From Classical to Consistent Query Answering under Existential Rules

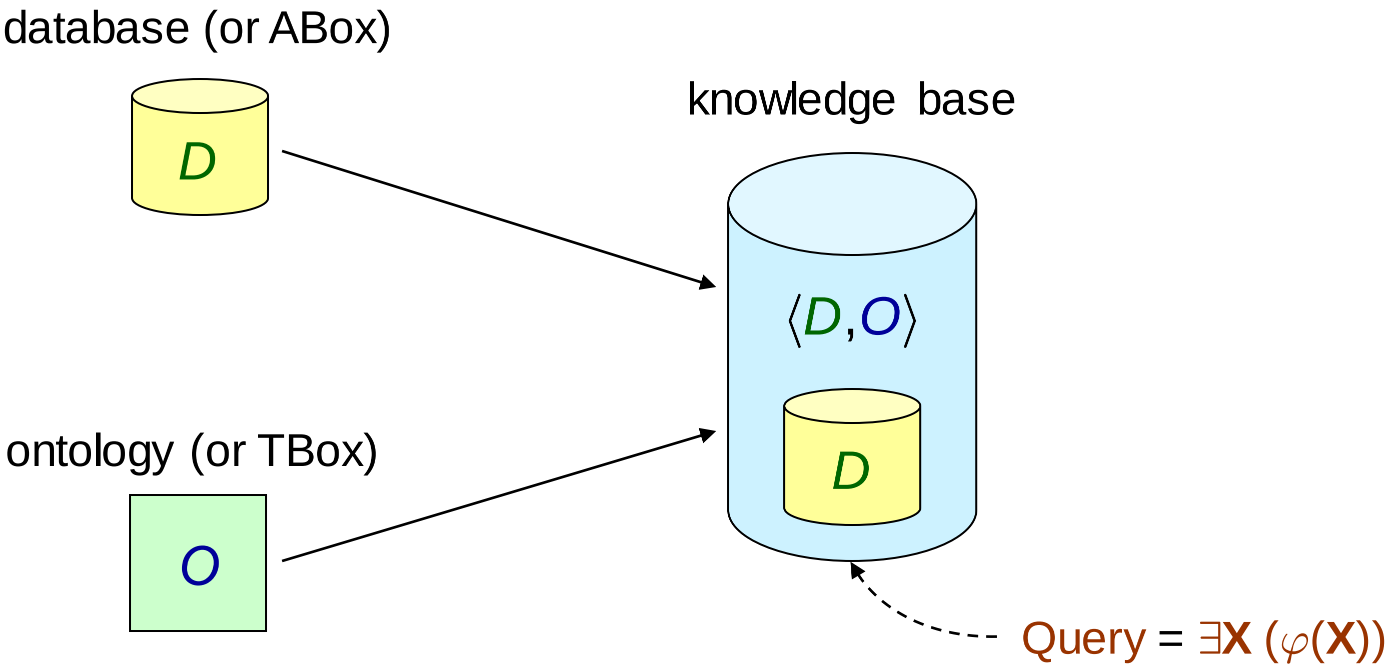
Andreas Pieris

Institute of Information Systems, Vienna University of Technology, Austria

Joint work with Thomas Lukasiewicz, Maria Vanina Martinez and Gerardo I. Simari

OntoLP, Argentina, Buenos Aires, July 25, 2015

Ontology-based Query Answering (OBQA)



$$\langle D, O \rangle \models \text{Query} \quad \Leftrightarrow \quad D \wedge O \models \text{Query}$$

A Simple Example

D =

professor(John)

fellow(John)

O =

$\forall X\ (\mathit{professor}(X) \rightarrow \exists Y\ (\mathit{faculty}(X) \wedge \mathit{teaches}(X,Y)))$

$\forall X\ (\mathit{fellow}(X) \rightarrow \mathit{faculty}(X))$

$\forall M \models \langle D, O \rangle\ :\ M$ = **...** *teaches*(John,#) **...**

{John $\rightarrow$ John, $X \rightarrow$ #}

$\exists X\ (\mathit{teaches}(\text{John}, X))$ ✓

A Simple Example

D =

professor(John)

fellow(John)

O =

$\forall X\ (professor(X) \rightarrow \exists Y\ (faculty(X) \wedge teaches(X,Y)))$

$\forall X\ (fellow(X) \rightarrow faculty(X))$

$\forall X\ (professor(X) \wedge fellow(X) \rightarrow \bot)$

no model $\Rightarrow$ every query is entailed

Handling Data Inconsistencies

- The data are likely to be inconsistent with the ontology

- Standard semantics fails: everything is inferred - not meaningful answers

- Two approaches to inconsistency-handling:
  - Resolve the inconsistencies - ideal, but not always possible
  - Live with the inconsistencies - inconsistency-tolerant semantics

