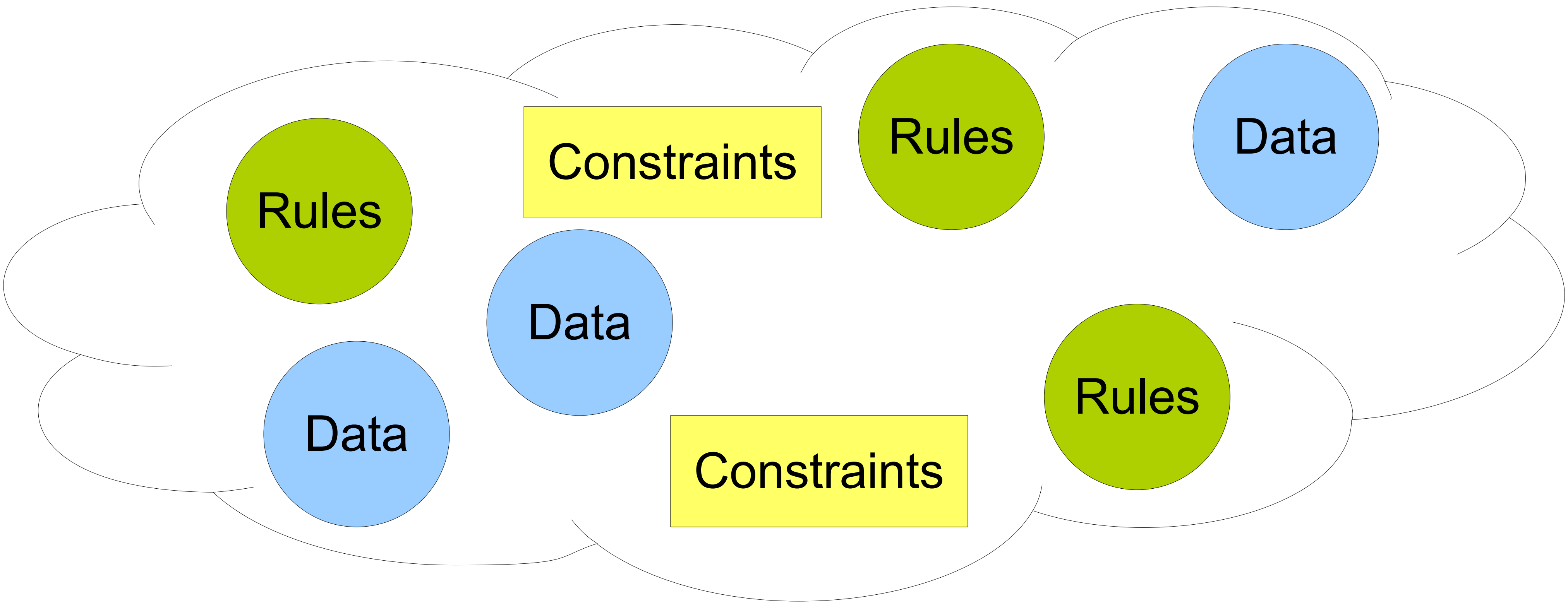
Inconsistency in Rule Processing

Eric Kao
Stanford University
erickao@stanford.edu

Joint work with Tim Hinrichs, Michael Kassoff, Michael Genesereth

Web Knowledge



Eric Kao
Stanford University

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

Eric Kao
Stanford University

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)

Eric Kao
Stanford University

Example

- Data:
  birth(adam, 1980)
  reces(1980)
  birth(cody, 1984)
  birth(cody, 1991)
  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)

Eric Kao
Stanford University

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)

Eric Kao
Stanford University

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)

Eric Kao
Stanford University

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)

