TECHNISCHE UNIVERSITÄT DRESDEN

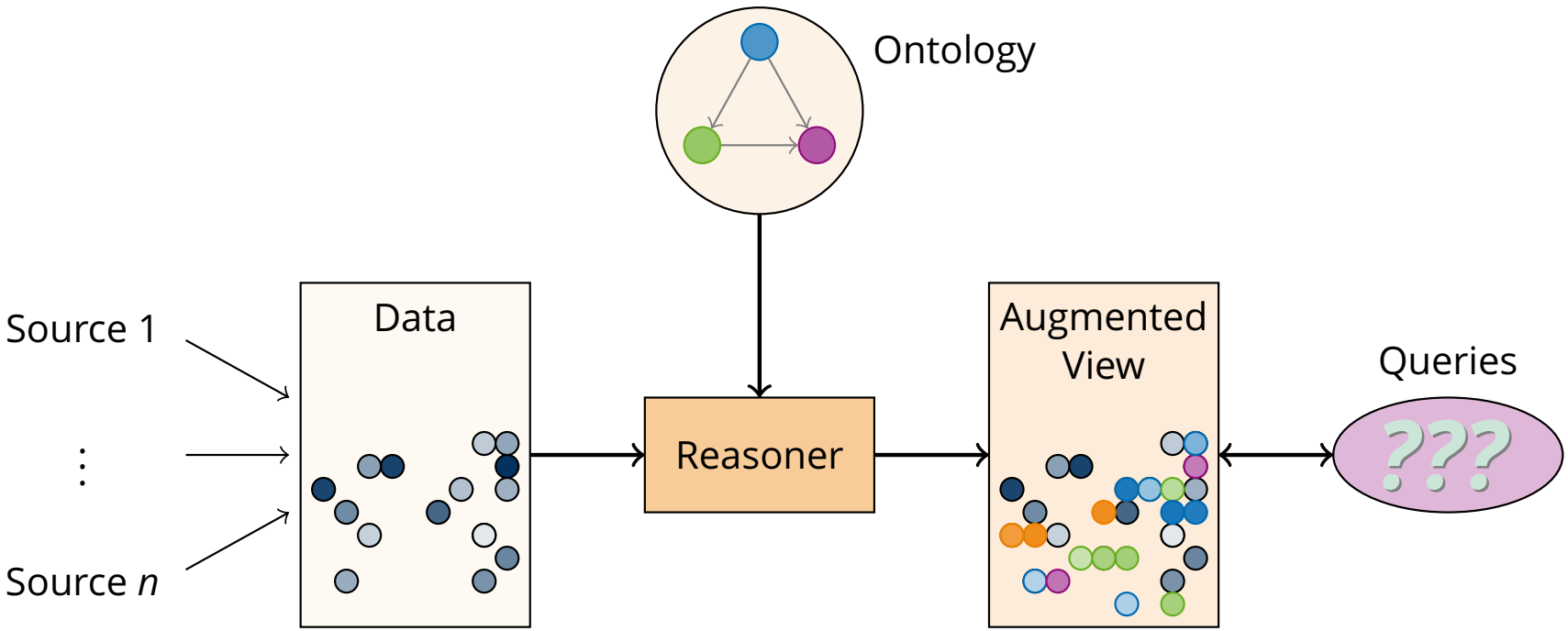
Patrick Koopmann

# Ontology-Based Query Answering for Probabilistic Temporal Data

Stream Reasoning Workshop, Linköping, 2019-04-17

# Ontology-Based Query Answering

- Data coming from various sources
- Ontology defines background knowledge
- Queries answered wrt. augmented view



# Example

## Data (ABox)

$$\text{Male}(\text{peter}) \qquad \text{hasChild}(\text{peter}, \text{tom})$$
$$\text{worksAt}(\text{peter}, \text{TUDresden})$$

## Ontology (TBox)

$$\text{Father} \equiv \text{Male} \sqcap \exists \text{child}.\top$$

## Conjunctive Query (CQ)

$$q(x) \leftarrow \exists y.\text{worksAt}(y, \mathbf{x}) \wedge \text{Father}(y)$$

