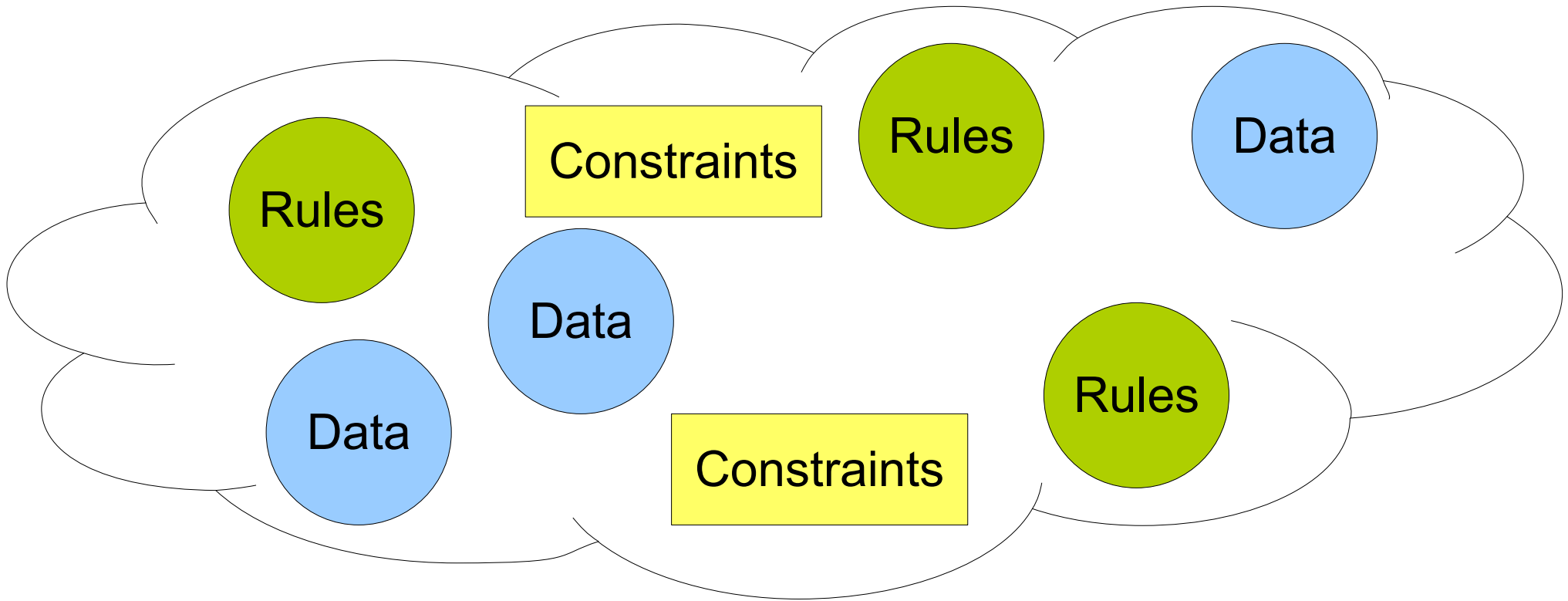


# Inconsistency in Rule Processing

Eric Kao  
Stanford University  
erickao@stanford.edu

Joint work with Tim Hinrichs, Michael Kassoff,  
Michael Genesereth

# Web Knowledge



# Inconsistencies

- Occasional errors and disagreements are unavoidable in real-world data.
  - Data acquisition error
  - Out-of-sync
  - Genuine disagreement: Julius Caesar birth year
  - Semantic disagreement: measuring GDP
  - Entity-resolution errors
  - Approximation – apparent contradictions

# Example

- Data:  
birth(adam, 1980)  
reces(1980)  
  
reces(1991)
- Constraint:  
:- birth(x,y) & birth(x,z) & y≠z
- Rules:  
adult(x) :- birth(x,z) & z < 1990  
rBorn(x) :- birth(x,y) & reces(y)  
rbAdult(x) :- rBorn(x) & adult(x)

