TECHNISCHE UNIVERSITÄT DRESDEN

Computational Logic Group

# Foundations of Knowledge Representation

## Lecture 7: Nonmonotonic Reasoning I

**Hannes Straß**

**based on slides of Bernardo Cuenca Grau, Ian Horrocks, and Przemysław Wałęga**

DRESDEN concept
Exzellenz aus Wissenschaft und Kultur

# A Fundamental Limitation of FOL

Humans constantly face the necessity of making decisions:

> What flight should I take?, What should I study?, What kind of surgery does this patient need?, . . .

Ideally, we would

1. Start with sufficient information about the problem
   > All direct and non-direct flights from London to Shanghai, their prices, flight length, and seat availability
2. Apply logical reasoning to draw a conclusion
   > The cheapest, shortest flight is with Virgin from LHR at 2pm.
3. Use the conclusion to make an informed decision
   > Buy tickets for the relevant flight

# A Fundamental Limitation of FOL

FOL Knowledge Representation addresses this ideal situation:

1. We gather information
2. We represent it in a knowledge base
3. We pose queries and get answers using reasoning

But in reality, we may not have sufficient information

Our decision making is sometimes based on common sense assumptions, rather than on FOL derivable conclusions:

- *If LHR website shows no departing flight from London to Shanghai at 2pm, then there is no such flight.*
- *Typically, the human heart is on the left side of the body.*

# A Fundamental Limitation of FOL

Consider the statement

"*Typically, humans have their heart on the left side of the body.*"

If I am a doctor and meet Mary Jones for the first time, I would conclude in the absence of additional information that

"*Mary Jones's heart is on the left side of her body*"

However, there is a rare condition, called *situs inversus*, in which the Heart is mirrored from its normal position.

If I examine her and discover that her heart is on the right, I should revise my previous conclusion and deduce that she has situs inversus.

# A Fundamental Limitation of FOL

Suppose I try to model the previous situation in FOL

$$\forall x.(\mathit{Human}(x) \rightarrow \exists y.(\mathit{hasOrg}(x,y) \wedge \mathit{Heart}(y)))$$
$$\forall x.(\mathit{Heart}(x) \leftrightarrow \mathit{NormalHeart}(x) \vee \mathit{SitInvHeart}(x))$$
$$\forall x.(\mathit{NormalHeart}(x) \leftrightarrow \mathit{Heart}(x) \wedge \mathit{hasLocation}(x, \mathit{left}))$$
$$\forall x.(\mathit{SitInvHeart}(x) \leftrightarrow \mathit{Heart}(x) \wedge \mathit{hasLocation}(x, \mathit{right}))$$
$$\forall x.(\mathit{hasLocation}(x, \mathit{left}) \wedge \mathit{hasLocation}(x, \mathit{right}) \rightarrow \bot)$$
$$\forall x.(\mathit{SitInvPatient}(x) \leftrightarrow \mathit{Human}(x) \wedge \exists y.(\mathit{hasOrg}(x,y) \wedge \mathit{SitInvHeart}(y)))$$
$$\mathit{Human}(\mathit{MaryJones})$$

# A Fundamental Limitation of FOL

Suppose I try to model the previous situation in FOL

$$\forall x.(Human(x) \rightarrow \exists y.(hasOrg(x,y) \wedge Heart(y)))$$
$$\forall x.(Heart(x) \leftrightarrow NormalHeart(x) \vee SitInvHeart(x))$$
$$\forall x.(NormalHeart(x) \leftrightarrow Heart(x) \wedge hasLocation(x, left))$$
$$\forall x.(SitInvHeart(x) \leftrightarrow Heart(x) \wedge hasLocation(x, right))$$
$$\forall x.(hasLocation(x, left) \wedge hasLocation(x, right) \rightarrow \bot)$$
$$\forall x.(SitInvPatient(x) \leftrightarrow Human(x) \wedge \exists y.(hasOrg(x,y) \wedge SitInvHeart(y))$$
$$Human(MaryJones)$$