# Motivating Scenario

**Application:**

Ontology-based hypertension management in smartphone app

# Motivating Scenario

**Application:**

Ontology-based hypertension management in smartphone app

- Sensor measures blood-pressure of patient
- Motion sensors indicate user activity
  - Walking
  - Cycling
  - Sitting
- Context + medical information
- Medical ontology as background knowledge

# Motivating Scenario

## Properties of Scenario

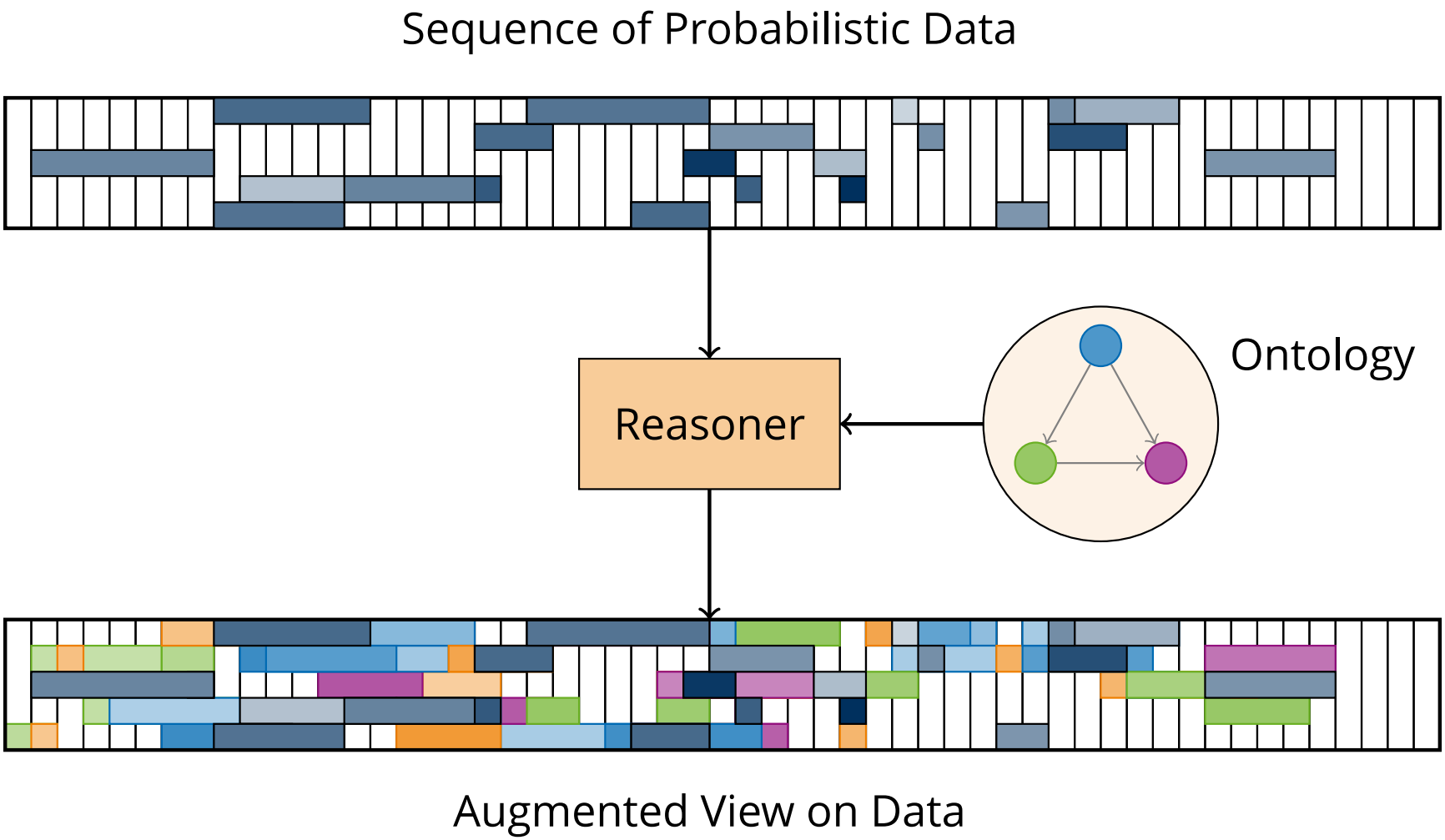
- History of observations relevant
  - Development of blood pressure
  - Recent activity