# Example

- Data:  
birth(adam, 1980)  
reces(1980)  
birth(cody, 1984)  
birth(cody, 1991)  
reces(1991)
- Rules:  
adult(x) :- birth(x,z) & z < 1990  
rBorn(x) :- birth(x,y) & reces(y)  
rbAdult(x) :- rBorn(x) & adult(x)

Constraint:

$:- \text{birth}(x,y) \ \& \ \text{birth}(x,z) \ \& \ y \neq z$

# Ignore constraints?

- Data:

birth(adam, 1980)

reces(1980)

birth(cody, 1984)

birth(cody, 1991)

reces(1991)

- Rules:

adult(x) :- birth(x,z) & z < 1990

rBorn(x) :- birth(x,y) & reces(y)

rbAdult(x) :- rBorn(x) & adult(x)

# Ignore constraints?

- Data:

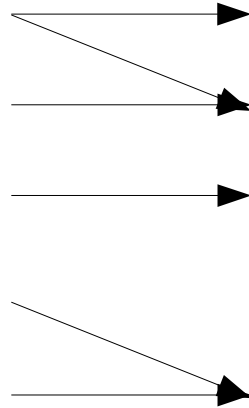
birth(adam, 1980)

reces(1980)

birth(cody, 1984)

birth(cody, 1991)

reces(1991)



adult(adam)

rBorn(adam)

adult(cody)

rBorn(cody)

- Rules:

adult(x) :- birth(x,z) & z < 1990

rBorn(x) :- birth(x,y) & reces(y)

rbAdult(x) :- rBorn(x) & adult(x)

# Ignore constraints?

- Data:

birth(adam, 1980)

reces(1980)

birth(cody, 1984)

birth(cody, 1991)

reces(1991)

- Rules:

adult(x) :- birth(x,z) & z < 1990

rBorn(x) :- birth(x,y) & reces(y)

rbAdult(x) :- rBorn(x) & adult(x)

adult(adam)

rBorn(adam)

adult(cody)

rBorn(cody)

rbAdult(adam)

rbAdult(cody)

# Ignore constraints?

- Data:

birth(adam, 1980)  
reces(1980)

birth(cody, 1984)  
birth(cody, 1991)  
reces(1991)

adult(cody)

rBorn(cody)

rbAdult(cody)

- Rules:

adult(x) :- birth(x,z) & z < 1990

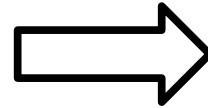
rBorn(x) :- birth(x,y) & reces(y)

rbAdult(x) :- rBorn(x) & adult(x)

# Existential Answers

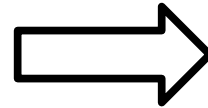
- Data:

birth(adam, 1980)  
reces(1980)



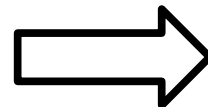
adult(adam)  
rBorn(adam)  
rbAdult(adam)

birth(cody, 1984)



adult(cody)

birth(cody, 1991)  
reces(1991)



rBorn(cody)