KB does not entail either of the following (not enough info):

$$SitInvPatient(MaryJones) \quad \neg SitInvPatient(MaryJones)$$

Nor does it entail either of:

$$hasLocation(MJH, right) \quad \neg hasLocation(MJH, right)$$

where *MJH* is Mary's heart

# A Fundamental Limitation of FOL

$$\forall x.(Human(x) \rightarrow \exists y.(hasOrg(x,y) \wedge Heart(y)))$$
$$\forall x.(Heart(x) \leftrightarrow NormalHeart(x) \vee SitInvHeart(x))$$
$$\forall x.(NormalHeart(x) \leftrightarrow Heart(x) \wedge hasLocation(x, left))$$
$$\forall x.(SitInvHeart(x) \leftrightarrow Heart(x) \wedge hasLocation(x, right))$$
$$\forall x.(hasLocation(x, left) \wedge hasLocation(x, right) \rightarrow \bot)$$
$$\forall x.(SitInvPatient(x) \leftrightarrow Human(x) \wedge \exists y.(hasOrg(x,y) \wedge SitInvHeart(y))$$
$$Human(MaryJones)$$

To deduce $\neg SitInvPatient(MaryJones)$, we could add the facts:

$$Heart(MJH) \quad hasOrg(MaryJones, MJH) \quad hasLocation(MJH, left)$$

But then, when I examine the patient I should add new evidence

$$hasLocation(MJH, right)$$

Problem: the KB is now unsatisfiable

# A Fundamental Limitation of FOL

In FOL, we cannot

1. Draw “default” or “common sense” conclusions
2. Withdraw conclusions when presented with new evidence

FOL’s knowledge is “certain”

- If a sentence is not entailed, it is not known

  Nothing we can assume “by default”

- If new information contradicts existing, we get unsatisfiability

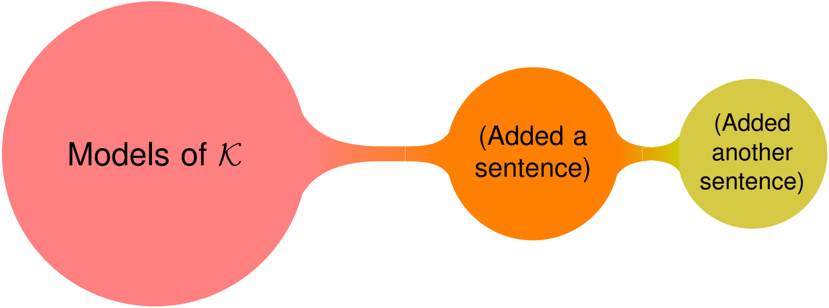
  Our only choice is to modify the KB manually

This is due to a property of FOL called Monotonicity

# Monotonicity of FOL

Take sets of FOL sentences $\mathcal{K}$ and $\mathcal{K}'$ with $\mathcal{K} \subseteq \mathcal{K}'$

1. The set of models of $\mathcal{K}'$ is contained in the set of models of $\mathcal{K}$
2. If $\mathcal{K} \models \alpha$, then $\mathcal{K}' \models \alpha$




By adding FOL sentences to a knowledge base we gain knowledge:

- Reduce number of models
- Increase number consequences (recall entailment definition)

# Introducing Nonmonotonicity: CWA

Departures data for all London airports between 1pm and 2pm:

$$\mathit{flight}(\mathit{LHR}, \mathit{Paris}) \quad \mathit{flight}(\mathit{LGW}, \mathit{Johannesburg})$$
$$\mathit{flight}(\mathit{LGW}, \mathit{Doha}) \quad \mathit{flight}(\mathit{LHR}, \mathit{Beijing})$$
$$\mathit{flight}(\mathit{LUT}, \mathit{Madrid}) \quad \mathit{flight}(\mathit{STD}, \mathit{Athens})$$

Since data does not include a Shanghai flight, reasonable to assume that there is no such flight; so-called Closed World Assumption (CWA)

CWA can be thought of as a "rule" that produces new consequences:

"If something is not provably true, then assume that it is false"