ABox Repair (AR) Semantics

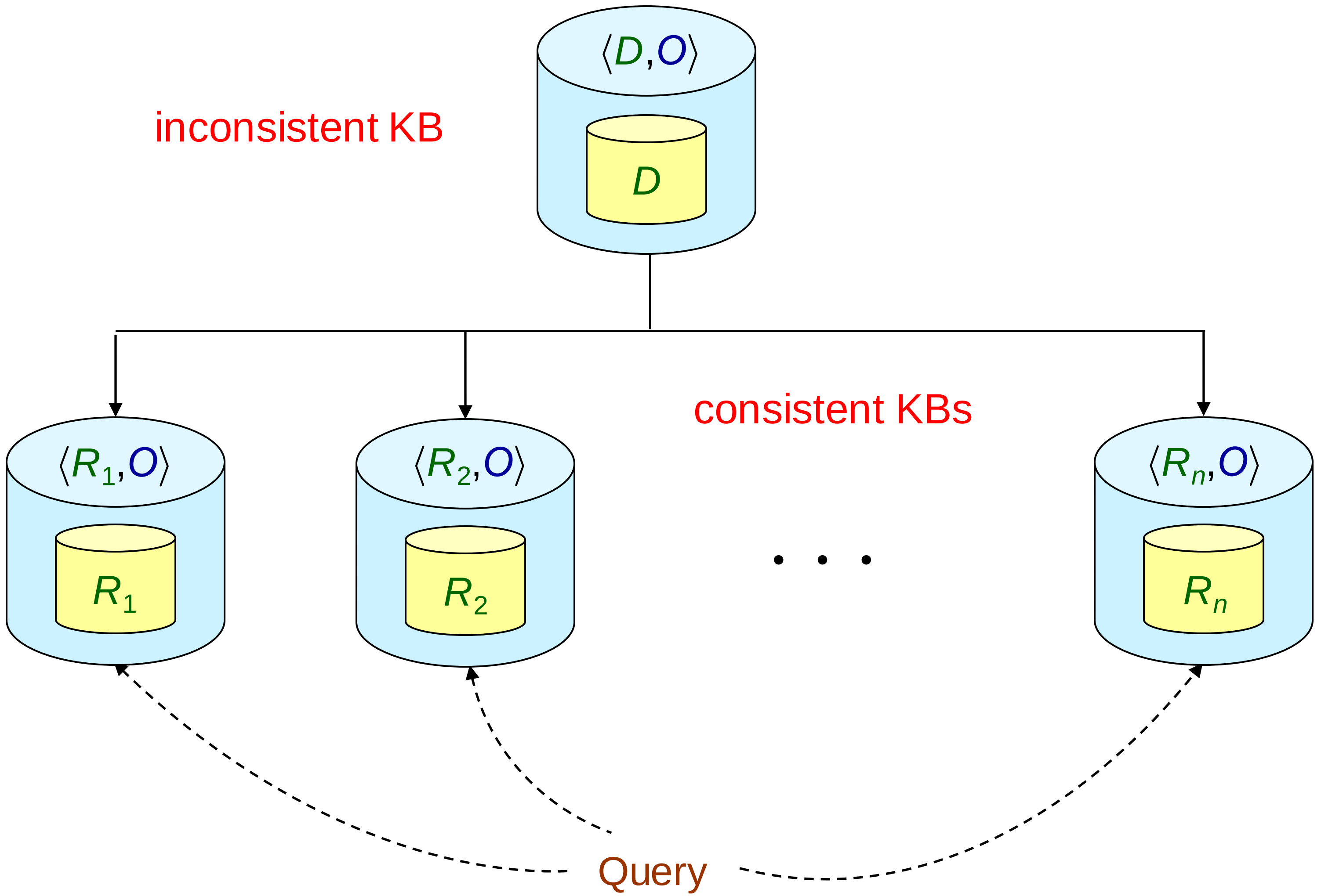
- Standard inconsistency-tolerant semantics

- **IDEA:** The query must be entailed by every database repair

$\subseteq$-maximal consistent subsets of the database

[Lembo et al., RR 2010]

ABox Repair (AR) Semantics



$$\langle D,O \rangle \models_{AR} \text{Query} \quad \Leftrightarrow \quad \forall R \in \{R_1,\ldots,R_n\}: \langle R,O \rangle \models \text{Query}$$

ABox Repair (AR) Semantics: Example

D =

professor(John)

fellow(John)

O =

$\forall X\, (professor(X) \rightarrow \exists Y\, (faculty(X) \wedge teaches(X,Y)))$

$\forall X\, (fellow(X) \rightarrow faculty(X))$

$\forall X\, (professor(X) \wedge fellow(X) \rightarrow \bot)$

$\langle D, O \rangle \models_{\mathrm{AR}} faculty(\mathrm{John})$ ✓

R_1 = *professor*(John)

$\langle R_1, O \rangle \models faculty(\mathrm{John})$ ✓

R_2 = *fellow*(John)

$\langle R_2, O \rangle \models faculty(\mathrm{John})$ ✓

ABox Repair (AR) Semantics: Example

D =

professor(John)

fellow(John)

O =

$\forall X\ (\mathit{professor}(X) \rightarrow \exists Y\ (\mathit{faculty}(X) \wedge \mathit{teaches}(X,Y)))$

$\forall X\ (\mathit{fellow}(X) \rightarrow \mathit{faculty}(X))$

$\forall X\ (\mathit{professor}(X) \wedge \mathit{fellow}(X) \rightarrow \bot)$

$\langle D,O \rangle \vDash_{\mathrm{AR}} \exists X\ (\mathit{teaches}(\mathrm{John},X))$ ✗

R_1 = *professor*(John)

$\langle R_1,O \rangle \vDash \exists X\ (\mathit{teaches}(\mathrm{John},X))$ ✓

R_2 = *fellow*(John)