Elvang-Goranson & Hunter  
Kassoff & Genesereth

~~rbAdult(cody)~~

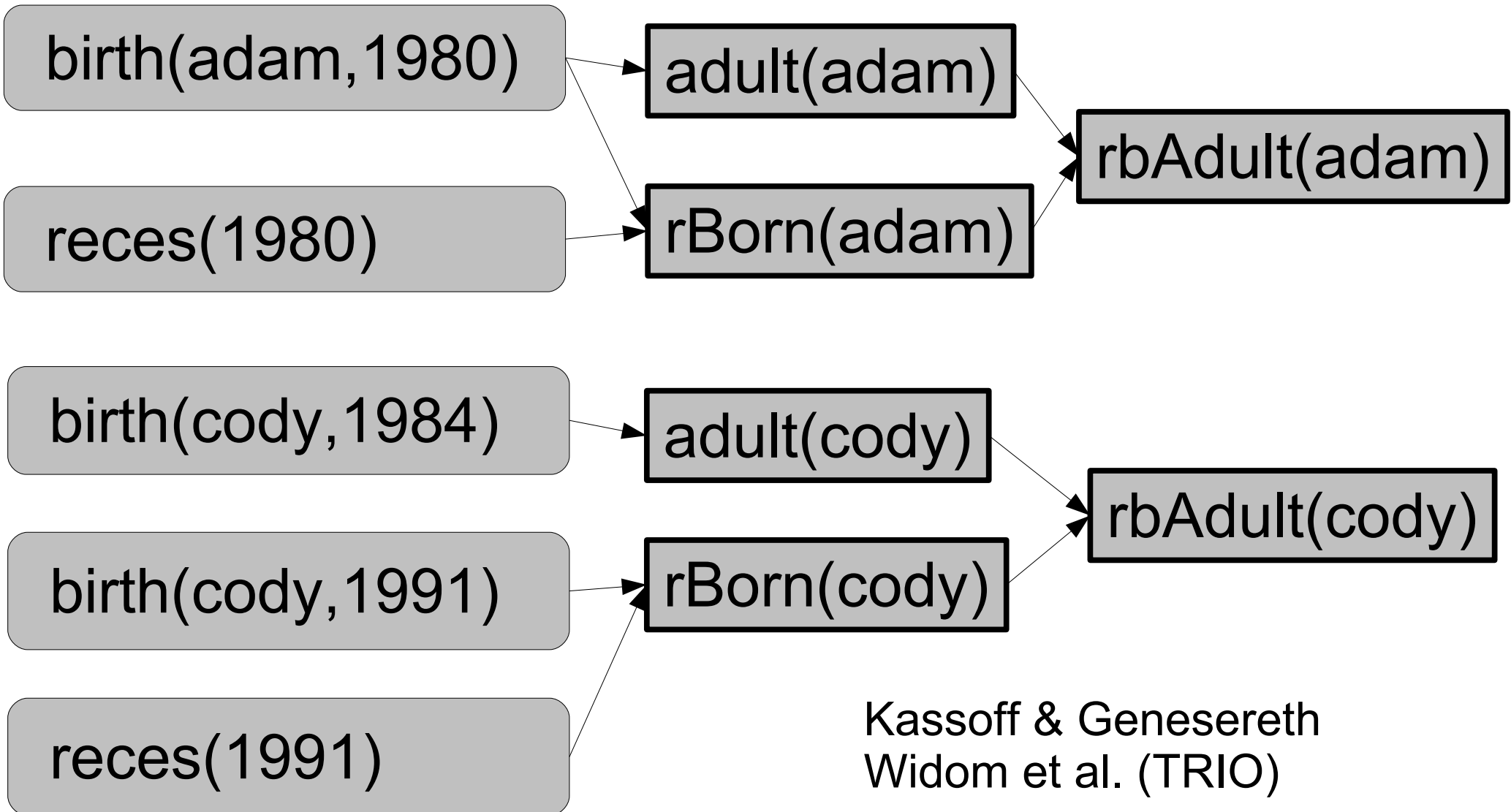
# Existential Answers

- Given
  - R: a set of rules (Datalog)
  - C: a set of constraints
  - D: a set of ground atoms
  - $D, R, C \models_E a$   
 $\Leftrightarrow$   
exists  $D^* \subseteq D$  s.t.
    - $D^*, R$  is consistent with C, and
    - $D^*, R \models a$