- A lot of observations *probabilistic* in nature
  - Uncertain measurements of sensor
  - Information inferred from motion sensor

⇒ Requires to handle data that are ***temporal***

⇒ Requires to handle data that are ***probabilistic***

# *Temporal Probabilistic* OBQA



# Temporal Knowledge Bases

- Ontology + Sequence of datasets (ABoxes)
- Signature divided in two parts:
  - ***Rigid*** names
    - Interpretation independent of time
    - e.g. gender, has-parent relation,
  - ***Non-rigid*** names
    - Interpretation may change over time
    - e.g. blood pressure level, user activity

# Example

## Sequence of ABoxes

| 0 | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| $\mathbf{bP(p, b)}$ | | $\mathbf{High(b)}$ | | $\mathbf{High(b)}$ | |

## TBox

$$\text{HighBloodPressurePatient} \equiv \exists \text{bloodPressure}.\text{High}$$

## Rigid Names

bloodPressure (bP)

$\Rightarrow$ HighBloodPressurePatient($p$) at 2 and 4

# Temporal Queries

- Well-investigated query language for temporal KBs
- Combine CQs with *Linear Temporal Logic* (LTL)

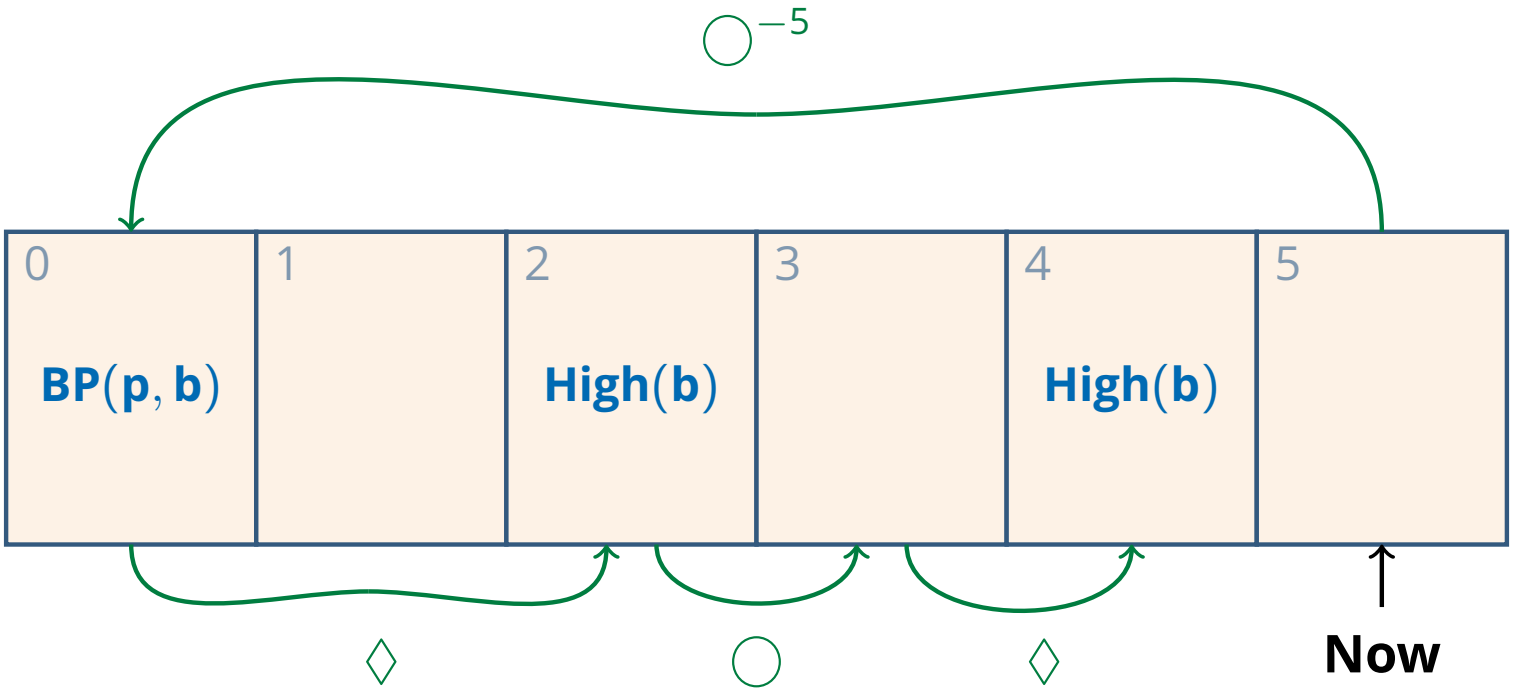
## Temporal queries (TQs)