$\langle R_2,O \rangle \vDash \exists X\ (\mathit{teaches}(\mathrm{John},X))$ ✗

AR Semantics

- Lots of recent work and complexity results for description logics

  [Lembo et al., RR 2010 / Rosati, IJCAI 2011 / Bienvenu, AAAI 2012 / Bienvenu & Rosati, IJCAI 2013]

- This talk is about existential rules + negative constraints

  [Lukasiewicz, Martinez & Simari, ODBASE 2013 / Lukasiewicz, Martinez, P. & Simari, AAAI 2015]

$$\forall \mathbf{X}\ (\varphi(\mathbf{X}) \rightarrow \exists \mathbf{Y}\ (\psi(\mathbf{X},\mathbf{Y}))) \quad + \quad \forall \mathbf{X}\ (\varphi(\mathbf{X}) \rightarrow \bot)$$

Our Goal

Perform an in-depth $\boxed{\text{complexity}}$ analysis of consistent query answering under the

main classes of existential rules + negative constraints

- Combined
- Bounded-arity combined
- Fixed-program combined
- Data

generic complexity results - from classical to consistent query answering

Combined Complexity

combined or ba-combined or fp-combined

class of $\exists$-rules

complexity class

$\mathbb{M}$ complexity of classical query answering under $\mathbb{L}$ is $\mathbb{C}$-complete

$$\Downarrow$$

$\mathbb{M}$ complexity of consistent query answering under $\mathbb{L}[\perp]$ is:

$\Pi_{P,2}$-complete if $\mathbb{C}$ = NP

$\mathbb{C}$-complete if $\mathbb{C} \supseteq$ PSPACE & $\mathbb{C}$ is deterministic

Combined Complexity: Upper Bounds

Guess and check algorithm (for the complement of the problem)

Input: D, $O \in \mathbb{L}[\perp]$, Q

1. Guess $R \subseteq D$ - a possible repair
2. Verify that R is a repair, i.e., $\langle R,O \rangle$ is consistent and R is $\subseteq$-maximal
3. Verify that $\langle R,O \rangle$ does not entail Q

no harder than classical query answering under $\mathbb{L}$

$$\Rightarrow \text{ our problem is in } \mathrm{coNP}^{\mathbb{C}} \Rightarrow \text{ in } \begin{cases} \mathrm{coNP}^{\mathrm{NP}} = \mathrm{co}\Sigma_{P,2} = \Pi_{P,2} & \text{if } \mathbb{C} = \mathrm{NP} \\ \mathrm{coNP}^{\mathbb{C}} = \mathrm{co}\mathbb{C} = \mathbb{C} & \text{if } \mathbb{C} \supseteq \mathrm{PSPACE},\ \mathbb{C} \text{ is deterministic} \end{cases}$$

Combined Complexity

combined or ba-combined or fp-combined   class of $\exists$-rules   complexity class

$\mathbb{M}$ complexity of classical query answering under $\mathbb{L}$ is $\mathbb{C}$-complete

$$\Downarrow$$

$\mathbb{M}$ complexity of consistent query answering under $\mathbb{L}[\perp]$ is:

$\Pi_{P,2}$-complete   if $\mathbb{C} = \mathrm{NP}$

$\mathbb{C}$-complete   if $\mathbb{C} \supseteq \mathrm{PSPACE}$ & $\mathbb{C}$ is deterministic

A Strong $\Pi_{P,2}$-hardness Result

Consistent query answering under the single constraint

$$\forall X \forall Y \forall Z \forall W \ (p(X,Y,Z) \wedge p(W,X,Z) \rightarrow \perp)$$

while the database and the query use only binary and ternary predicates

(by reduction from satisfiability of 2QBF formulas)

$$\Downarrow$$

For every class $\mathbb{L}$ of existential rules, the fp-combined complexity of consistent query answering under $\mathbb{L}[\perp]$ is $\Pi_{P,2}$-hard

Combined Complexity

combined or ba-combined or fp-combined

class of $\exists$-rules

complexity class

$\mathbb{M}$ complexity of classical query answering under $\mathbb{L}$ is $\mathbb{C}$-complete

$$\Downarrow$$

$\mathbb{M}$ complexity of consistent query answering under $\mathbb{L}[\perp]$ is:

$\Pi_{P,2}$-complete if $\mathbb{C} = \mathrm{NP}$

$\mathbb{C}$-complete if $\mathbb{C} \supseteq \mathrm{PSPACE}$ & $\mathbb{C}$ is deterministic

Data Complexity

class of ∃-rules

complexity class

data complexity of classical query answering under $\mathbb{L}$ is $\mathbb{C}$-complete

$$\Downarrow$$

data complexity of consistent query answering under $\mathbb{L}[\perp]$ is:

coNP-complete if $\mathbb{C} \subseteq$ PTIME

Data Complexity: Upper Bounds

Guess and check algorithm (for the complement of the problem)

Input: D, $O \in \mathbb{L}[\perp]$, Q

1. Guess $R \subseteq D$ - a possible repair

2. Verify that R is a repair, i.e., $\langle R, O \rangle$ is consistent and R is $\subseteq$-maximal

3. Verify that $\langle R, O \rangle$ does not entail Q

no harder than classical query answering under $\mathbb{L}$

$\Rightarrow$ our problem is in $\mathrm{coNP}^{\mathbb{C}}$ $\Rightarrow$ in coNP (since $\mathrm{NP}^{\mathrm{PTIME}} = \mathrm{NP}$)