# Naïve Method

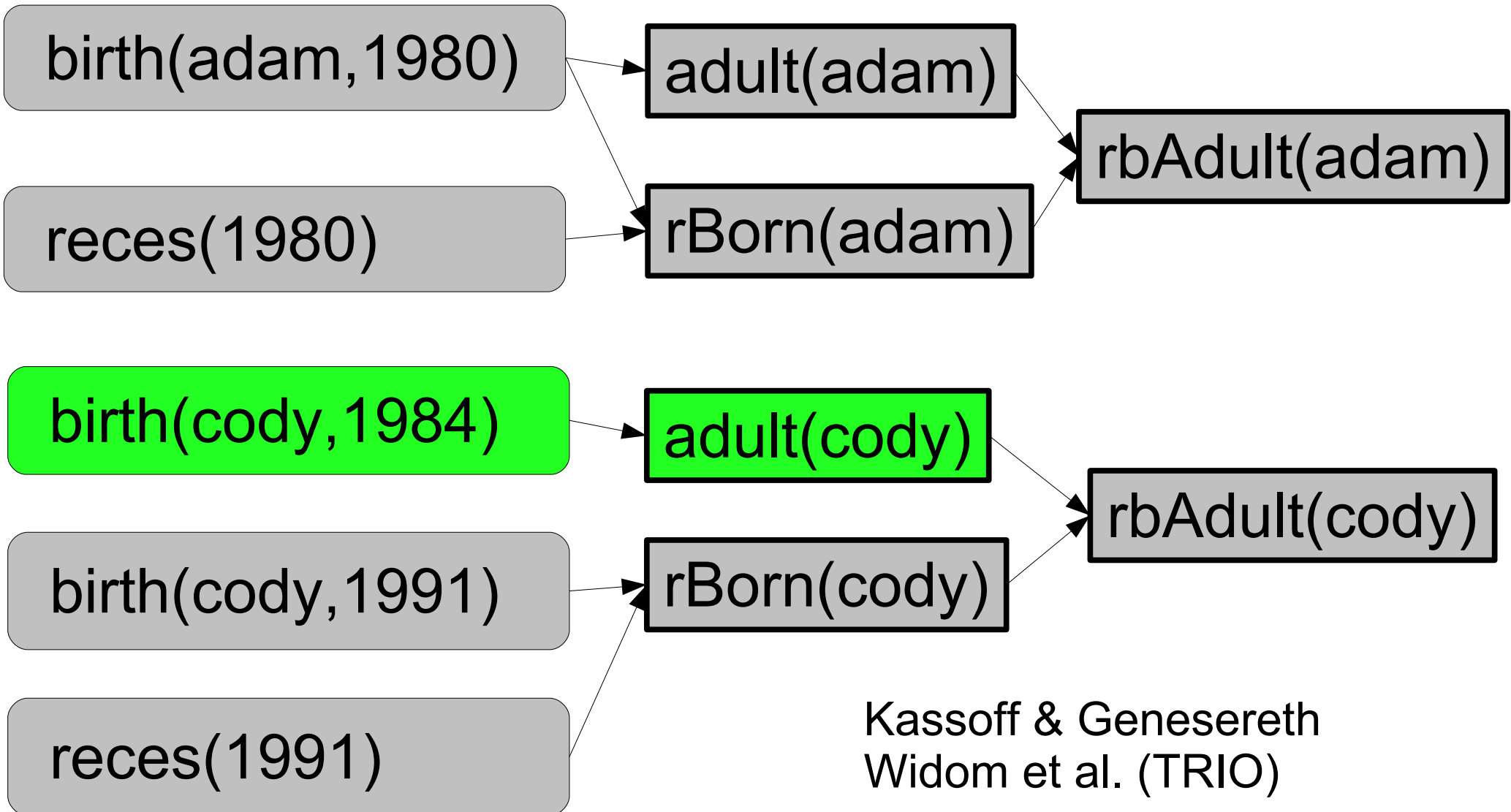
- Consider every subset  $D^*$
- If it is consistent with  $C$ , test if  $D^*, R \models a$
- Impractical because there are  $2^{|D|}$  many subsets
- Check only maximal, consistent subsets  
Still exponential in worst case.
  - e.g.  $2^{|D|/2}$