$$\exists \vec{x}.Q(\vec{x},\vec{y}) \quad | \quad \neg q \quad | \quad q_1 \wedge q_2 \quad | \quad q_1 \vee q_2$$
$$\bigcirc q \quad | \quad \bigcirc^{-} q \quad | \quad \Diamond q \quad | \quad \Diamond^{-} q$$
$$\Box q \quad | \quad \Box^{-} q \quad | \quad q_1 \mathcal{U} q_2 \quad | \quad q_1 \mathcal{S} q_2$$

# Example Temporal Queries

Twice HighBloodPressurePatient within last 5 time units:

$$q(x) \leftarrow \bigcirc^{-5} \Diamond (\mathrm{HBPP}(x) \wedge \bigcirc \Diamond \mathrm{HBPP}(x))$$



# Probabilistic OBQA

- Probabilistic ABoxes: Based on Probabilistic Databases
- Simplest approach: Assign probabilities to ABox axioms

$$\mathit{Exercising}(\mathit{patient}) : 0.6 \qquad \mathit{HighBloodPressure}(\mathit{patient}) : 0.8$$

- Define *probability measure* over *possible worlds*
- Assume statistical independence
  - More advanced models assign formulae over statistical variables

  $\Rightarrow$ Complexities presented here still apply

# Temporal Probabilistic KBs

Temporal Probabilistic ABox:

| 0 | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| $\mathbf{BP(p, b)}$ | $\mathbf{High(b)}$:**0.7** | | $\mathbf{High(b)}$:**0.9** | | $\mathbf{High(b)}$:**0.6** |

# Temporal Probabilistic KBs

Temporal Probabilistic ABox:

| 0 | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| **BP(p, b)** | **High(b):0.7** | | **High(b):0.9** | | **High(b):0.6** |

Possible worlds:

| | | | | | | |
|---|---|---|---|---|---|---|
| $BP(p,b)$ | $High(b)$ | | $High(b)$ | | $High(b)$ | 0.378 |
| $BP(p,b)$ | $High(b)$ | | $High(b)$ | | | 0.162 |
| $BP(p,b)$ | $High(b)$ | | | | $High(b)$ | 0.042 |
| $BP(p,b)$ | $High(b)$ | | | | | 0.018 |
| $BP(p,b)$ | | | $High(b)$ | | $High(b)$ | 0.252 |
| $BP(p,b)$ | | | $High(b)$ | | | 0.108 |
| $BP(p,b)$ | | | | | $High(b)$ | 0.028 |
| $BP(p,b)$ | | | | | | 0.012 |