A Strong coNP-hardness Result

Consistent query answering under the single constraint

$$\forall X \;\; (p(X) \wedge s(X) \; \rightarrow \; \bot)$$

while the query is fixed

(by reduction from 2+2UNSAT)

$$\Downarrow$$

For every class $\mathbb{L}$ of existential rules, the data complexity of consistent query answering under $\mathbb{L}[\bot]$ is coNP-hard

Data Complexity

class of $\exists$-rules complexity class

data complexity of classical query answering under $\mathbb{L}$ is $\mathbb{C}$-complete

$$\Downarrow$$

data complexity of consistent query answering under $\mathbb{L}[\perp]$ is:

coNP-complete if $\mathbb{C} \subseteq$ PTIME

From Classical to Consistent Query Answering

(ba-/fp)combined complexity:

in NP	$\rightarrow$	$\Pi_{P,2}$-complete
$\mathbb{C}$-complete, $\mathbb{C} \supseteq$ PSPACE & $\mathbb{C}$ is deterministic	$\rightarrow$	$\mathbb{C}$-complete

data complexity:

in $\mathbb{C} \subseteq$ PTIME	$\rightarrow$	coNP-complete

an (almost) complete picture for the main classes of existential rules + negative constraints

Existential Rules

$$\forall \mathbf{X}\ (\varphi(\mathbf{X}) \rightarrow \exists \mathbf{Y}\ (\psi(\mathbf{X},\mathbf{Y})))$$

conjunctions of atoms

- Classical query answering under existential rules is undecidable
  see, e.g., [Beeri & Vardi, ICALP 1981]

- Expressive decidable fragments - field of intense research

  (e.g., Montpellier, Dresden, Calabria, Oxford, Vienna, …)

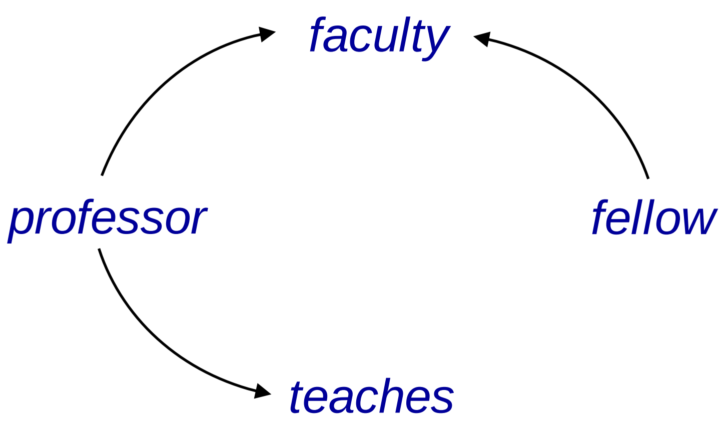
- Main decidability paradigms: acyclicity, guardedness & stickiness

Acyclic Existential Rules

- The predicate graph is acyclic

$$\forall X\,(professor(X) \;\rightarrow\; \exists Y\,(faculty(X) \wedge teaches(X,Y)))$$

$$\forall X\,(fellow(X) \;\rightarrow\; faculty(X))$$



(Frontier-)Guarded Existential Rules

- Frontier-guardedness: There exists a body-atom that contains the frontier

$$\forall X \forall Y \forall Z\, (\boldsymbol{supervisorOf(X,Y)} \wedge supervisorOf(Y,Z) \rightarrow manager(X))$$

- Guardedness: There exists a body-atom that contains all the $\forall$-variables

$$\forall X \forall Y\ (\boldsymbol{supervisorOf(X,Y)} \wedge emp(Y) \rightarrow emp\,(X))$$

- Linearity: There exists only one atom in the body

$$\forall X\ (\boldsymbol{employee(X)}\ \rightarrow\ \exists Y\ (supervisorOf(Y,X) \wedge employee(Y)))$$

