MANCHESTER
1824

The University of Manchester

From standard reasoning problems to non-standard reasoning problems and one step further

Uli Sattler
University of Manchester, University of Oslo
uli.sattler@manchester.ac.uk

Some Advertisement

Description Logics (DLs) have a long tradition in computer science and knowledge representation, being designed so that domain knowledge can be described and so that computers can reason about this knowledge. DLs have recently gained increased importance since they form the logical basis of widely used ontology languages, in particular the web ontology language OWL.

Written by four renowned experts, this is the first textbook on Description Logic. It is suitable for self-study by graduates and as the basis for a university course. Starting from a basic DL, the book introduces the reader to their syntax, semantics, reasoning problems and model theory, and discusses the computational complexity of these reasoning problems and algorithms to solve them. It then explores a variety of reasoning techniques, knowledge-based applications and tools, and describes the relationship between DLs and OWL.

Franz Baader is a professor in the Institute of Theoretical Computer Science at TU Dresden.

Ian Horrocks is a professor in the Department of Computer Science at the University of Oxford.

Carsten Lutz is a professor in the Department of Computer Science at the University of Bremen.

Uli Sattler is a professor in the Information Management Group within the School of Computer Science at the University of Manchester.

Cover illustration: The Description Logic logo.
Courtesy of Enrico Franconi.

Designed by Zoe Naylor.

CAMBRIDGE
UNIVERSITY PRESS
www.cambridge.org

ISBN 978-0-521-87361-1

9 780521 873611

Baader, Horrocks, Lutz, Sattler

An Introduction to Description Logic

CAMBRIDGE

An Introduction to Description Logic

Franz Baader
Ian Horrocks
Carsten Lutz
Uli Sattler

be described and so that computers can reason about this knowledge. DLs have recently gained increased importance since they form the logical basis of widely used ontology languages, in particular the web ontology language OWL.

Written by four renowned experts, this is the first textbook on Description Logic. It is suitable for self-study by graduates and as the basis for a university course. Starting from a basic DL, the book introduces the reader to their syntax, semantics, reasoning problems and model theory, and discusses the computational complexity of these reasoning problems and algorithms to solve them. It then explores a variety of reasoning techniques, knowledge-based applications and tools, and describes the relationship between DLs and OWL.

Department of Computer Science at the University of Oxford.

Carsten Lutz is a professor in the Department of Computer Science at the University of Bremen.

Uli Sattler is a professor in the Information Management Group within the School of Computer Science at the University of Manchester.

Cover illustration: The Description Logic logo. Courtesy of Enrico Franconi.

Designed by Zoe Naylor.

CAMBRIDGE UNIVERSITY PRESS
www.cambridge.org

ISBN 978-0-521-87361-1

ocks,

An Introduction to Description Logic

CAMBRIDGE

An Introduction to Description Logic

Franz Baader
Ian Horrocks
Carsten Lutz
Uli Sattler

MANCHESTER 1824

The University of Manchester

Standard
&
Non-Standard
Reasoning Problems

MANCHESTER 1824

The University of Manchester

Standard Reasoning Problems

we all know them: given $C, D, \mathcal{O}, \mathcal{T}, \mathcal{A}, \ldots$ decide/compute

- consistency/satisfiability
- subsumption
- classification
- query answering

- …all only involve entailment checks:

$$\mathcal{O} \models \alpha$$

- …possibly many (classification!)

Non-Standard Reasoning Problems

we all know them: given $C, D, \mathcal{O}, \mathcal{T}, \mathcal{A}, \ldots$

- msc$(a, \mathcal{O})$, lcs$(C, D, \mathcal{O}), \ldots$
- Justs$(\alpha, \mathcal{O})$, PinPoint$(\alpha, \mathcal{O}), \ldots$
- match$(C, P, \mathcal{O})$, unify$(P_1, P_2, \mathcal{O}), \ldots$
- x-mod$(\Sigma, \mathcal{O}), \ldots$

- …involve finding *extreme* X such that …
  - *subset-minimal* or
  - *maximally/minimally strong*
- …possibly many such Xs

Non-Standard Reasoning Problems

we all know them: given $C, D, \mathcal{O}, \mathcal{T}, \mathcal{A}, \ldots$

- msc$(a, \mathcal{O})$, lcs$(C, D, \mathcal{O}), \ldots$
- Justs$(\alpha, \mathcal{O})$, PinP...
- match$(C, \ldots$
- x-mod$(\Sigma, \ldots$

Are
- (conservative) rewritability
- (query) inseparability

also standard reasoning problems?