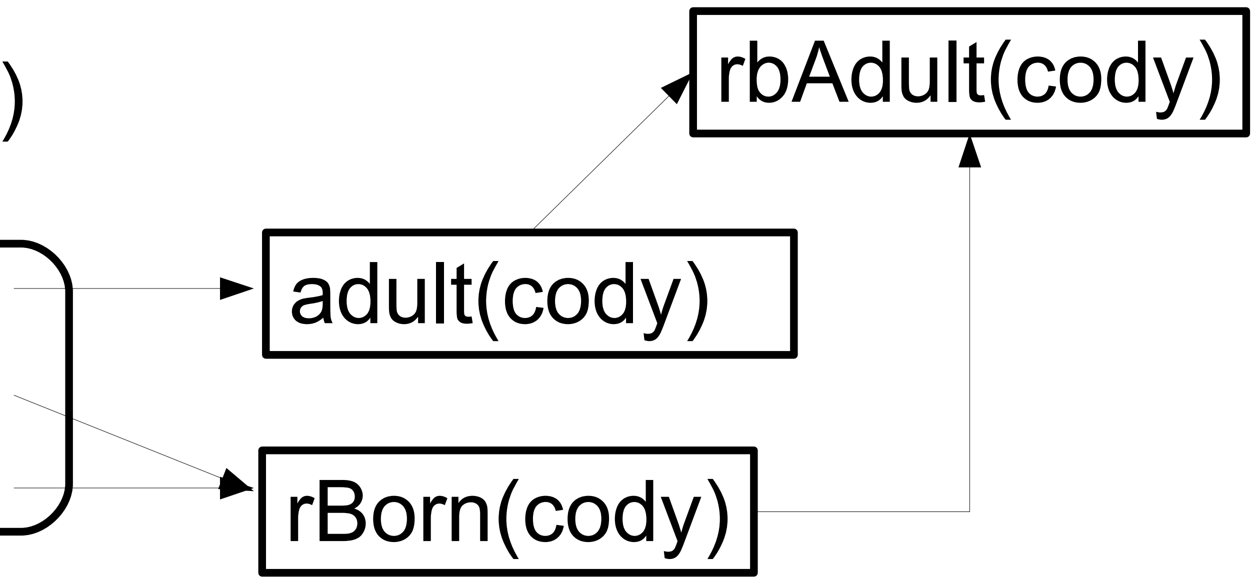
adult(adam)
rBorn(adam)
adult(cody)
rBorn(cody)

rbAdult(adam)
rbAdult(cody)

Eric Kao
Stanford University

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)

Eric Kao
Stanford University

Existential Answers

- Data:

birth(adam, 1980)
reces(1980)

$\Rightarrow$

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

birth(cody, 1984)

$\Rightarrow$

adult(cody)

birth(cody, 1991)
reces(1991)

$\Rightarrow$

rBorn(cody)