# Lineage Approach



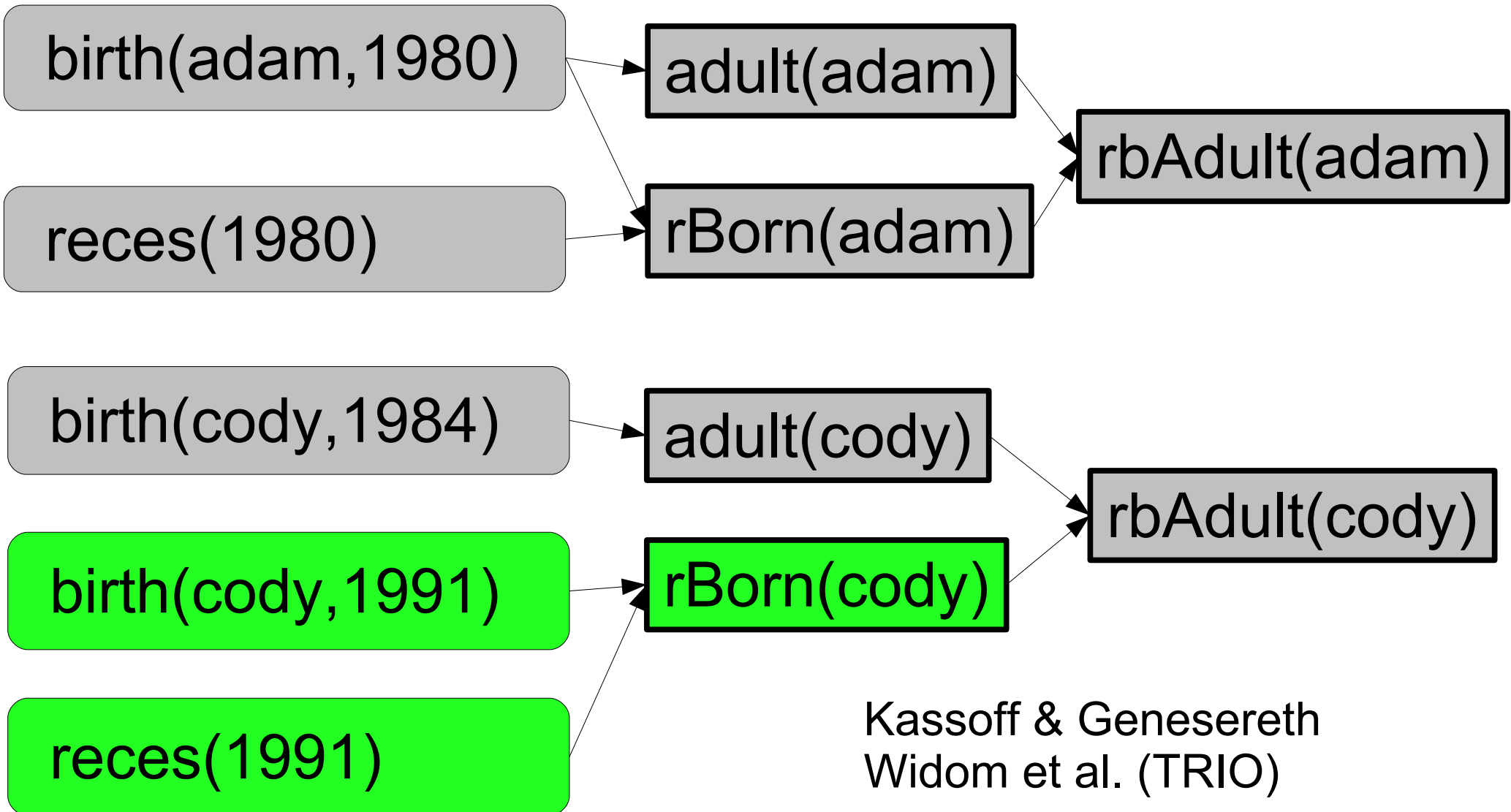
Kassoff & Genesereth  
Widom et al. (TRIO)