- …involve finding *extreme* X such that …
  - *subset-minimal* or
  - *maximally/minimally strong*
- …possibly many such Xs

(Non-)Standard Reasoning: we know how to

understand problems:

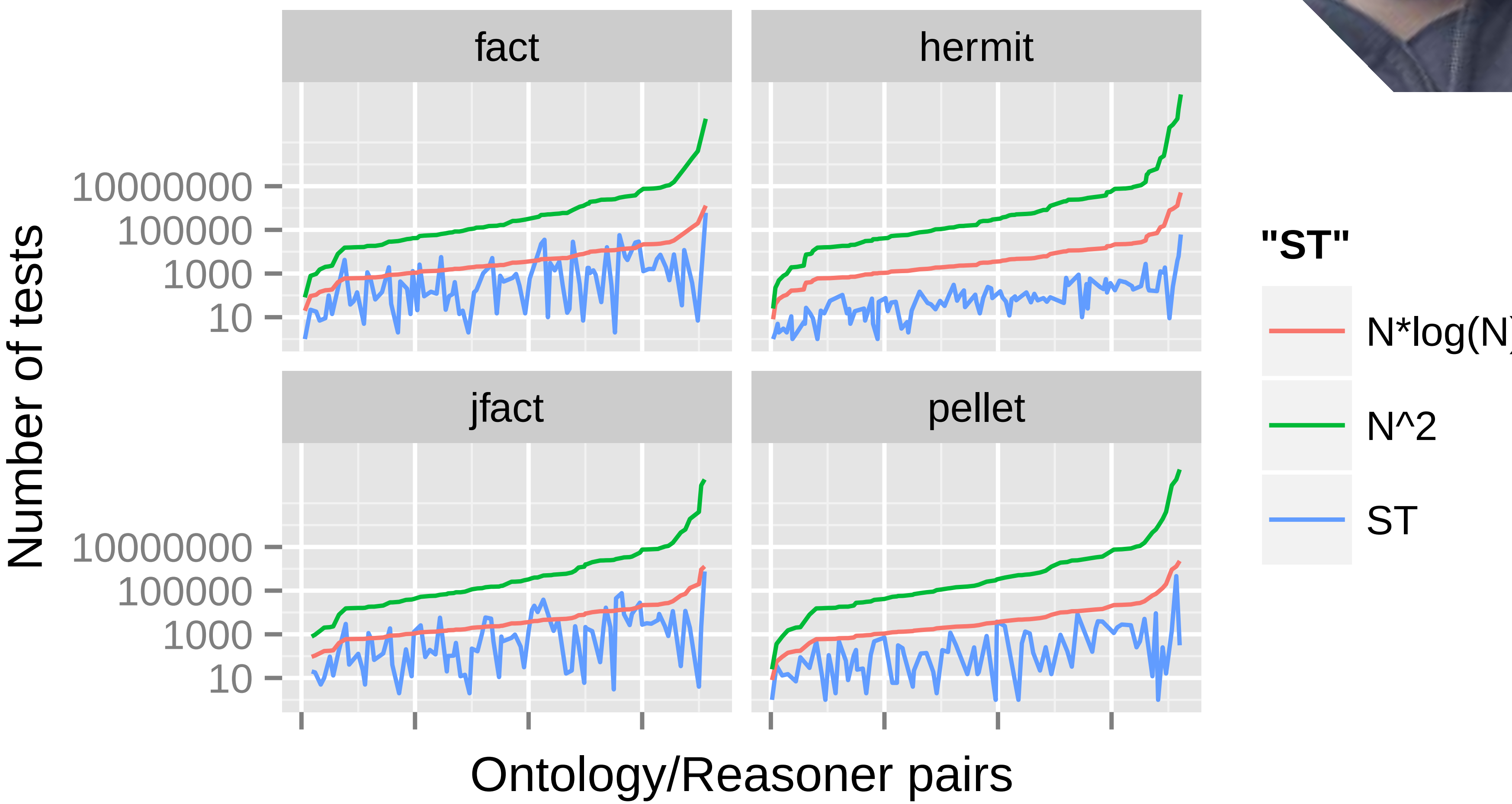
- decidability & computational complexity
  - worst case
  - data
  - parametrised
  - …