~~rbAdult(cody)~~

Elvang-Goranson & Hunter
Kassoff & Genesereth

Eric Kao
Stanford University

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

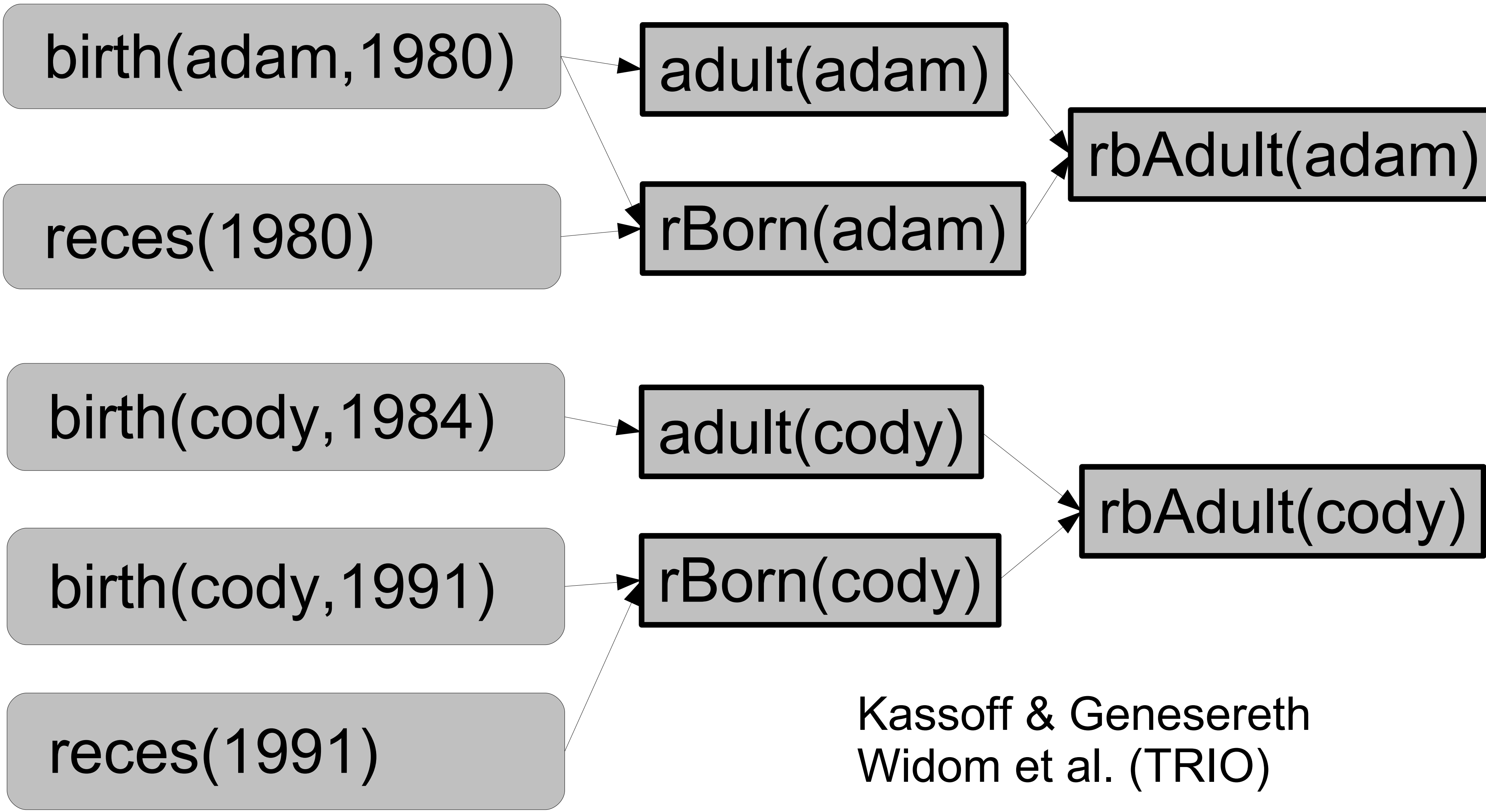
Eric Kao
Stanford University

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. $2^{|D|/2}$

Eric Kao
Stanford University

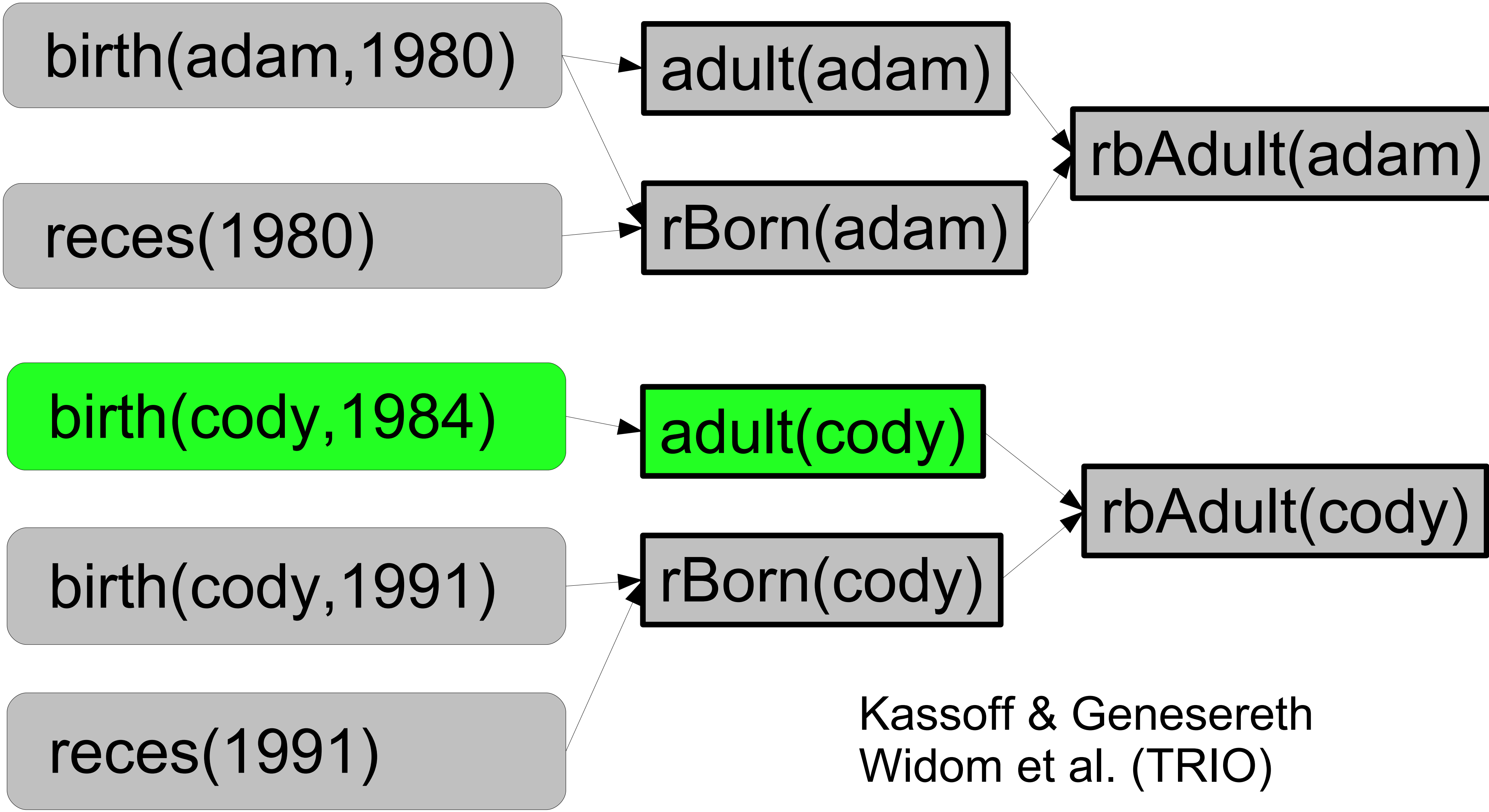
Lineage Approach



Kassoff & Genesereth
Widom et al. (TRIO)