# Lineage Approach



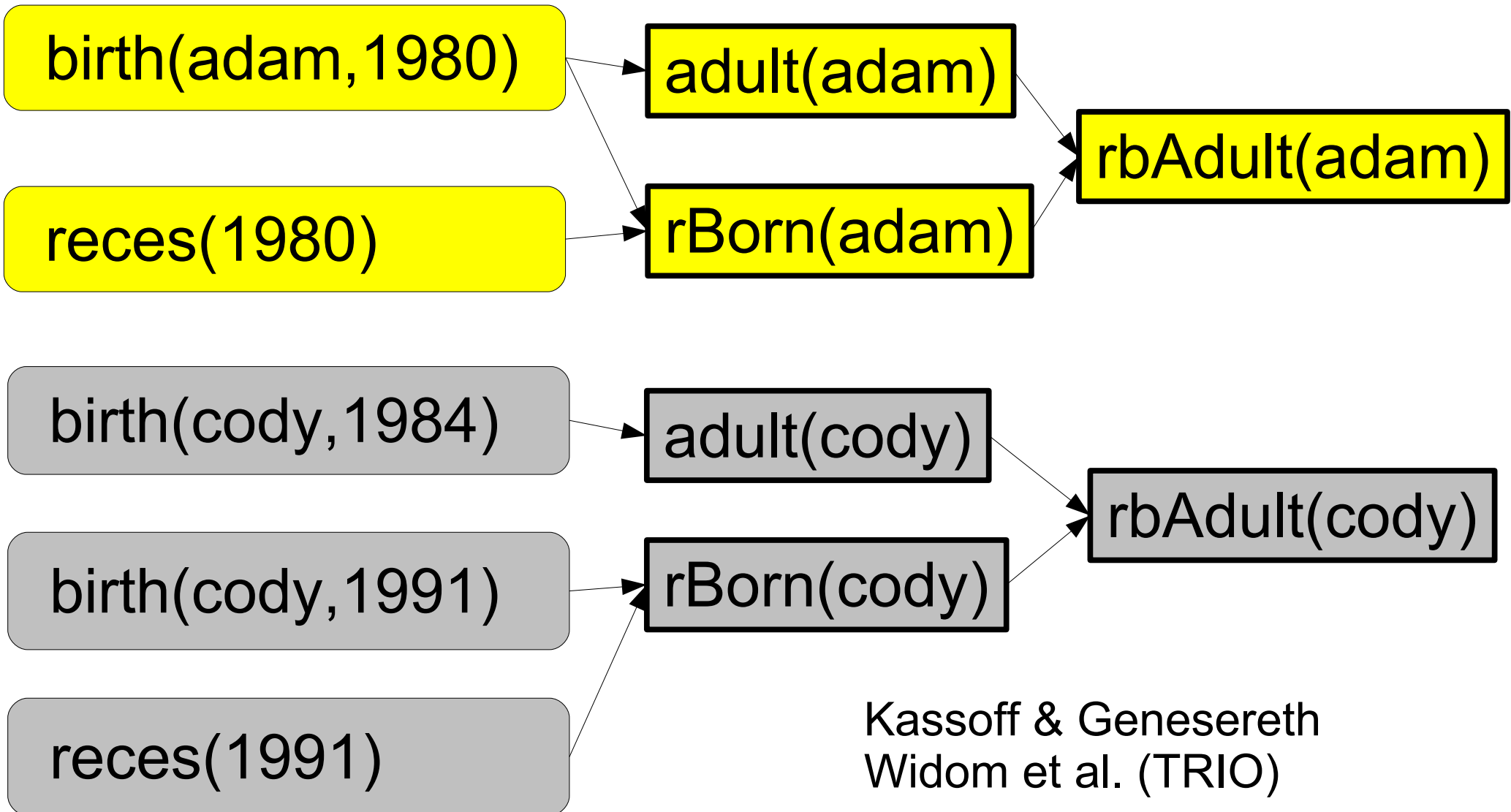
Kassoff & Genesereth  
Widom et al. (TRIO)

# Lineage Approach



Kassoff & Genesereth  
Widom et al. (TRIO)

# Lineage Approach



Kassoff & Genesereth  
Widom et al. (TRIO)

# Lineage Approach

birth(adam, 1980)

reces(1980)

adult(adam)

rBorn(adam)

rbAdult(adam)

birth(cody, 1984)

birth(cody, 1991)

reces(1991)

adult(cody)

rBorn(cody)