understand solutions:

- soundness, completeness, termination
- relations between them
- complexity
  - see above
- practicability
  - worst case complexity ≠ best case complexity
  - amenable to optimisation
  - empirical evaluation

An interesting side note

from our empirical evaluation:
how many subsumption does classification involve?



Not always that straightforward

- Which problem/solution to consider when?
  - e.g., $\text{x-mod}(\Sigma, \mathcal{O}), \ldots$
    - minimal/top/bottom/semantic/…
    - depends on size, signature, application, …
  - but we know properties of/relations between solutions
    - smallest
    - self-contained
    - unique
    - depleting
    - …
- How to measure practicability?
  - benchmarks, ORE,..

Not always that straightforward

- Which problem/solution to consider when?
  - e.g., x-mod$(\Sigma, \mathcal{O}), \ldots$
    - minimal/top/bottom/… tic/…
    - depe… application, …
  - bu… ions between solutions
    -
    - s
    - u
    - de… g
    - …
- How to measure practicability?
  - benchmarks, ORE,..

Extension/variants of DLs
- probabilistic
- non-monotonic
- rules
- …

Subjective Ontology-Based Problems

Subjective Ontology-Based Problems

are problems that are based on

- $C, \; D, \; \mathcal{O}, \; \mathcal{T}, \; \mathcal{A}, \; \models, \ldots$ plus
- additional parameter(s)

Subjective Ontology-Based Problems

are problems that are based on

- $C, \ D, \ \mathcal{O}, \ \mathcal{T}, \ \mathcal{A}, \ \models, \ldots$ plus
- additional parameter(s)

- because **objective solution** is
  - not feasible/computable
  - or makes little sense
  - e.g. in $\mathcal{SROIQ}$: $\mathsf{ComSubs}(C, \ D, \{ \ C \sqsubseteq \forall R.(A \sqcap C), \ D \sqsubseteq \forall R.(A \sqcap D)\})$

Subjective Ontology-Based Problems

are problems that are based on

- $C, \; D, \; \mathcal{O}, \; \mathcal{T}, \; \mathcal{A}, \; \models, \ldots$ plus
- additional parameter(s)

- because **objective solution** is
  - not feasible/computable
  - or makes little sense
  - e.g. in $\mathcal{SROIQ}$: ComSubs$(C, \; D, \{ \; C \sqsubseteq \forall R.(A \sqcap C), \; D \sqsubseteq \forall R.(A \sqcap D)\})$

$$= \{\forall R.A, \; \forall R.\forall R.A, \; \forall R.\forall R.\forall R.A, \; \ldots\}$$

Subjective Ontology-Based Problems

are problems that are based on

- $C, \; D, \; \mathcal{O}, \; \mathcal{T}, \; \mathcal{A}, \; \models, \ldots$ plus
- additional parameter(s)

- because **objective solution** is
  - not feasible/computable
  - or makes little sense
  - e.g. in $\mathcal{SROIQ}$: ComSubs$(C, \; D, \{ \; C \sqsubseteq \forall R.(A \sqcap C), \; D \sqsubseteq \forall R.(A \sqcap D)\})$