Eric Kao
Stanford University

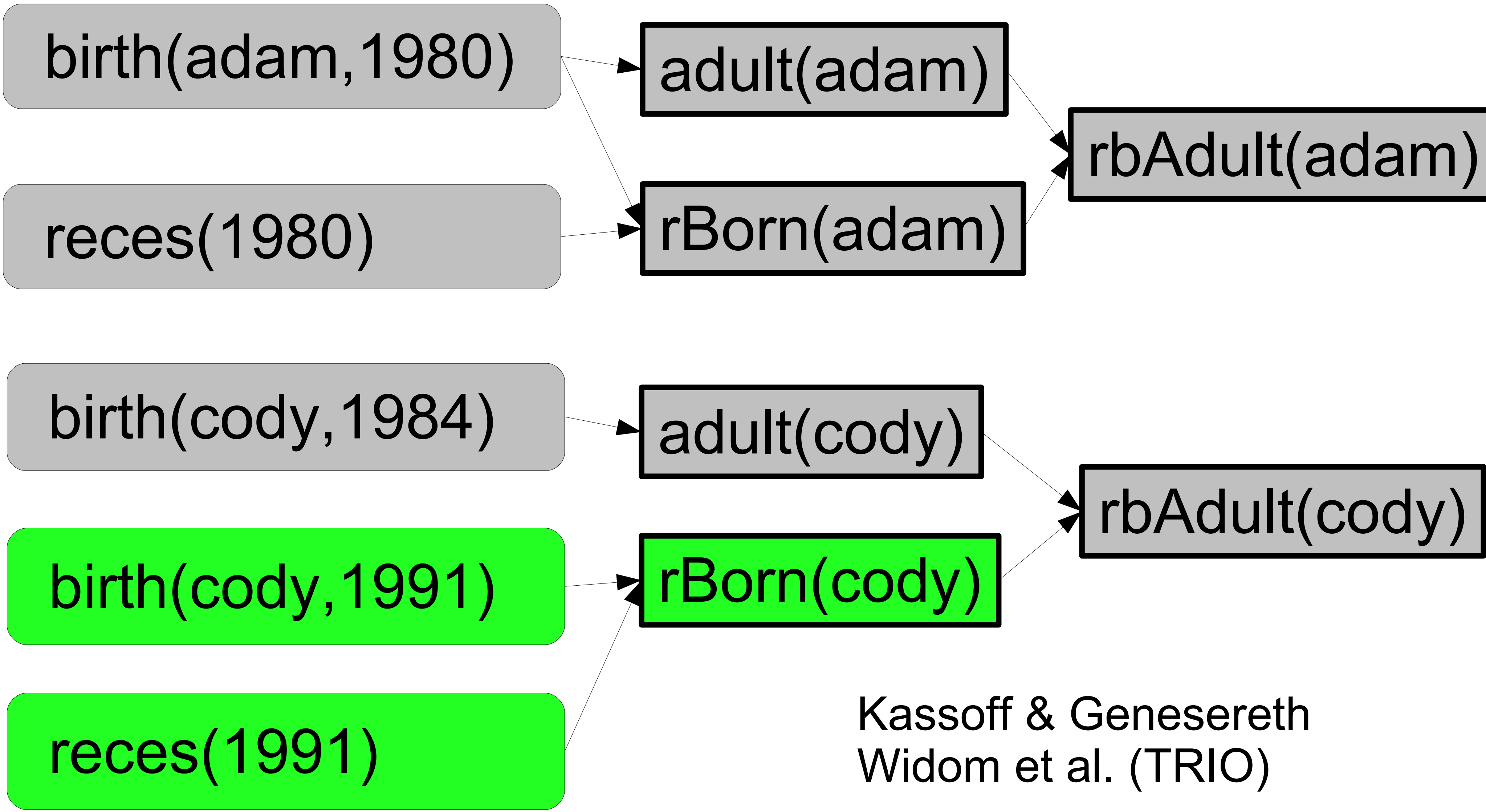
Lineage Approach



Kassoff & Genesereth
Widom et al. (TRIO)