We can use CWA rule to deduce, for example

$$\neg \mathit{flight}(\mathit{LHR}, \mathit{Shanghai})$$

CWA rule is non-monotonic:

If the data is extended with the fact $\mathit{flight}(\mathit{LHR}, \mathit{Shanghai})$, then CWA no longer justifies above deduction

# Introducing Nonmonotonicity: Defaults

Consider our example about situs inversus. Suppose we extend our original KB with the following Default Rule:

$$\frac{\mathit{hasOrg}(x, y) \wedge \mathit{Heart}(y) \text{ and not provably true } \neg \mathit{hasLocation}(y, \mathit{left})}{\text{deduce } \mathit{hasLocation}(y, \mathit{left})}$$

Formalises the fact that "*typically, the Human Heart is on the left side*"

If we extend our KB with such rule we would infer

$$\neg \mathit{SitInvPatient}(\mathit{MaryJones})$$

Default rules are non-monotonic:

> If we find out that $\mathit{hasLocation}(\mathit{MJH}, \mathit{right})$, previous entailment no longer holds w.r.t. our KB and default rule

# The Need for Non-monotonic Logics

In formal terms,

- FOL has a monotonic entailment relation $\models$:

$$\mathcal{K} \subseteq \mathcal{K}' \text{ and } \mathcal{K} \models \alpha \quad \Rightarrow \quad \mathcal{K}' \models \alpha$$

- A non-monotonic entailment relation $\mid\!\approx$ is one such that there exist $\mathcal{K}'$, $\mathcal{K} \subseteq \mathcal{K}'$ and $\alpha$ such that $\mathcal{K} \mid\!\approx \alpha$, but $\mathcal{K}' \not\mid\!\approx \alpha$.

Intuitively, there is nothing esoteric about non-monotonic reasoning
In fact our everyday reasoning is often non-monotonic

But as logicians we should insist on:

- well defined syntax and semantics
- well understood computational properties
- semantics that induce a "reasonable" non-monotonic entailment relation—one that is consistent with our intuitions

Our next question is, how to define such a logic?

# The Semantics of FOL Entailment

There are many ways to define a non-monotonic logic from (a fragment of) FOL. We will focus only on one of them.

Idea: Take into account only a subset of preferred models of $\mathcal{K}$ (instead of all models) when checking whether $\mathcal{K} \models \alpha$.

- Monotonic entailment:

  $\mathcal{K} \models \alpha$ iff every (FOL) model of $\mathcal{K}$ is a model of $\alpha$.

- Non-monotonic entailment:

  $\mathcal{K} \mid\approx \alpha$ iff every preferred (FOL) model of $\mathcal{K}$ is a model of $\alpha$.

# The Semantics of FOL Entailment

There are many ways to define a non-monotonic logic from (a fragment of) FOL. We will focus only on one of them.

Idea: Take into account only a subset of preferred models of $\mathcal{K}$ (instead of all models) when checking whether $\mathcal{K} \models \alpha$.

- Monotonic entailment:

  $\mathcal{K} \models \alpha$ iff every (FOL) model of $\mathcal{K}$ is a model of $\alpha$.

- Non-monotonic entailment:

  $\mathcal{K} \mid\approx \alpha$ iff every preferred (FOL) model of $\mathcal{K}$ is a model of $\alpha$.

The non-monotonic entailment relation $\mid\approx$ is supra-classical:

Using $\mid\approx$ we will always derive more consequences than using $\models$.

Key problem: How to specify which models are preferred?