Sticky Existential Rules

- Join-variables stick to the inferred atoms

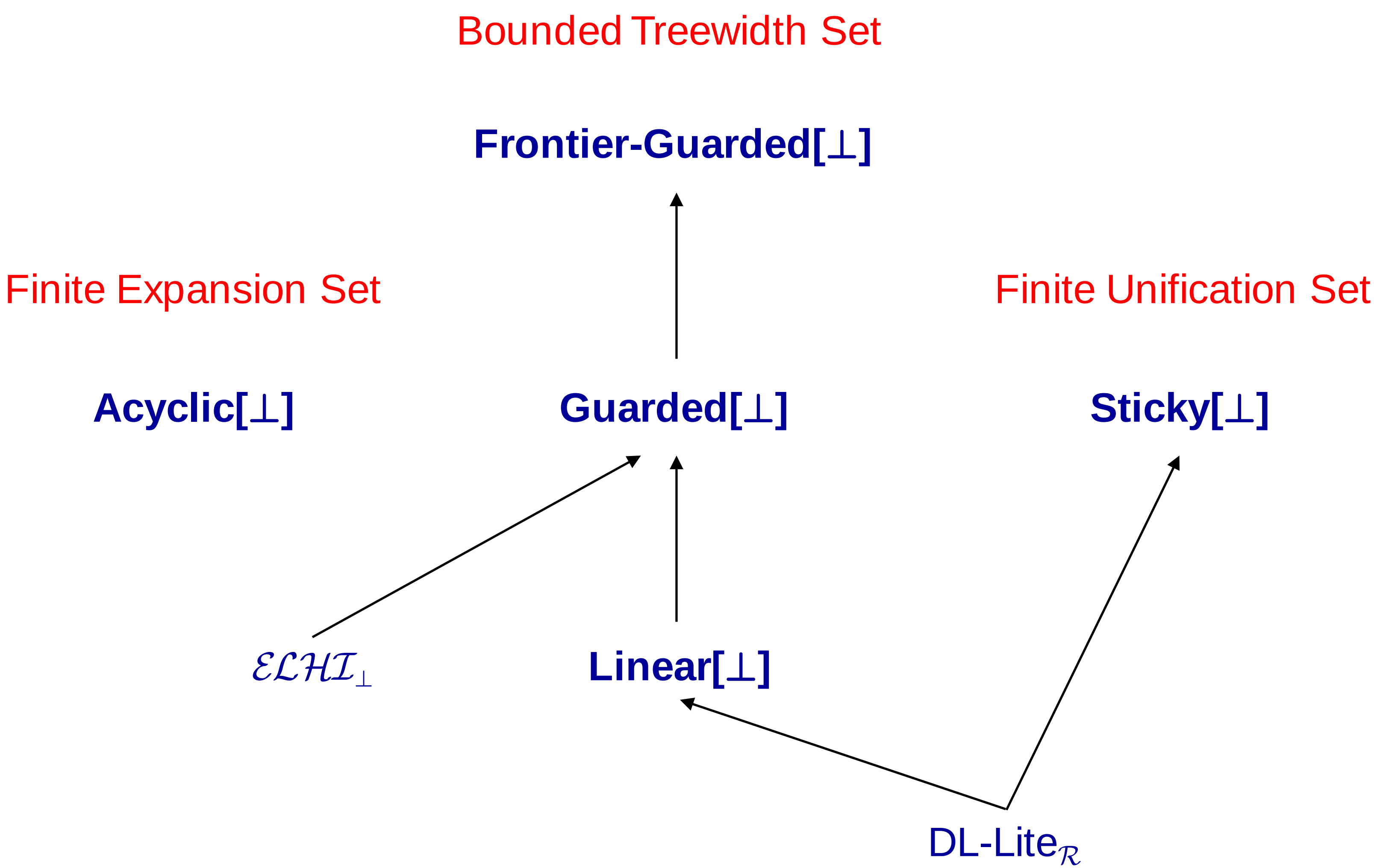
$$\forall X \forall Y \forall Z\, (q(X,Y) \wedge p(Y,Z) \rightarrow \exists W\, (t(X,Y,W)))$$

$$\forall X \forall Y \forall Z\, (t(X,Y,Z) \rightarrow \exists W\, (s(Y,W)))$$

$$\forall X \forall Y \forall Z\, (q(X,Y) \wedge p(Y,Z) \rightarrow \exists W\, (t(X,Y,W)))$$

$$\forall X \forall Y \forall Z\, (t(X,Y,Z) \rightarrow \exists W\, (s(X,W)))$$

Existential Rules + Negative Constraints



From Classical to Consistent Query Answering

(ba-/fp)combined complexity:

in NP $\rightarrow$ $\Pi_{p,2}$-complete

$\mathbb{C}$-complete, $\mathbb{C} \supseteq$ PSPACE & $\mathbb{C}$ is deterministic $\rightarrow$ $\mathbb{C}$-complete

data complexity:

in $\mathbb{C} \subseteq$ PTIME $\rightarrow$ coNP-complete

we simply need to exploit existing results on classical query answering

Classical Query Answering

	Combined	ba-combined	fp-combined	Data
Acyclic[⊥]	NEXPTIME	NEXPTIME	NP	in AC_0
Frontier-Guarded[⊥]	2EXPTIME	2EXPTIME	NP	PTIME
Guarded[⊥]	2EXPTIME	EXPTIME	NP	PTIME
Linear[⊥]	PSPACE	NP	NP	in AC_0
Sticky[⊥]	EXPTIME	NP	NP	in AC_0

Classical Query Answering

	Combined	ba-combined	fp-combined	Data
Acyclic[⊥]	NEXPTIME	NEXPTIME	NP	in AC_0
Frontier-Guarded[⊥]	2EXPTIME	2EXPTIME	NP	PTIME
Guarded[⊥]	2EXPTIME	EXPTIME	NP	PTIME
Linear[⊥]	PSPACE	NP	NP	in AC_0
Sticky[⊥]	EXPTIME	NP	NP	in AC_0

- Until recently, it was generally believed that it is EXPTIME
- The obvious algorithm does not work - models of double-exponential size

Classical Query Answering

	Combined	ba-combined	fp-combined	Data
Acyclic[⊥]	NEXPTIME	NEXPTIME	NP	in AC_0
Frontier-Guarded[⊥]	2EXPTIME	2EXPTIME	NP	PTIME
Guarded[⊥]	2EXPTIME	EXPTIME	NP	PTIME
Linear[⊥]	PSPACE	NP	NP	in AC_0
Sticky[⊥]	EXPTIME	NP	NP	in AC_0

- Upper bound: non-deterministically construct a proof of the query
- Lower bound: by reduction from a TILING problem