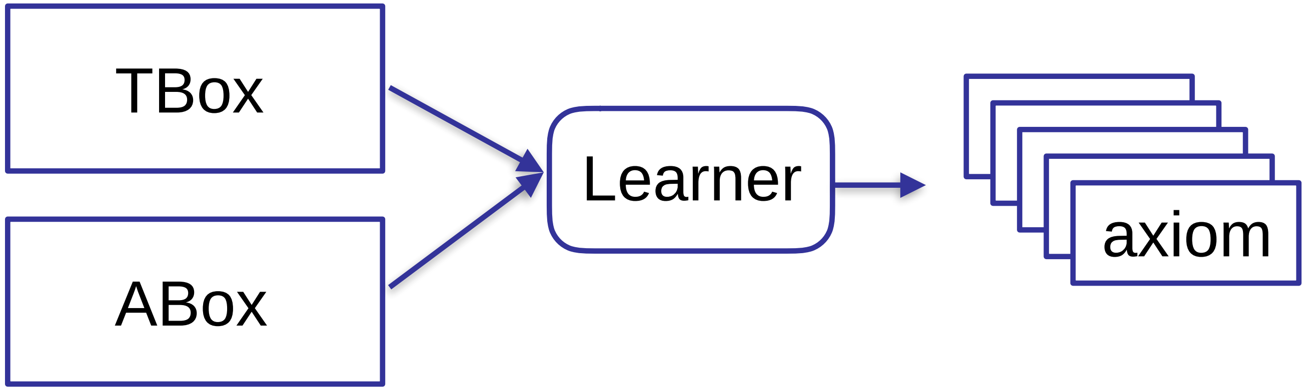
$$= \{\forall R.A, \; \forall R.\forall R.A, \; \forall R.\forall R.\forall R.A, \; \ldots\}$$

- or we want to capture quality criteria
  - interestingness
  - readability
  - relevance …

A subjective OB problem:
Mining TBox Axioms from KBs
or
Finding Interesting Correlations

Mining TBox axioms from KBs

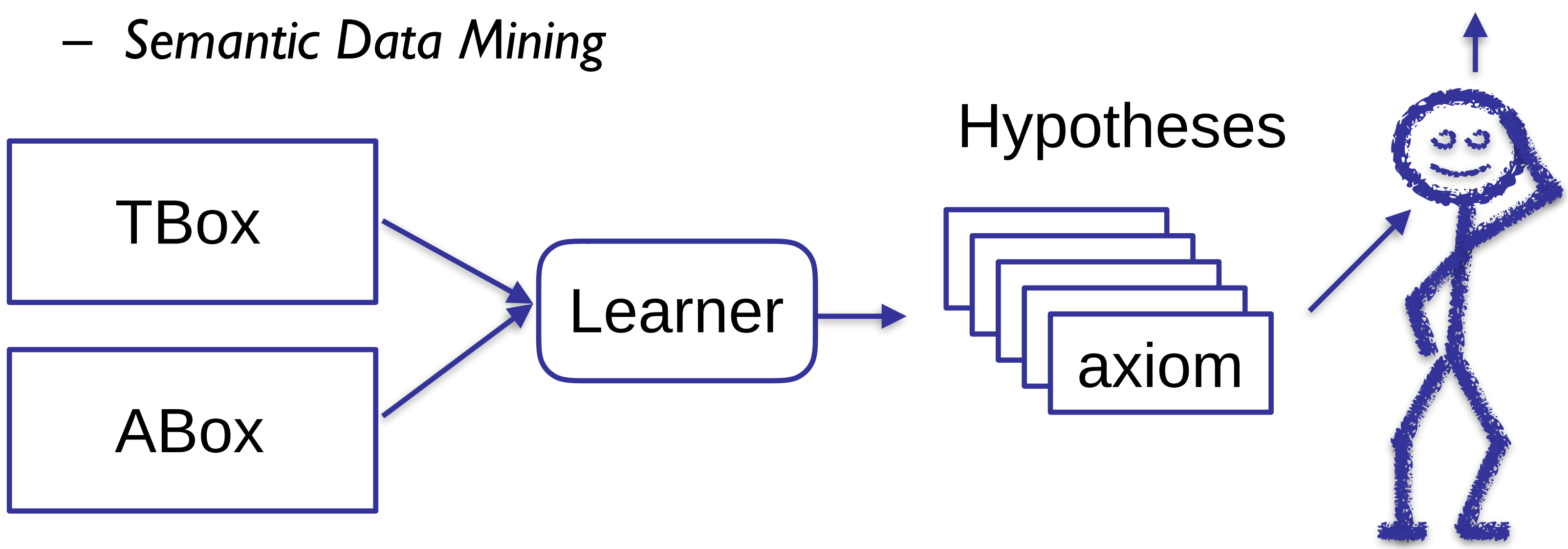
- learn (implicit) correlations in our data
- get interesting insights into domain