# Temporal Probabilistic Queries

## Temporal Probabilistic Query Language

*Conjunctive Queries* + *LTL operators* + *prob. operators*

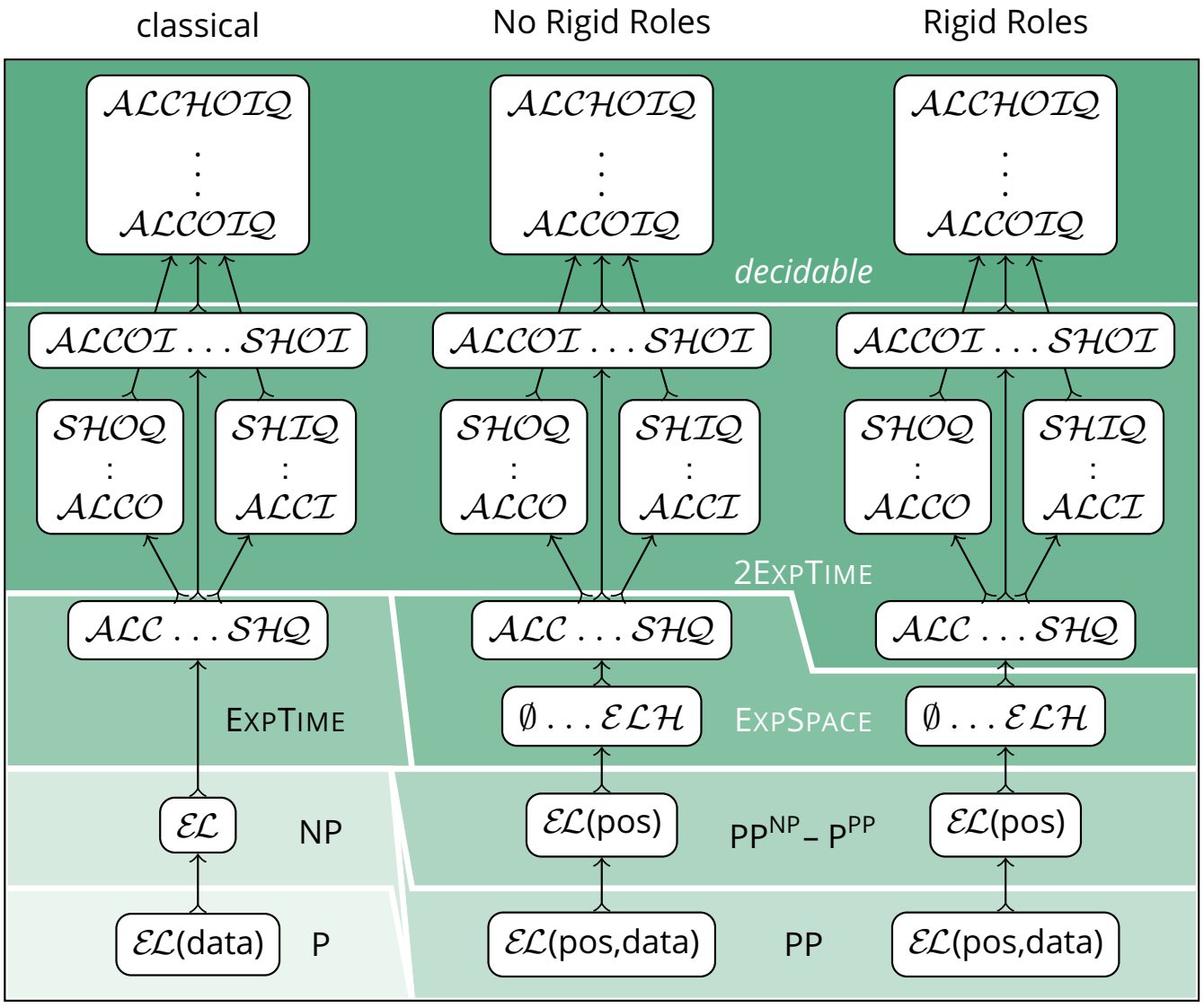
$$P_{>p}Q \qquad P_{=p}Q \qquad P_{<p}Q$$