Classical Query Answering

	Combined	ba-combined	fp-combined	Data
Acyclic[⊥]	NEXPTIME	NEXPTIME	NP	in AC_0
Frontier-Guarded[⊥]	2EXPTIME	2EXPTIME	NP	PTIME
Guarded[⊥]	2EXPTIME	EXPTIME	NP	PTIME
Linear[⊥]	PSPACE	NP	NP	in AC_0
Sticky[⊥]	EXPTIME	NP	NP	in AC_0

(ba-/fp)combined complexity:

in NP $\rightarrow$ $\Pi_{P,2}$-complete

$\mathbb{C}$-complete, $\mathbb{C} \supseteq$ PSPACE & $\mathbb{C}$ is deterministic $\rightarrow$ $\mathbb{C}$-complete

data complexity:

in $\mathbb{C} \subseteq$ PTIME $\rightarrow$ coNP-complete

Consistent Query Answering

	Combined	ba-combined	fp-combined	Data
Acyclic[$\perp$]	?	?	$\Pi_{P,2}$	coNP
Frontier-Guarded[$\perp$]	2EXPTIME	2EXPTIME	$\Pi_{P,2}$	coNP
Guarded[$\perp$]	2EXPTIME	EXPTIME	$\Pi_{P,2}$	coNP
Linear[$\perp$]	PSPACE	$\Pi_{P,2}$	$\Pi_{P,2}$	coNP
Sticky[$\perp$]	EXPTIME	$\Pi_{P,2}$	$\Pi_{P,2}$	coNP

(ba-/fp)combined complexity:

in NP $\rightarrow$ $\Pi_{P,2}$-complete

$\mathbb{C}$-complete, $\mathbb{C} \supseteq$ PSPACE & $\mathbb{C}$ is deterministic $\rightarrow$ $\mathbb{C}$-complete

data complexity:

in $\mathbb{C} \subseteq$ PTIME $\rightarrow$ coNP-complete

Complexity of Acyclic[$\perp$]

- The guess and check algorithm gives a $\mathrm{coNP}^{\mathrm{NEXPTIME}}$ upper bound

- The class $\mathrm{NP}^{\mathrm{NEXPTIME}}$ lies at a higher level of the strong exponential hierarchy

- The SEH collapses to its Δ_2 level $\Rightarrow$ $\mathrm{NP}^{\mathrm{NEXPTIME}} = \mathrm{P}^{\mathrm{NE}}$
  [Hemachandra, J. Comput. Syst. Sci. 1989]

- P^{NE} is a deterministic class $\Rightarrow$ $\mathrm{coP}^{\mathrm{NE}} = \mathrm{P}^{\mathrm{NE}}$

Consistent Query Answering

	Combined	ba-combined	fp-combined	Data
Acyclic[⊥]	NEXP - P^{NE}	NEXP - P^{NE}	$\Pi_{P,2}$	coNP
Frontier-Guarded[⊥]	2EXPTIME	2EXPTIME	$\Pi_{P,2}$	coNP
Guarded[⊥]	2EXPTIME	EXPTIME	$\Pi_{P,2}$	coNP
Linear[⊥]	PSPACE	$\Pi_{P,2}$	$\Pi_{P,2}$	coNP
Sticky[⊥]	EXPTIME	$\Pi_{P,2}$	$\Pi_{P,2}$	coNP

$$P^{NE} \subseteq \text{coNEXPTIME}^{NP}$$

[Hemachandra, J. Comput. Syst. Sci. 1989]

Consistent Query Answering

	Combined	ba-combined	fp-combined	Data
Acyclic[⊥]	NEXP - P^{NE}	NEXP - P^{NE}	$\Pi_{P,2}$	coNP
Frontier-Guarded[⊥]	2EXPTIME	2EXPTIME	$\Pi_{P,2}$	coNP
Guarded[⊥]	2EXPTIME	EXPTIME	$\Pi_{P,2}$	coNP
Linear[⊥]	PSPACE	$\Pi_{P,2}$	$\Pi_{P,2}$	coNP
Sticky[⊥]	EXPTIME	$\Pi_{P,2}$	$\Pi_{P,2}$	coNP

Conjecture: Consistent query answering under Acyclic[⊥] is coNEXPTIME^{NP}-c

Data Intractable

	Combined	ba-combined	fp-combined	Data
Acyclic[⊥]	NEXP - P^{NE}	NEXP - P^{NE}	$\Pi_{P,2}$	coNP
Frontier-Guarded[⊥]	2EXPTIME	2EXPTIME	$\Pi_{P,2}$	coNP
Guarded[⊥]	2EXPTIME	EXPTIME	$\Pi_{P,2}$	coNP
Linear[⊥]	PSPACE	$\Pi_{P,2}$	$\Pi_{P,2}$	coNP
Sticky[⊥]	EXPTIME	$\Pi_{P,2}$	$\Pi_{P,2}$	coNP

but, what about tractability results w.r.t. the data complexity?