Eric Kao
Stanford University

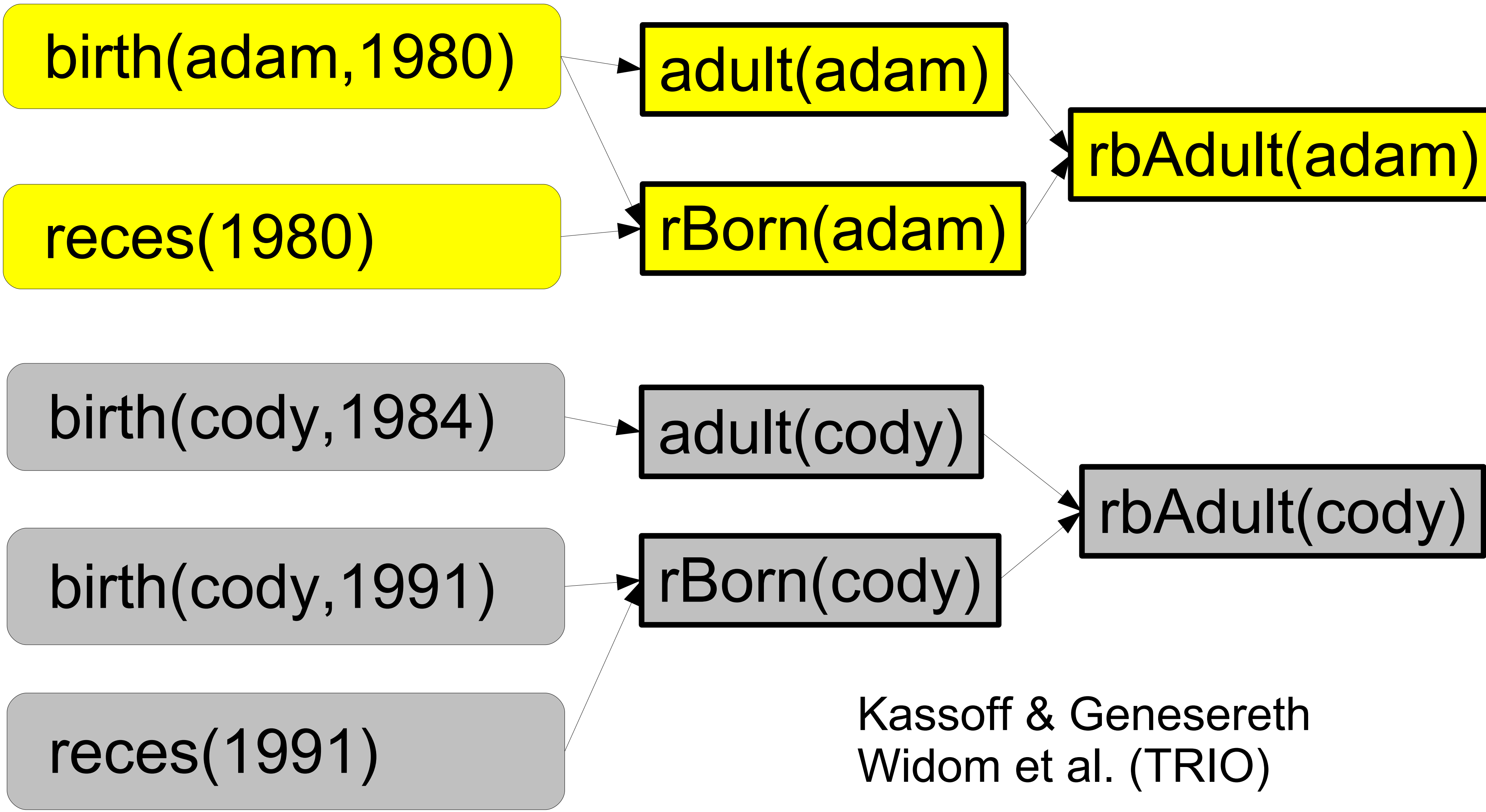
Lineage Approach



Kassoff & Genesereth
Widom et al. (TRIO)