# Preferred Models

Coming back to our flight knowledge base $\mathcal{K}$:

$$\begin{array}{rr} \mathit{flight}(\mathit{LHR}, \mathit{Paris}) & \mathit{flight}(\mathit{LGW}, \mathit{Johannesburg}) \\ \mathit{flight}(\mathit{LGW}, \mathit{Doha}) & \mathit{flight}(\mathit{LHR}, \mathit{Beijing}) \\ \mathit{flight}(\mathit{LUT}, \mathit{Madrid}) & \mathit{flight}(\mathit{STD}, \mathit{Athens}) \end{array}$$

This set of ground literals has infinitely many FOL models, and

$$\mathcal{K} \not\models \neg\mathit{flight}(\mathit{LHR}, \mathit{Shanghai})$$

Here is a (Herbrand) counter-model $\mathcal{I} \not\models \neg\mathit{flight}(\mathit{LHR}, \mathit{Shanghai})$:

$$\begin{array}{rcl} \mathit{flight}^{\mathcal{I}} & = & \{(\mathit{LHR}, \mathit{Paris}), (\mathit{LGW}, \mathit{Johannesburg}) \\ & & (\mathit{LGW}, \mathit{Doha}), (\mathit{LHR}, \mathit{Beijing}) \\ & & (\mathit{LUT}, \mathit{Madrid}), (\mathit{STD}, \mathit{Athens}) \\ & & (\mathit{LHR}, \mathit{Shanghai})\} \end{array}$$

# Preferred Models

Note, however, that there is a special Herbrand model, namely the one that coincides with the set of facts:

$$
\begin{aligned}
\mathit{flight}^{\mathcal{I}_{min}} \quad = \quad & \{(\mathit{LHR}, \mathit{Paris}), (\mathit{LGW}, \mathit{Johannesburg}) \\
& (\mathit{LGW}, \mathit{Doha}), (\mathit{LHR}, \mathit{Beijing}) \\
& (\mathit{LUT}, \mathit{Madrid}), (\mathit{STD}, \mathit{Athens})\}
\end{aligned}
$$

This model $\mathcal{I}_{min}$ is the intersection of all Herbrand models of $\mathcal{K}$, and it is called the least Herbrand model of $\mathcal{K}$

Suppose we specify $\mid\!\approx$ by defining the set of preferred models as

$$\text{Preferred}(\mathcal{K}) = \{\mathcal{I}_{min}\}$$

Clearly $\mathcal{K} \mid\!\approx \neg \mathit{flight}(\mathit{LHR}, \mathit{Shanghai})$

Our entailment relation $\mid\!\approx$ captures the CWA

# Preferred Models

Our strategy of selecting the least Herbrand model as preferred seemed plausible:

We correctly captured CWA (in this example)

Our example was, however, a bit too simplistic

Our KB only contained positive, ground literals!!

Problem: FOL KBs may not have least Herbrand models

$$\forall x.(Mammal(x) \to Male(x) \lor Female(x))$$
$$\forall x.(Male(x) \land Female(x) \to \bot)$$
$$Mammal(a)$$

We have only two Herbrand models:

$$\mathcal{I}_1 \quad : \quad Mammal^{\mathcal{I}_1} = \{a\}, Male^{\mathcal{I}_1} = \{a\}$$
$$\mathcal{I}_2 \quad : \quad Mammal^{\mathcal{I}_2} = \{a\}, Female^{\mathcal{I}_2} = \{a\}$$

Their intersection is not a model of our KB.

# Coming Back to Datalog

Bad news: Things could get much more complicated

Good news: Datalog is a nice logic with least Herbrand Models

$$\forall x.(JuvArthritis(x) \rightarrow JuvDisease(x))$$
$$\forall x.(\forall y.(JuvDisease(x) \wedge Affects(x,y) \rightarrow Child(y)))$$
$$JuvArthritis(JRA)$$
$$Affects(JRA, John)$$

The above KB has a least Herbrand model:

$$\begin{aligned} \mathcal{I}_{min} \quad : \quad & JuvArthritis^{\mathcal{I}_{min}} = \{JRA\}, Affects^{\mathcal{I}_{min}} = \{(JRA, John)\}, \\ & JuvDisease^{\mathcal{I}_{min}} = \{JRA\}, Child^{\mathcal{I}_{min}} = \{John\} \end{aligned}$$

And we have a way to compute it: forward-chaining.