Do not confuse with
(exact) *learning of TBoxes* (via probing queries)

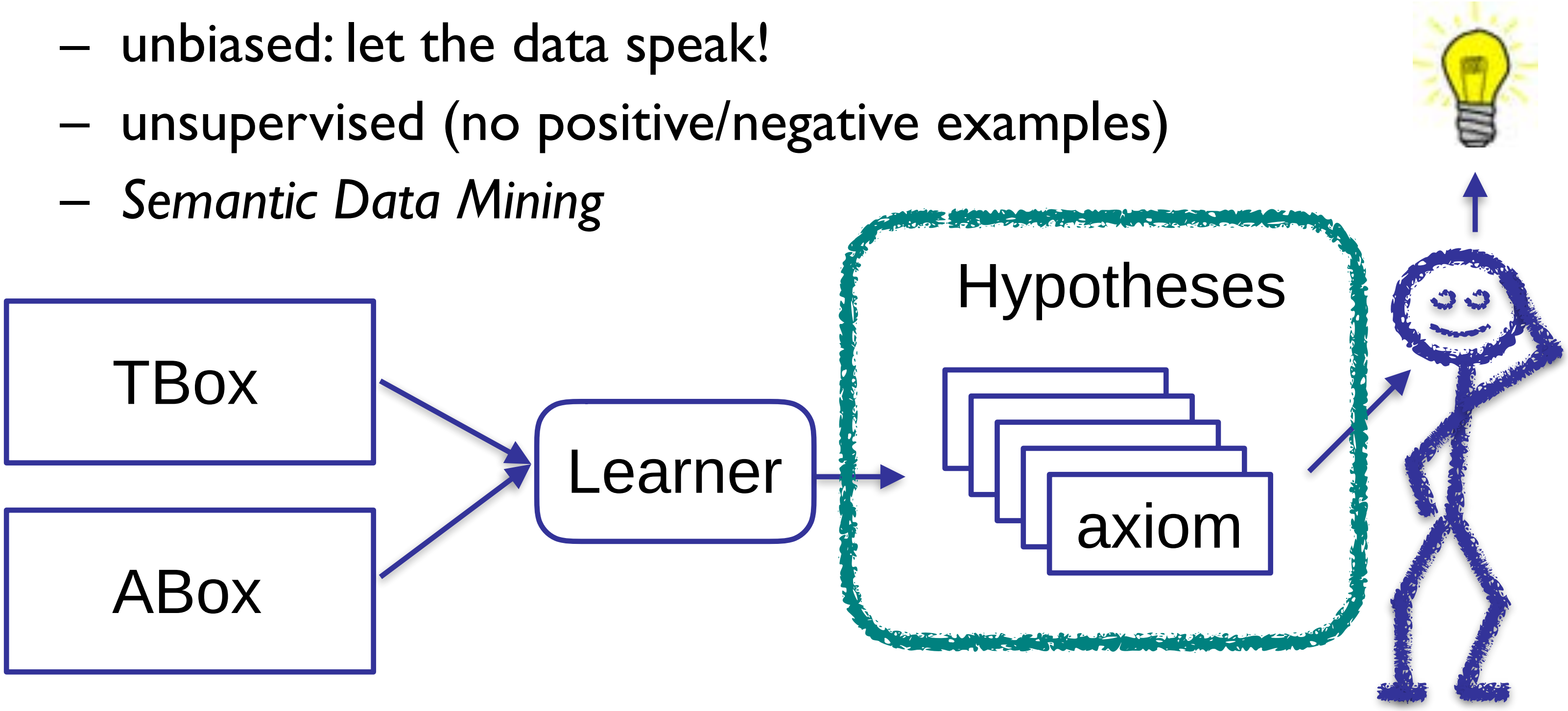
Mining TBox axioms from KBs

- Correlations in KB = classical machine learning
- automatic generation of knowledge from data
  - taking background knowledge in KB into account
  - unbiased: let the data speak!
  - unsupervised (no positive/negative examples)
  - *Semantic Data Mining*

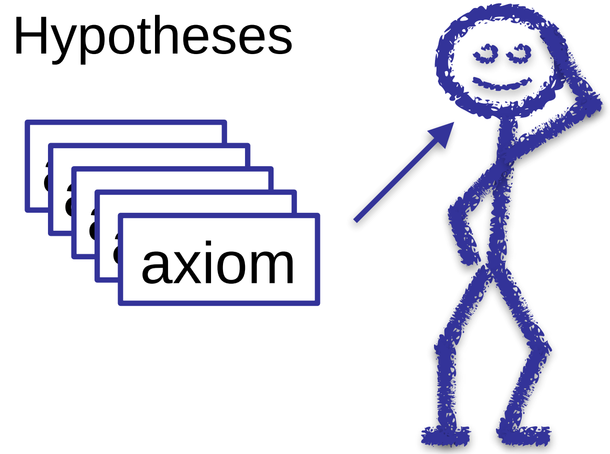


Mining TBox axioms from KBs

- Correlations in KB = classical machine learning
- automatic generation of knowledge from data
  - taking background knowledge in KB into account
  - unbiased: let the data speak!
