

# Weaken NRA

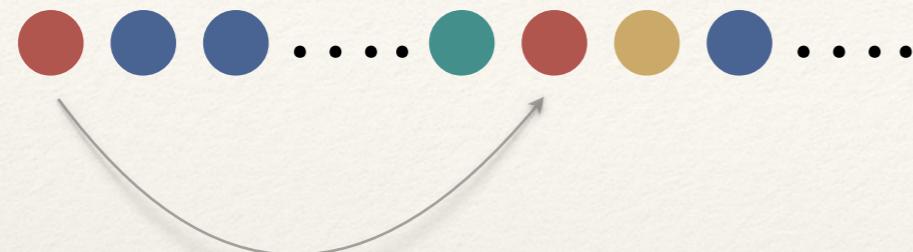


Only allow small amount of non-determinism

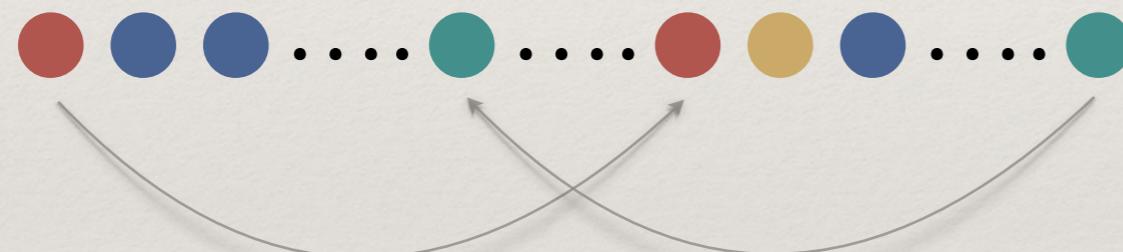
$\text{U}_{\text{nambiguousRA}} = \forall w \ \exists^{\leq 1} \text{ accepting run on } w$



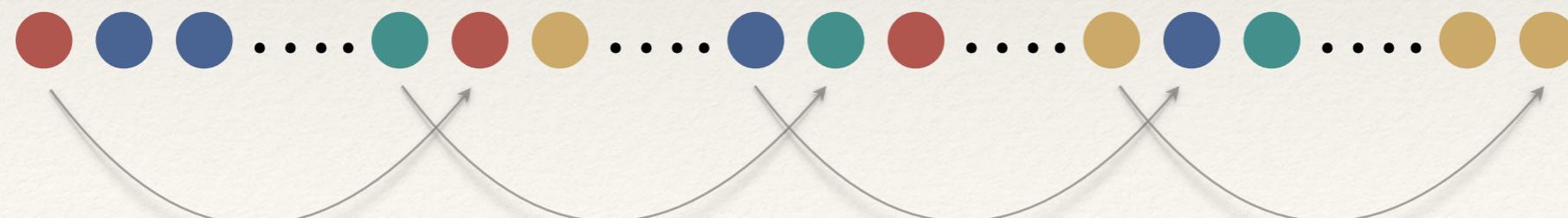
“first or last value reappears” ([co]DRA)



“first value reappears after occurrence of last value” (cascade products)



“bridges”: (bounded crossing 2wayDRA)



Main Conjecture:

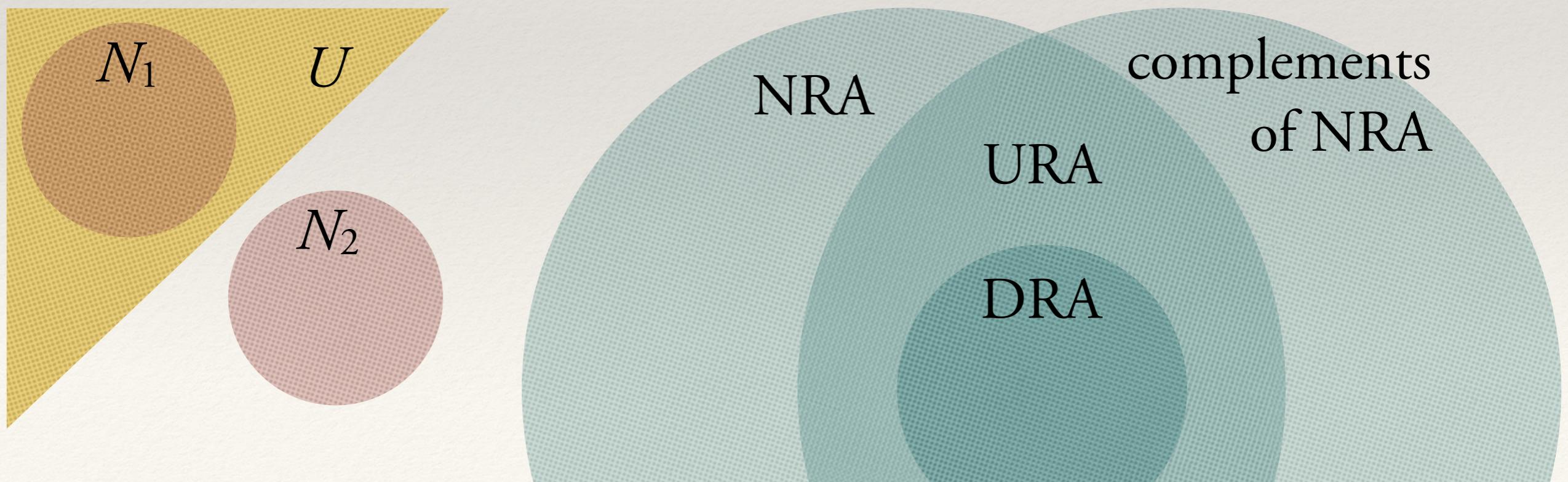
URA form the *largest subclass* of NRA **closed under complement**

Sub-conjecture 1:

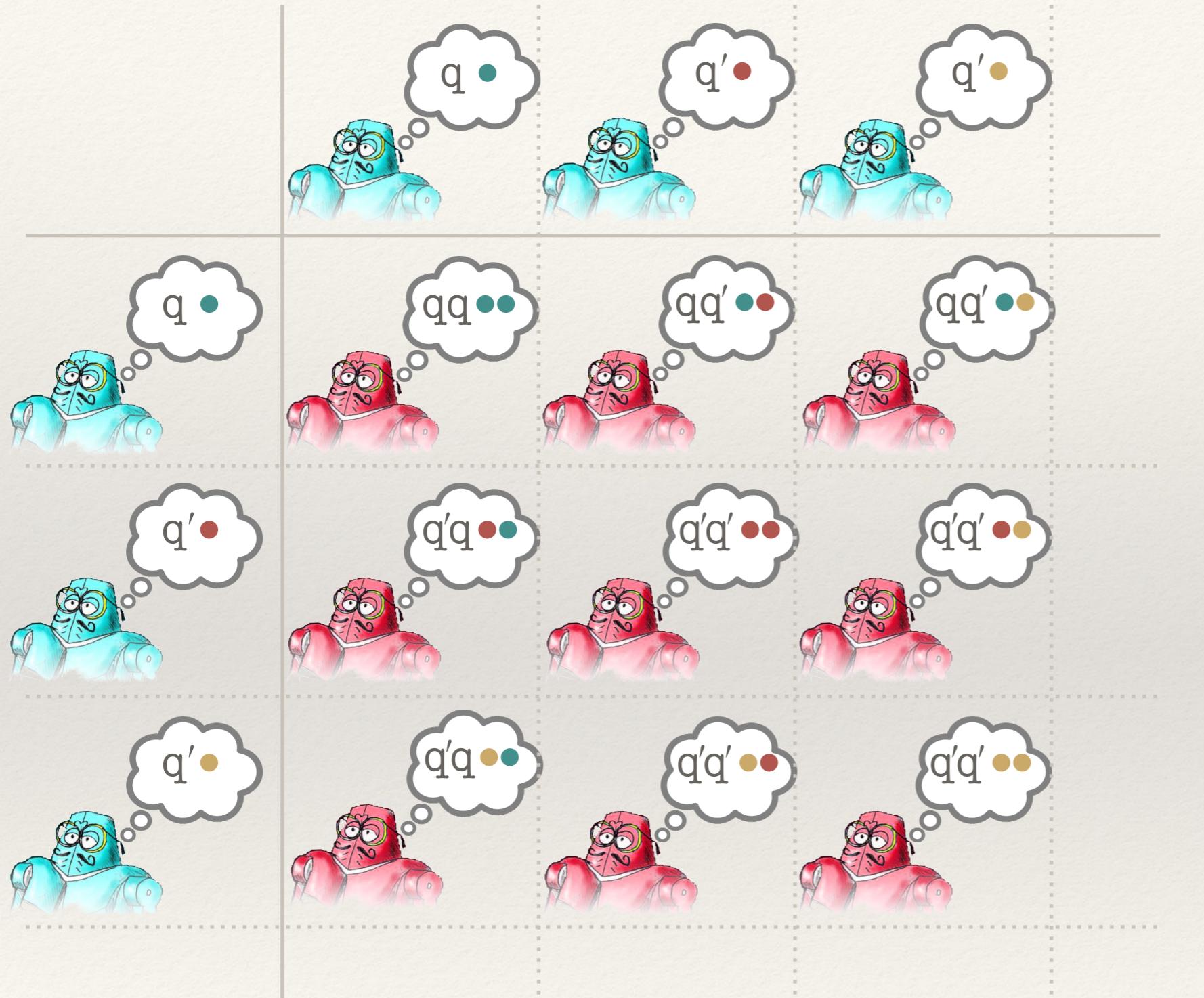
URA are closed under complement

Sub-conjecture 2:

Disjoint NRA can be *separated* by URA



## Abstracted square automaton



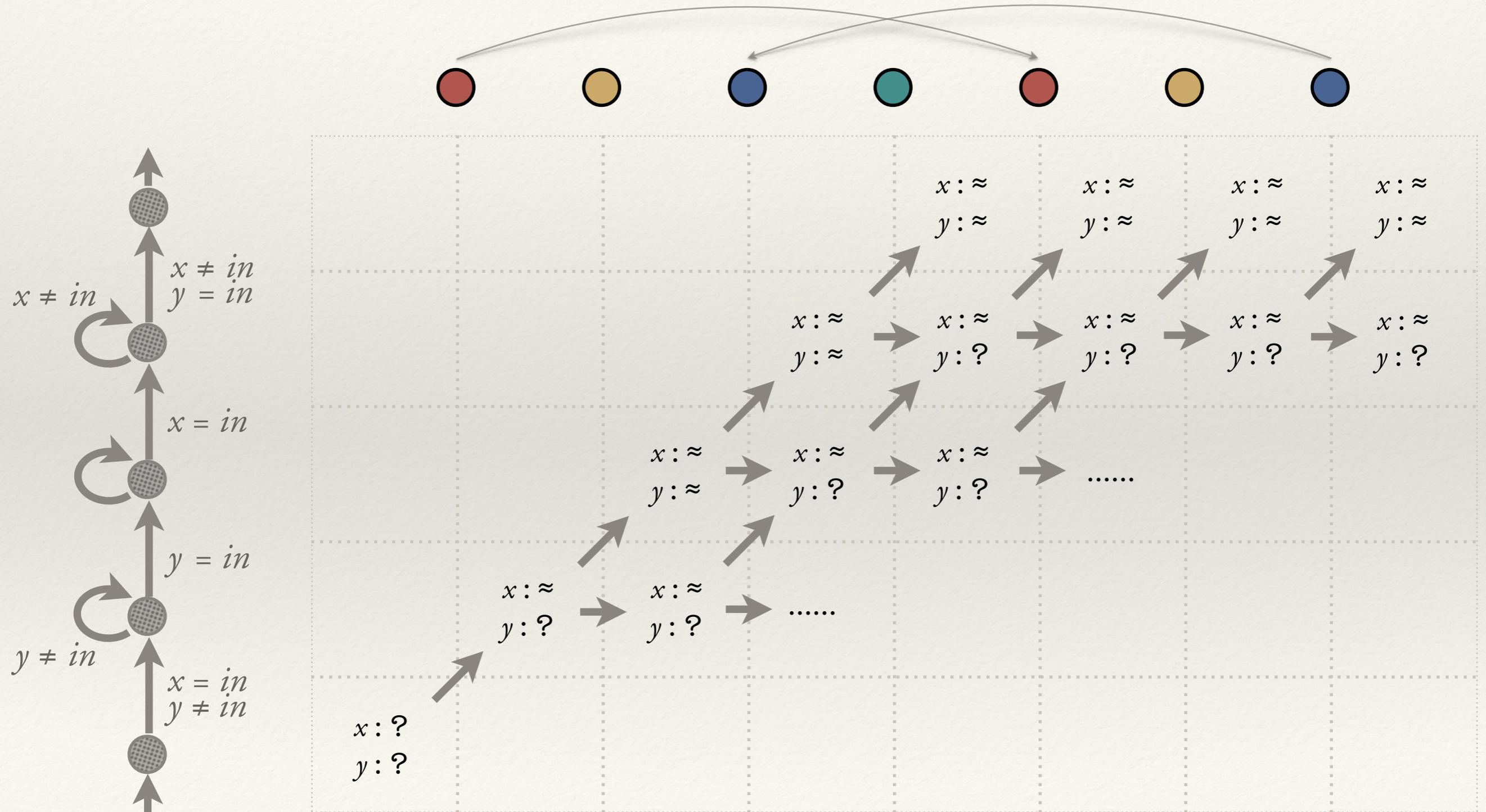
$qq \approx$   
 $qq' \neq q'q \approx$   
 $q'q' ?$