## Example

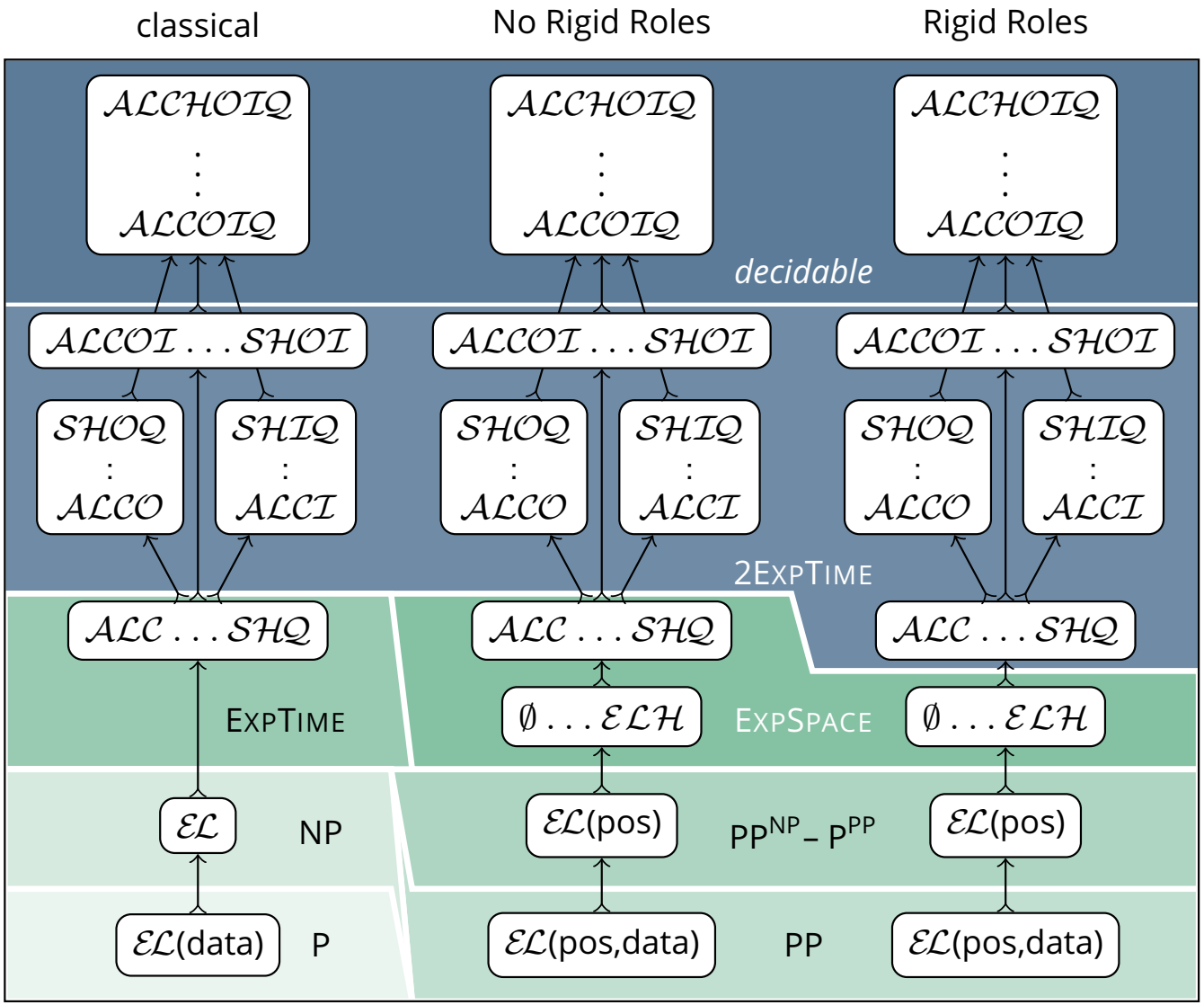
$$q(x) \leftarrow \bigcirc^{-10} \left(P_{<0.2}(\text{Exercising}(x)) \,\mathcal{U}\, P_{>0.8}(\text{HighBPP}(x))\right)$$

- During the ***last 10 time units***,
- patient was with ***low probability*** exercising
- ***until*** with ***high probability*** he had high blood pressure