Eric Kao
Stanford University

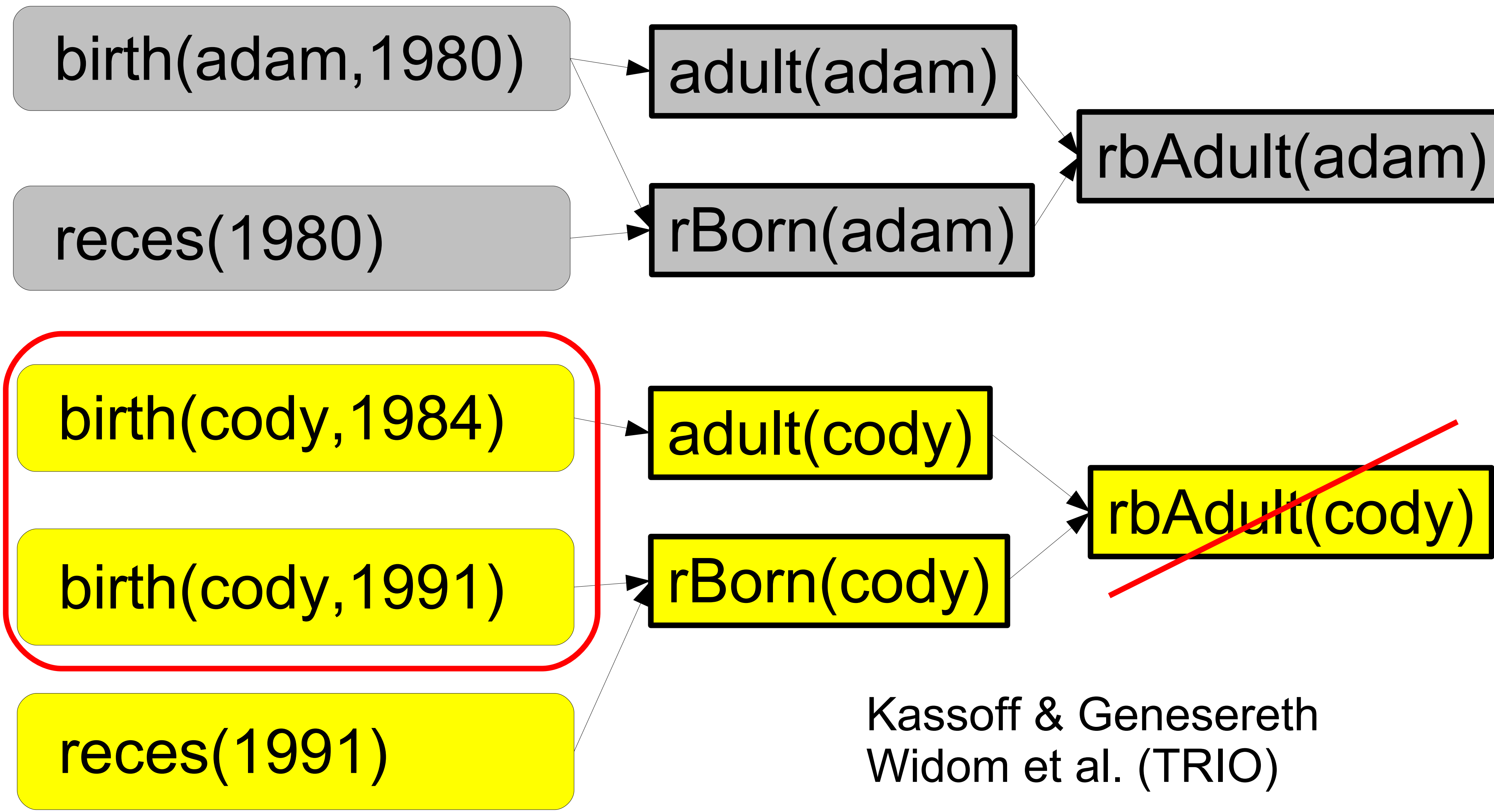
Lineage Approach



Kassoff & Genesereth
Widom et al. (TRIO)

Eric Kao
Stanford University

Lineage Approach



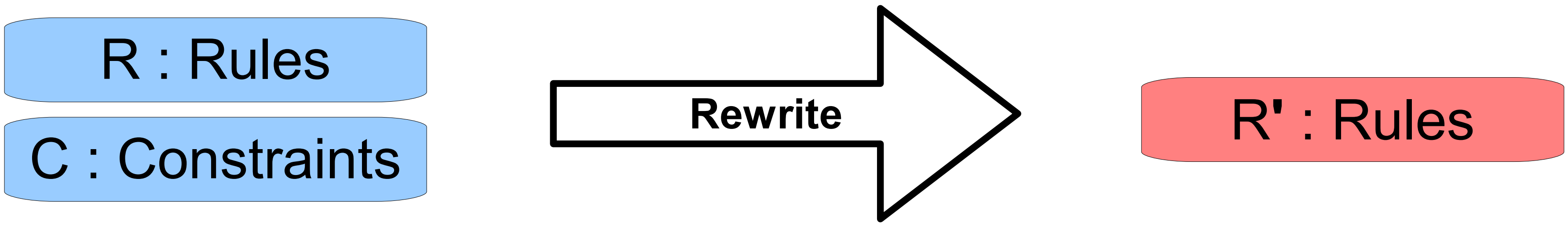
Kassoff & Genesereth
Widom et al. (TRIO)