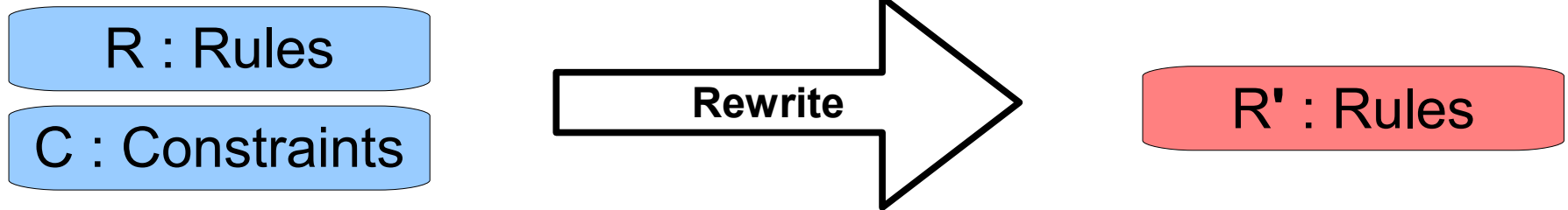
~~rbAdult(cody)~~

Kassoff & Genesereth  
Widom et al. (TRIO)

# Features

- Advantage:
  - Polynomial data-complexity for nonrecursive Datalog
- Disadvantage:
  - Nontrivial changes to rule processing infrastructure

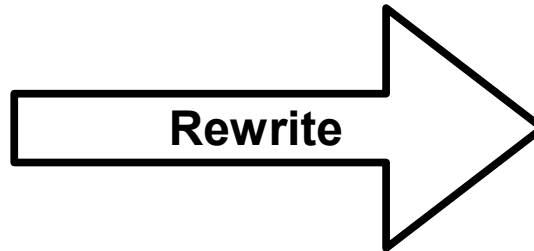
# Rule-rewriting approach



# Rule-rewriting approach

R : Rules

C : Constraints



R' : Rules



$D, R, C \models_E a$



$D, R' \models a$

Constraint:  $\text{:- birth}(x,y) \ \& \ \text{birth}(x,z) \ \& \ y \neq z$

- $\text{adult}(x) \text{ :- birth}(x,z) \ \& \ z < 1990$
- $\text{rBorn}(x) \text{ :- birth}(x,y) \ \& \ \text{reces}(y)$
- $\text{rbAdult}(x) \text{ :- rBorn}(x) \ \& \ \text{adult}(x)$





# Augment for Inconsistency

- rule:  $r(Z) :- p(X,Y) \ \& \ q(Z,1)$
- Rule body  
b:  $p(X,Y) \ \& \ q(Z,1)$
- Constraint  
c:  $:- p(2,U) \ \& \ q(U,V)$
- c,b inconsistent  $\Leftrightarrow X = 2 \ \& \ Y = Z$
- Augmented rule:

$r(Z) :- p(X,Y) \ \& \ q(Z,1) \ \& \ \neg[X = 2 \ \& \ Y = Z]$

# Features

- Polynomial data complexity
- Rewrite and send
- Reuse on different/evolving data
- Limitation:
  - Does not work for recursive rules

# Recursive Rules

- R:

$\text{reach}(X, Y) \text{ :- link}(X, Z, T) \ \& \ \text{reach}(Z, Y)$   
 $\text{reach}(X, Y) \text{ :- link}(X, Y, T)$

- C:

$\text{ :- p}(X, Y, T) \ \& \ \text{p}(X', Y', T) \ \& \ X \neq X' \ \& \ Y \neq Y'$

- Appears to be no Datalog rewriting

- Theorem:

Unless  $P = NP$ , some recursive rule sets do not have existential answers rewriting in Datalog

# Other Answer Semantics

- $\text{Con}(D, R, C) := \{D^* \subseteq D \mid D^*, R \text{ consistent w } C\}$
- $\text{MaxCon}(D, R, C) := \text{Maximal sets in } \text{Con}(D, R, C)$
- $\text{Free}(D, R, C) := \text{intersection of } \text{MaxCon}(D, R, C)$