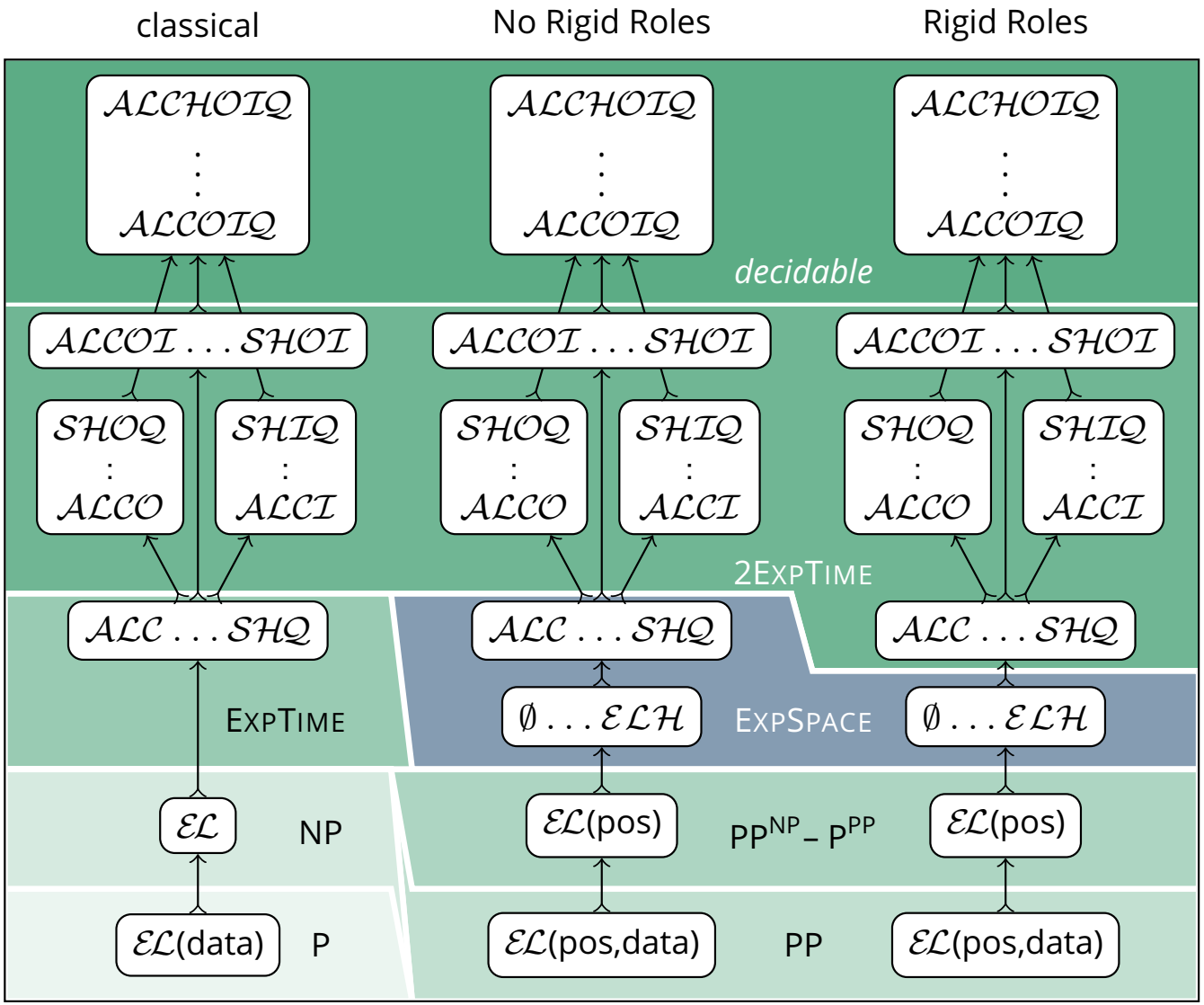
# Complexities



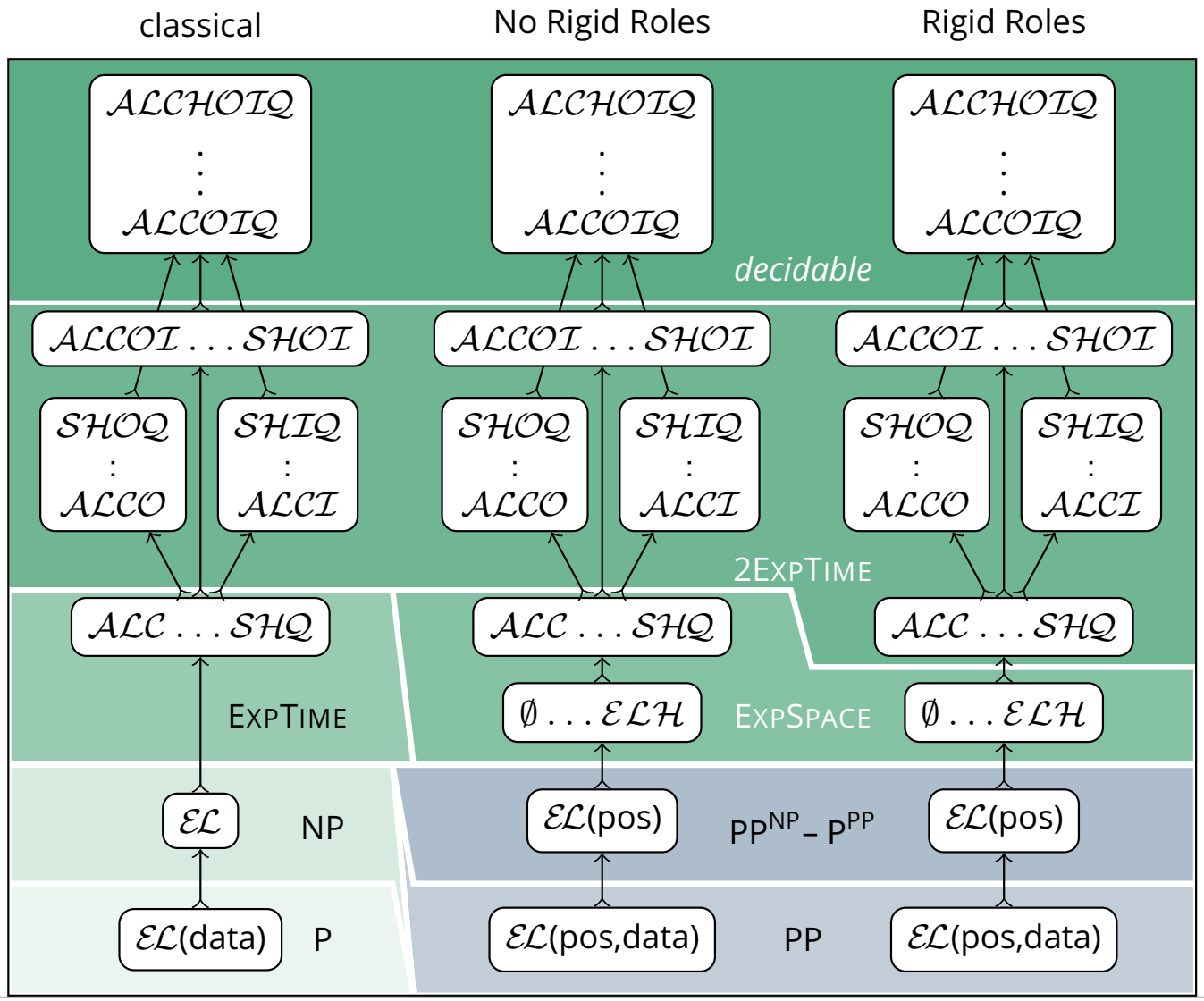
# Complexities



# Complexities



# Complexities Probabilistic Temporal OBQA



# Outlook

## Main results

- Framework combining temporal and probabilistic OBQA
- ExpSpace-hard, already without ontology
- Source of complexity: negation
  - Without, not much harder as probabilistic data access (PP/$PP^{NP}$)

## Current Research

- Prototypical implementation
  - positive queries, *DL-Lite*
- Extend also ontology language
  - prob.+temp. concept+axiom operators
  - prob. + temp. on concepts makes 2-ExpSpace-hard