Features

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

Eric Kao
Stanford University

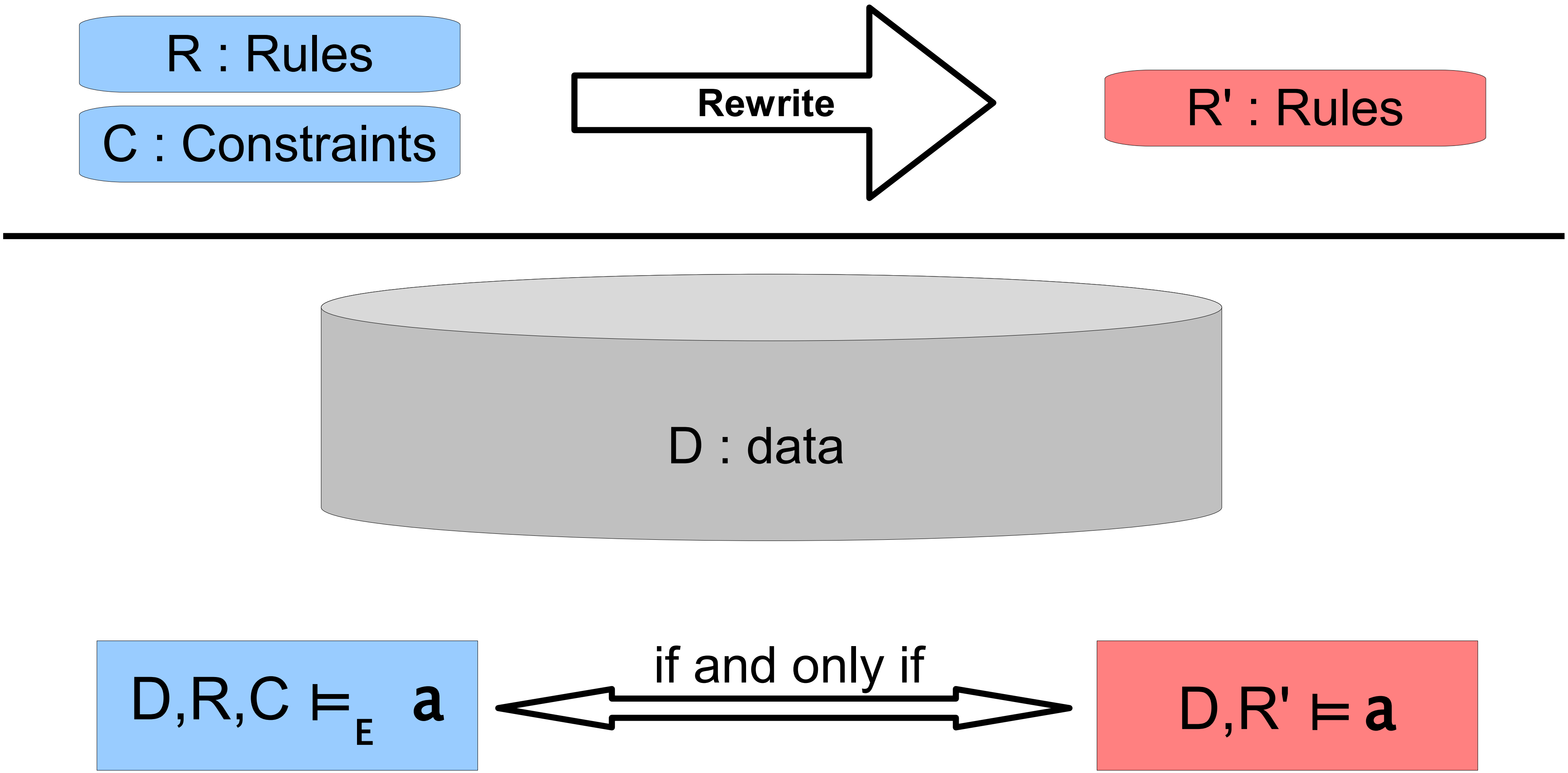
Rule-rewriting approach



Hinrichs, Kao & Genesereth

Eric Kao
Stanford University

Rule-rewriting approach



Hinrichs, Kao & Genesereth

Eric Kao
Stanford University

Constraint: :- birth(x,y) & birth(x,z) & y≠z

- adult(x) :- birth(x,z) & z < 1990
  rBorn(x) :- birth(x,y) & reces(y)
  rbAdult(x) :- rBorn(x) & adult(x)

Eric Kao
Stanford University

Constraint: :- birth(x,y) & birth(x,z) & y≠z

- adult(x) :- birth(x,z) & z < 1990
  rBorn(x) :- birth(x,y) & reces(y)
  rbAdult(x) :- rBorn(x) & adult(x)

unroll

- rbAdult(x) :- birth(x,z) & z < 1990
  & birth(x,y) & reces(y)

Constraint: :- birth(x,y) & birth(x,z) & y≠z

- adult(x) :- birth(x,z) & z < 1990
  rBorn(x) :- birth(x,y) & reces(y)
  rbAdult(x) :- rBorn(x) & adult(x)

flatten

- rbAdult(x) :- birth(x,z) & z < 1990
  & birth(x,y) & reces(y)

rewrite

- rbAdult(x) :- birth(x,z) & z < 1990
  & birth(x,y) & reces(y)
  & y=z

Eric Kao
Stanford University

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 ⇔ X = 2 & Y = Z
- Augmented rule:

```
r(Z) :- p(X,Y) & q(Z,1)& ¬[X = 2 & Y = Z]
```


Eric Kao
Stanford University

Features

- Polynomial data complexity
- Rewrite and send
- Reuse on different/evolving data
- Limitation:

  Does not work for recursive rules

Eric Kao
Stanford University

Recursive Rules

- R:

  reach(X,Y) :- link(X,Z,T) & reach(Z,Y)
  reach(X,Y) :- link(X,Y,T)