# Coming Back to Datalog

So, what is the difference with Datalog under monotonic semantics?

$$\forall x.(JuvArthritis(x) \rightarrow JuvDisease(x))$$
$$\forall x.(\forall y.(JuvDisease(x) \land Affects(x,y) \rightarrow Child(y)))$$
$$JuvArthritis(JRA)$$
$$Affects(JRA, John)$$

$$\mathcal{I}_{min} \quad : \quad JuvArthritis^{\mathcal{I}_{min}} = \{JRA\}, Affects^{\mathcal{I}_{min}} = \{(JRA, John)\},$$
$$JuvDisease^{\mathcal{I}_{min}} = \{JRA\}, Child^{\mathcal{I}_{min}} = \{John\}$$

$\mathcal{K} \quad Child(John)$ $\qquad$ $\mathcal{K} \quad Child(John)$

$\mathcal{K} \quad \neg Child(JRA)$ $\qquad$ $\mathcal{K} \quad \neg Child(JRA)$

# Coming Back to Datalog

So, what is the difference with Datalog under monotonic semantics?

$$\forall x.(JuvArthritis(x) \rightarrow JuvDisease(x))$$
$$\forall x.(\forall y.(JuvDisease(x) \wedge Affects(x, y) \rightarrow Child(y)))$$
$$JuvArthritis(JRA)$$
$$Affects(JRA, John)$$

$$\mathcal{I}_{min} \quad : \quad JuvArthritis^{\mathcal{I}_{min}} = \{JRA\}, Affects^{\mathcal{I}_{min}} = \{(JRA, John)\},$$
$$JuvDisease^{\mathcal{I}_{min}} = \{JRA\}, Child^{\mathcal{I}_{min}} = \{John\}$$

No difference with respect to entailment of positive literals:

$\mathcal{K} \models Child(John)$ $\qquad$ $\mathcal{K} \mathrel{|}\!\approx Child(John)$

$\mathcal{K} \quad \neg Child(JRA)$ $\qquad$ $\mathcal{K} \quad \neg Child(JRA)$

# Coming Back to Datalog

So, what is the difference with Datalog under monotonic semantics?

$$\forall x.(\mathit{JuvArthritis}(x) \rightarrow \mathit{JuvDisease}(x))$$
$$\forall x.(\forall y.(\mathit{JuvDisease}(x) \wedge \mathit{Affects}(x,y) \rightarrow \mathit{Child}(y)))$$
$$\mathit{JuvArthritis}(\mathit{JRA})$$
$$\mathit{Affects}(\mathit{JRA}, \mathit{John})$$

$$\mathcal{I}_{min} \quad : \quad \mathit{JuvArthritis}^{\mathcal{I}_{min}} = \{\mathit{JRA}\}, \mathit{Affects}^{\mathcal{I}_{min}} = \{(\mathit{JRA}, \mathit{John})\},$$
$$\mathit{JuvDisease}^{\mathcal{I}_{min}} = \{\mathit{JRA}\}, \mathit{Child}^{\mathcal{I}_{min}} = \{\mathit{John}\}$$

No difference with respect to entailment of positive literals:

$$\mathcal{K} \models \mathit{Child}(\mathit{John}) \qquad \mathcal{K} \;|\!\approx\; \mathit{Child}(\mathit{John})$$

Very different w.r.t. entailment of negative literals (CWA)

$$\mathcal{K} \not\models \neg\mathit{Child}(\mathit{JRA}) \qquad \mathcal{K} \;|\!\approx\; \neg\mathit{Child}(\mathit{JRA})$$

# Limitations

We have successfully formalised CWA in a useful FOL fragment.

However, we have just seen the tip of the iceberg:

1. We still don't know what to do with FOL fragments not having least Herbrand models
2. Datalog with non-monotonic semantics is not sufficiently expressive to represent default statements

$$\frac{\mathit{hasOrg}(x, y) \wedge \mathit{Heart}(y) \text{ \& not provably true } \neg \mathit{hasLocation}(y, \mathit{left})}{\text{deduce } \mathit{hasLocation}(y, \mathit{left})}$$

So, we have reached a crossroads:

1. We need logics beyond Datalog to express defaults
2. Not clear how to define $\models\!\!\!\approx$ for those fragments