Used to check

- ❖ unambiguity
- ❖ disjointness

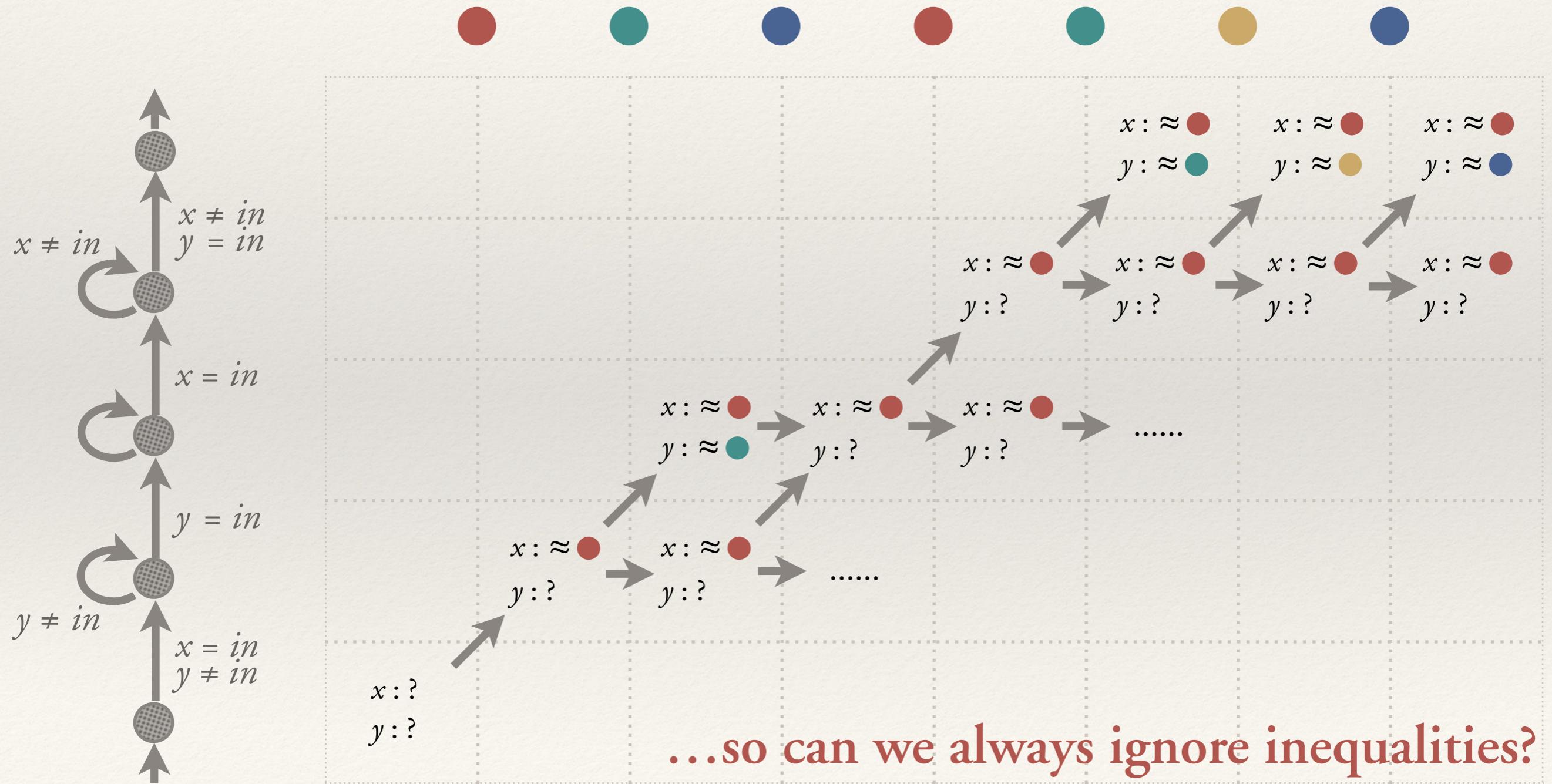


The abstracted square automaton can be turned into a **proof system** that propagates (in)equalities *forth and back*





The abstracted square automaton can be turned into a **proof system** that propagates (in)equalities *forth and back* (moreover, emerging equalities can be annotated with values)