…consider approximations of the AR semantics

Intersection ABox Repair (IAR) Semantics

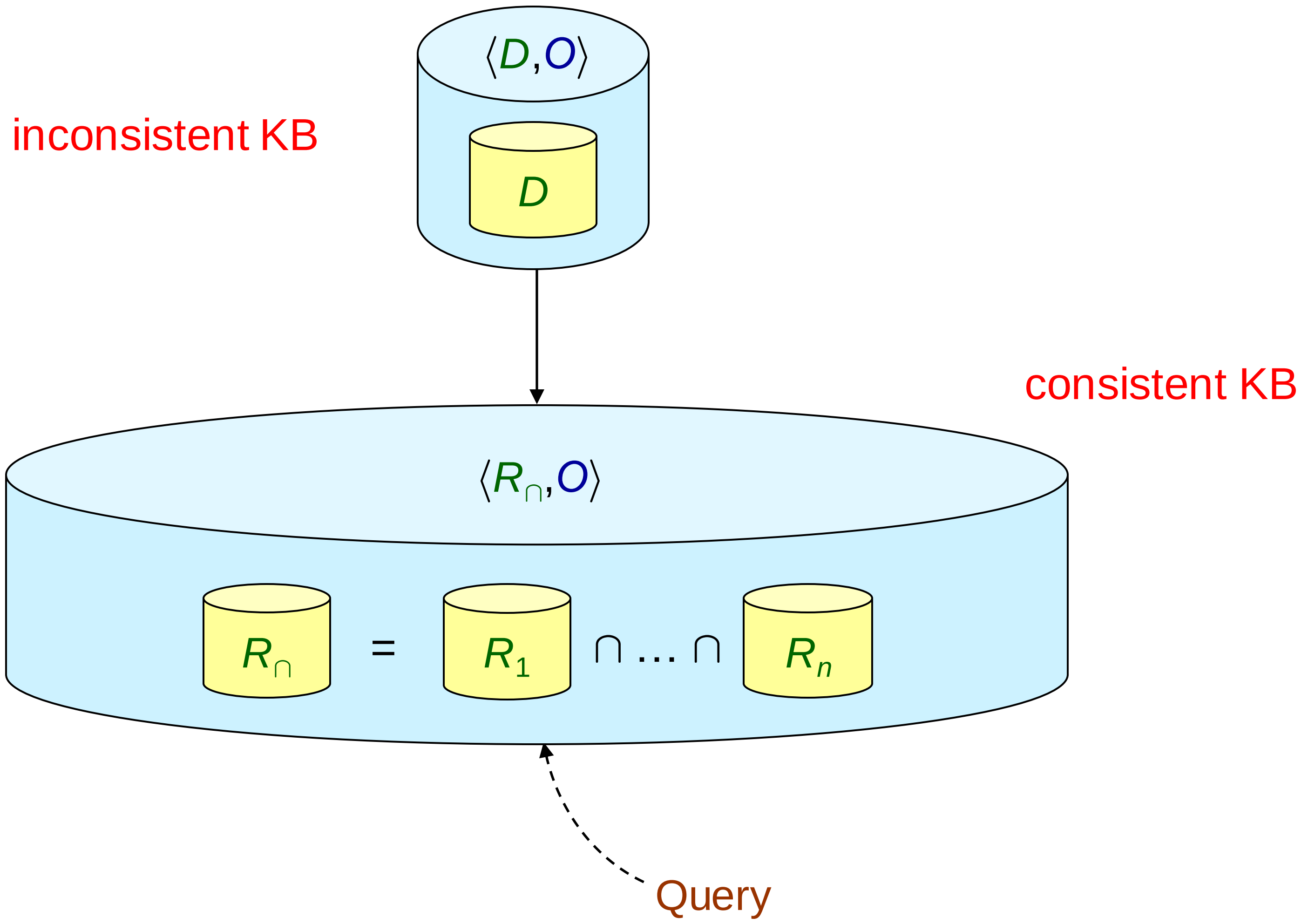
- One of the basic sound approximations of the AR semantics

- **IDEA:** The query must be entailed by the intersection of the database repairs

$\subseteq$-maximal consistent subsets of the database

[Lembo et al., RR 2010]

Intersection ABox Repair (IAR) Semantics



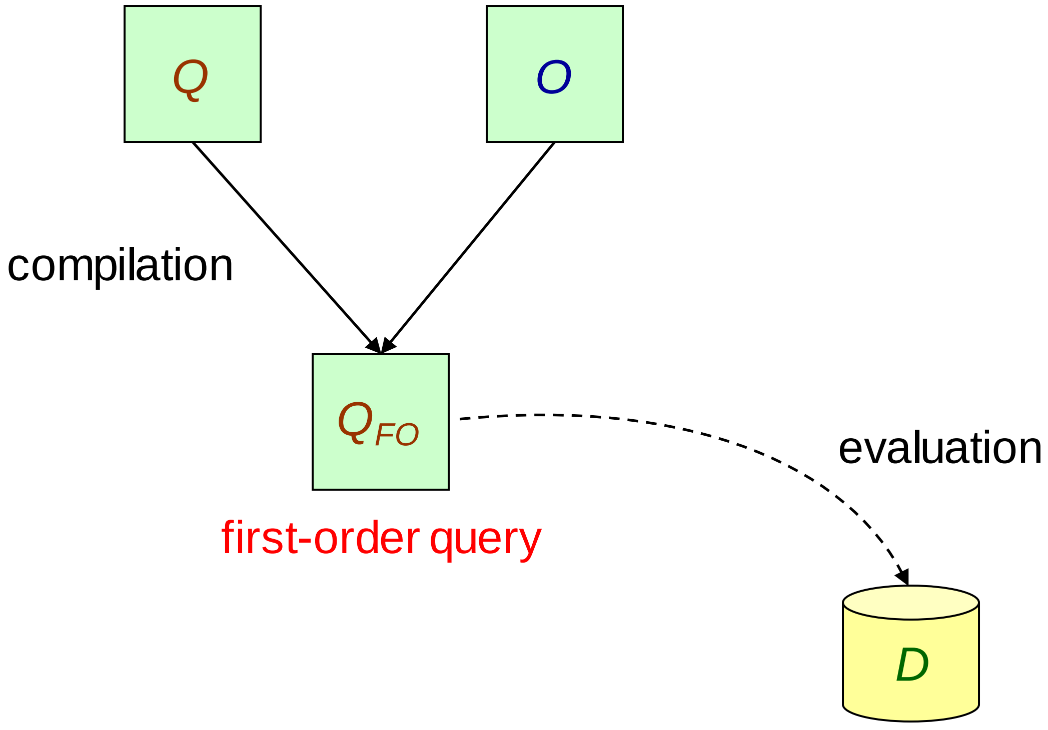
$$\langle D, O\rangle \models_{\mathrm{IAR}} \text{Query} \quad \Leftrightarrow \quad \langle R_\cap, O\rangle \models \text{Query}$$