- $D, R, C \models_U a$

$\Leftrightarrow$

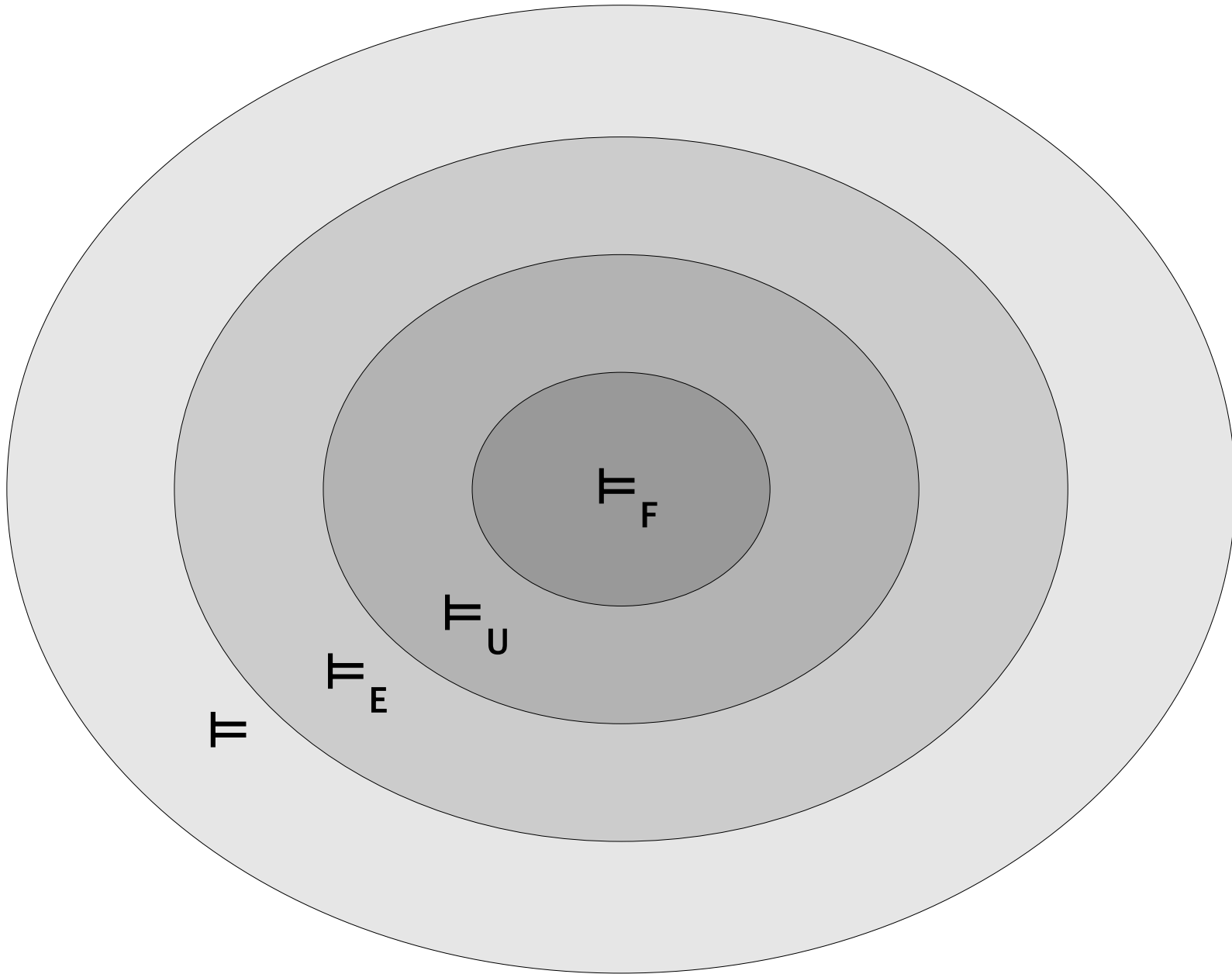
for all  $D^*$  in  $\text{MaxCon}(D, R, C)$ ,

$D^*, R \models a$

- $D, R, C \models_F a$

$\Leftrightarrow$

$\text{Free}(D, R, C), R \models a$



\*Assume  $\text{Con}(D,R,C)$  non-empty  
28 of 29

# Summary

- Existential Answers Semantics
- Lineage approach
- Rewriting approach
- Future
  - Rewriting approach for some recursive rule sets
  - Hybrid approach?
  - Degree of trustworthiness