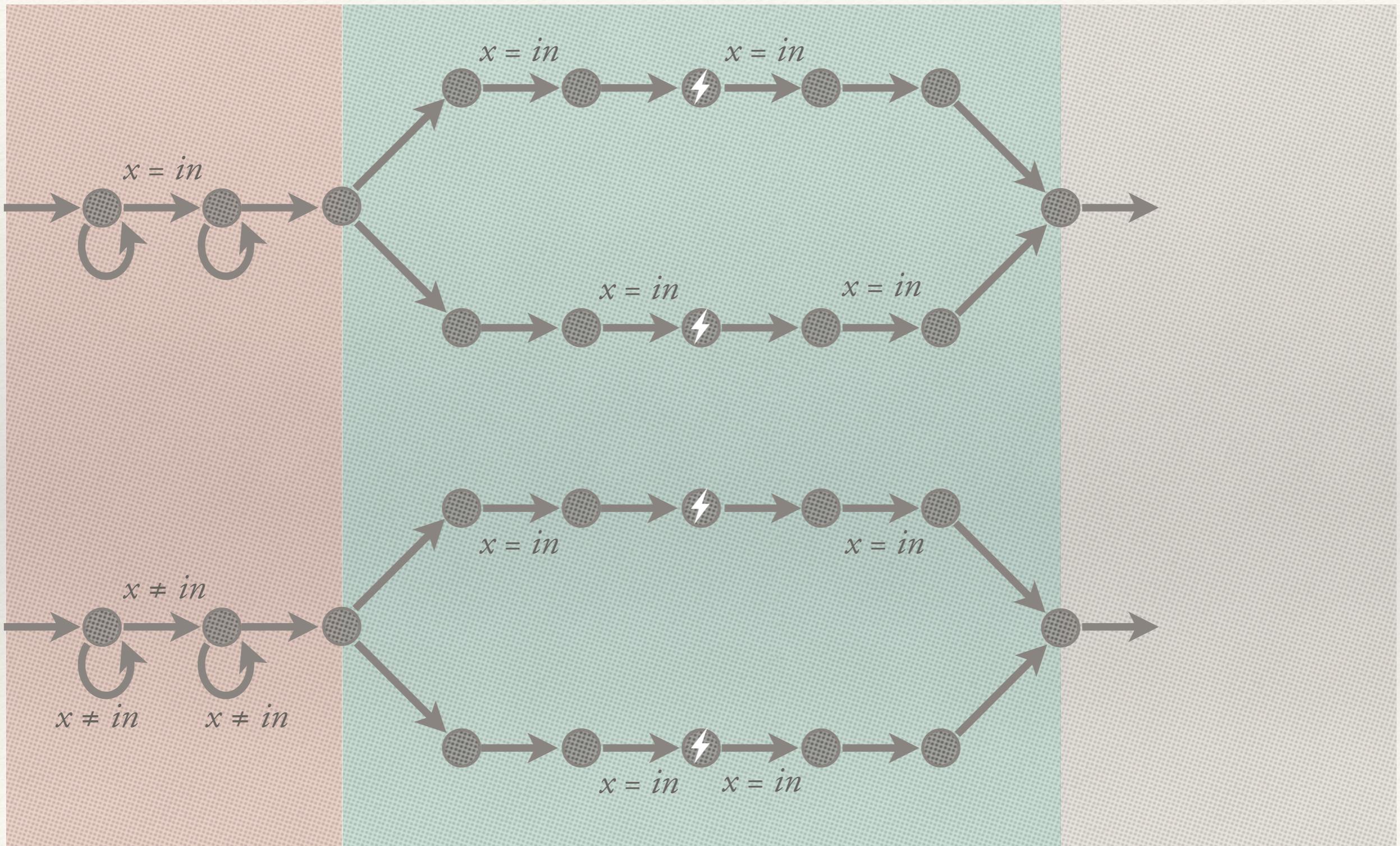
$x : \not\approx$

$x : \approx$

$\vee$

$x : \approx$

$x : ?$



What we have:

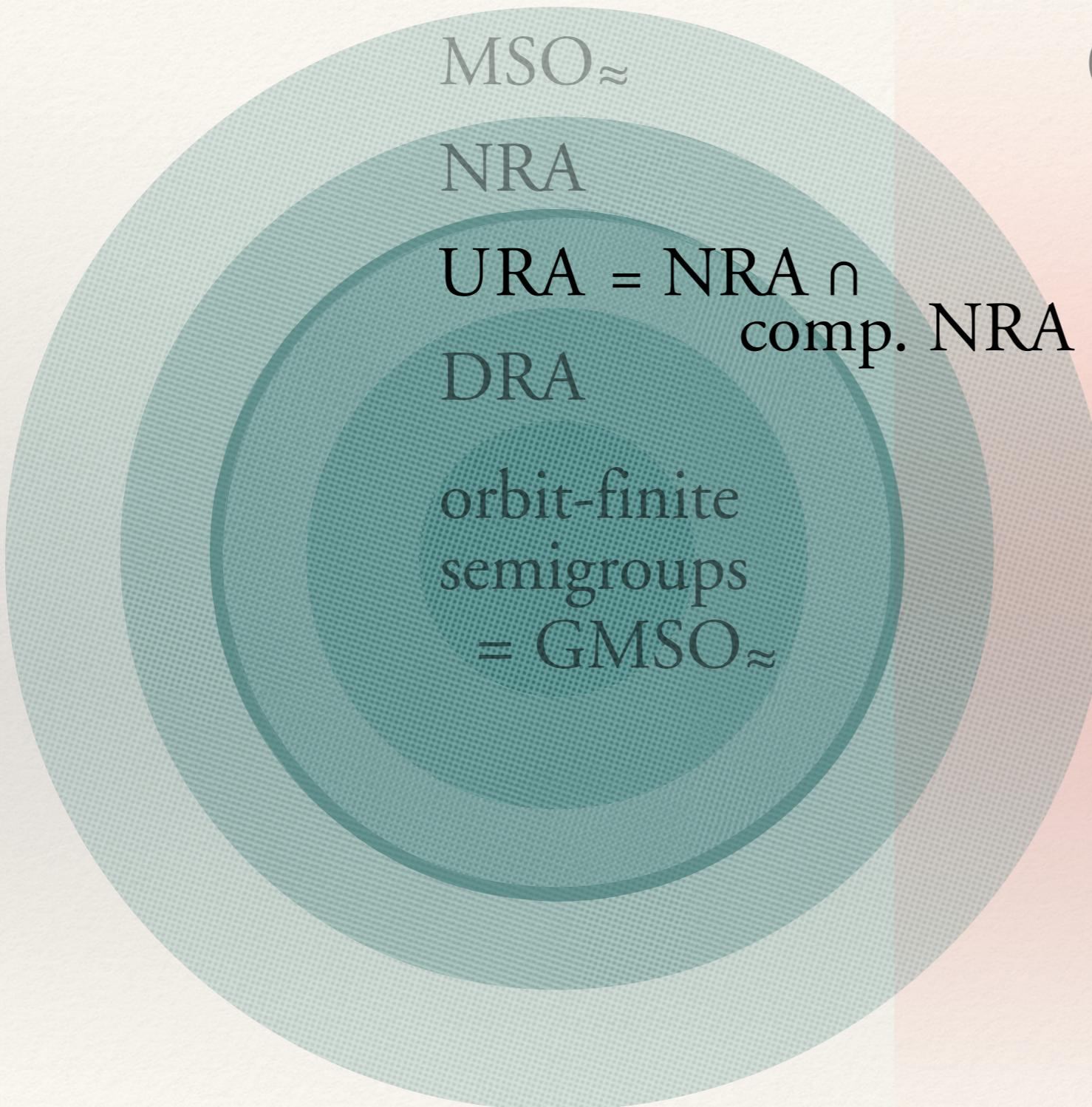
- ❖ Proof system “ $\mathcal{A}$ ” with  $\approx, \not\approx$  (sound & complete)
- ❖ Proof system “ $\mathcal{B}$ ” with only  $\approx$  (sound & complete w.r.t.  $\approx$ )

What we would like to have:

- ❖ Proof system “ $\mathcal{C}$ ” that receives input  $w$  and returns *set of pseudo-runs*  $\mathcal{C}(w)$  s.t.  
 $(\text{Runs-accepting-}w)^2 \subseteq \mathcal{C}^2(w) \subseteq \mathcal{B}(w)$
- ❖ Bi-unambiguous RA recognising both  $\{w : \mathcal{C}^2(w) \neq \emptyset\}$  and its complement

How we would like to use this:

- ❖ for URA  $U$ ,  $\mathcal{C}(w) = \emptyset$  ( $U$  rejects  $w$ )  
or  $\mathcal{C}(w)$  is singleton ( $U$  accepts  $w$ )  $\implies$  **complementation** of URA!
- ❖ for union  $N_1 + N_2$  of disjoint NRA,  
 $\exists i=1,2 \quad \mathcal{C}(w) \cap \text{Runs-of-}N_i = \emptyset$  ( $N_i$  rejects  $w$ )  $\implies$  **separation** by URA!



Unambiguous RA  
(assuming conjectures hold)

- decidable
- robust
- expressive  
(*best we can expect*)
- logical/algebraic characterisations