  - unsupervised (no positive/negative examples)
  - *Semantic Data Mining*



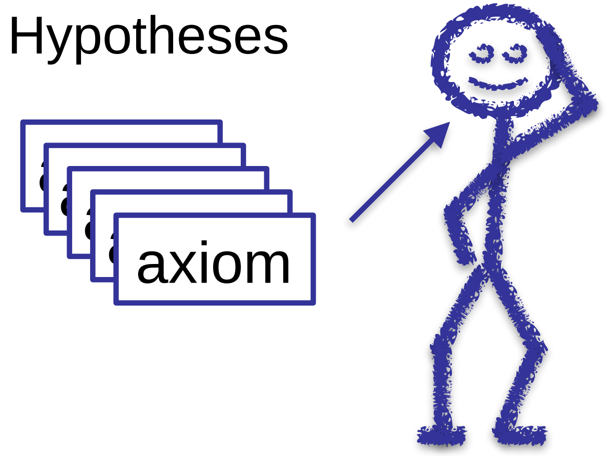
Mining TBox axioms from KBs



- Which kind of hypotheses to capture correlations in KB?
  1. expressive: GCIs, role inclusions
  2. readable
  3. logically sound
  4. statistically sound

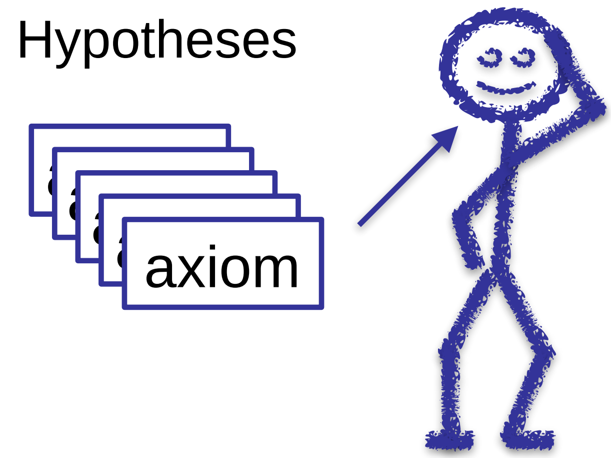
2. Readable Hypotheses

- A *hypothesis* is
  - a small set of short axioms
    - fewer than $n_{\max}$ axioms
    - with concepts shorter than $\ell_{\max}$
  - in a suitable DL: $\mathcal{ALCHI} \ldots \mathcal{SROIQ}$
  - free of redundancy
    - no superfluous parts
    - ✓ preferred laconic justifications