- C:

  :- p(X,Y,T) & p(X',Y',T) & X≠X' & Y≠Y'

- Appears to be no Datalog rewriting
- Theorem:

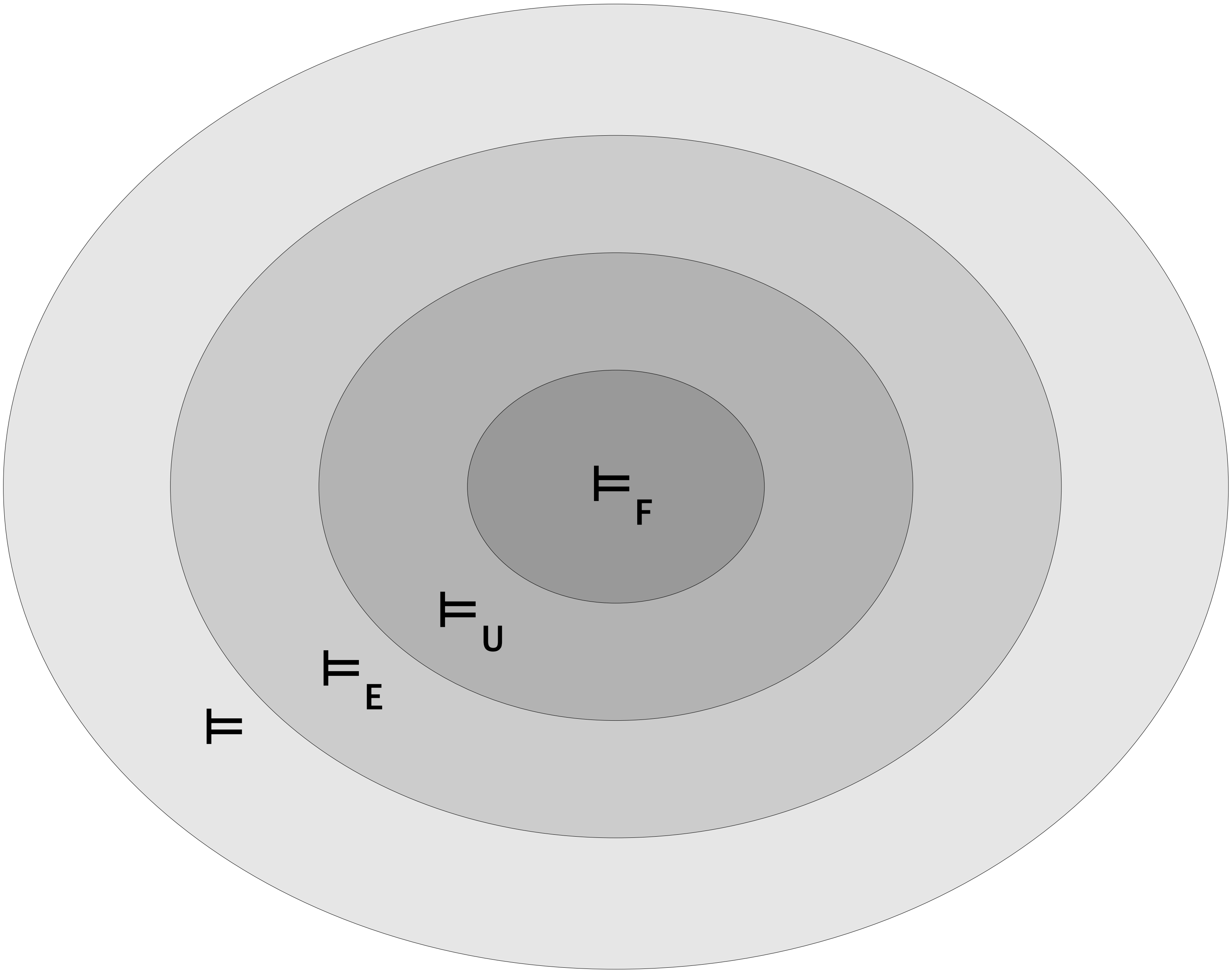
  Unless P = NP, some recursive rule sets do not have existential answers rewriting in Datalog

Eric Kao
Stanford University

Other Answer Semantics

- Con(D,R,C) := {D* ⊆ D | D*,R consistent w C}
- MaxCon(D,R,C) := Maximal sets in Con(D,R,C)
- Free(D,R,C) := intersection of MaxCon(D,R,C)
- D,R,C $\models_U$ a
  $\Leftrightarrow$
  for all D* in MaxCon(D,R,C),
  D*, R $\models$ a
- D,R,C $\models_F$ a
  $\Leftrightarrow$
  Free(D,R,C), R $\models$ a

Eric Kao
Stanford University



*Assume Con(D,R,C) non-empty

Eric Kao
Stanford University

Summary

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

Eric Kao
Stanford University