Data Complexity under the IAR Semantics

Acyclic[⊥]	in AC_0
Frontier-Guarded[⊥]	coNP
Guarded[⊥]	coNP
Linear[⊥]	in AC_0
Sticky[⊥]	in AC_0

via first-order rewritability - a generic result can be established

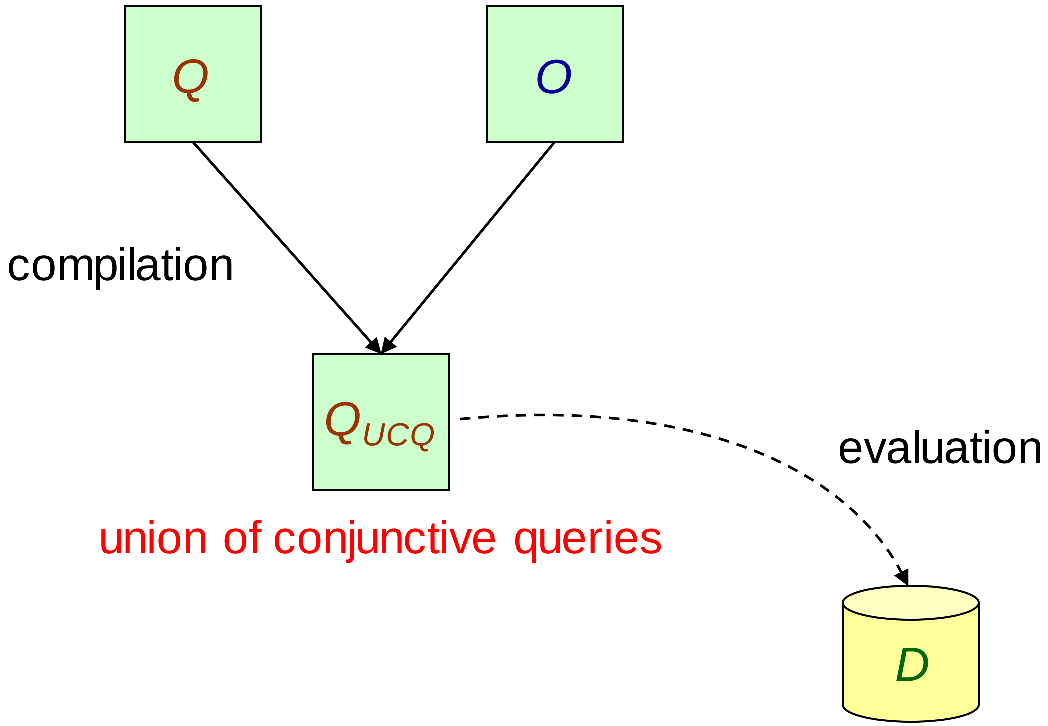
First-Order Rewritability (FO-Rewritability)



$$\forall D \;:\; \langle D, O \rangle \models Q \quad \Leftrightarrow \quad D \models Q_{FO}$$

$$\forall D \;:\; \langle D, O \rangle \models_{\mathrm{IAR}} Q \quad \Leftrightarrow \quad D \models Q_{FO}$$

UCQ-Rewritability



$$\forall D : \langle D, O \rangle \models Q \quad \Leftrightarrow \quad D \models Q_{UCQ}$$

$$\forall D : \langle D, O \rangle \models_{\mathrm{IAR}} Q \quad \Leftrightarrow \quad D \models Q_{UCQ}$$

From UCQ-Rewritability to FO-Rewritability

class of $\exists$-rules

classical query answering under $\mathbb{L}$ is UCQ-Rewritable

$\Downarrow$

consistent query answering under the IAR semantics for $\mathbb{L}[\perp]$ is FO-Rewritable

Data Complexity under the IAR Semantics

Acyclic[⊥]	in AC_0
Frontier-Guarded[⊥]	coNP
Guarded[⊥]	coNP
Linear[⊥]	in AC_0
Sticky[⊥]	in AC_0

via first-order rewritability - a generic result can be established

Key Message

We can transfer complexity results

from classical to consistent query answering

in a generic and uniform way

…with some unexpected exceptions - Acyclic[$\perp$]

Thank you!