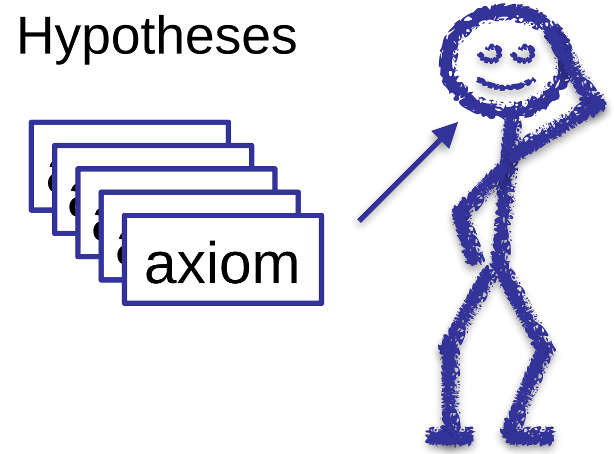


3. Logically Sound Hypotheses

- A hypothesis H should be
  - ✓ informative: $\forall \alpha \in H : \mathcal{O} \not\models \alpha$
    - ✓ we want to mine new axioms
  - ✓ consistent: $\mathcal{O} \cup H \not\models \top \sqsubseteq \bot$
  - ✓ non-redundant among all hypotheses:
    - there is no $H', H \in \mathbb{H} : H \neq H'$ and $H' \equiv H$



3. Logically Sound Hypotheses



- A hypothesis H should be
  - ✓ informative: $\forall \alpha \in H : \mathcal{O} \not\models \alpha$
    - ✓ we want to mine new axioms
  - ✓ consistent: $\mathcal{O} \cup H \not\models \top \sqsubseteq \bot$
  - ✓ non-redundant among all hypotheses:
    - there is no $H', H \in \mathbb{H} : H \neq H'$ and $H' \equiv H$

- Different hypotheses can be compared wrt. their
  - ✓ logical strength:
  - ? maximally strong?
    - no: overfitting!
  - ? minimally strong?
    - no: under-fitting
  - ✓ reconciliatory power
    - brings together terms so far only loosely related

4. Statistically Sound Hypotheses

- we need to assess *data support* of hypothesis
- introduce metrics that capture *quality* of an axiom
  - learn from association rule mining (ARM):
    - count individuals that *support* a GCI
  - count instances, neg instances, non-instances
    - using standard DL semantics, OWA, TBox, entailments,….
    - no 'artificial closure'
  - make sure you treat a GCI as an axiom and not as a rule
    - contrapositive!
  - coverage, support, …, lift

4. Statistically Sound Hypotheses

Some useful notation:

- $\mathsf{Inst}(C, \mathcal{O}) := \{a \mid \mathcal{O} \models C(a)\}$
- $\mathsf{UnKn}(C, \mathcal{O}) := \mathsf{Inst}(\top, \mathcal{O}) \setminus (\mathsf{Inst}(C, \mathcal{O}) \cup \mathsf{Inst}(\neg C, \mathcal{O}))$
- relativized: $\mathsf{P}(C, \mathcal{O}) := \#\,\mathsf{Inst}(C, \mathcal{O}) / \#\,\mathsf{Inst}(\top, \mathcal{O})$
- projection tables:

	C1	C2	C3	C4	…
Ind1	X	X	X	?	…
Ind2	0	X	X	0	…
Ind3	?	?	X	?	…
Ind4	?	0	?	?	…
…	…	…	…	…	…

4. Statistically Sound Hypotheses: Axioms

some axiom measures easily adapted from ARM:

for a GCI $C \sqsubseteq D$ define its metrics as follows:

	basic	relativized
Coverage	$\#\,\mathsf{Inst}(C, \mathcal{O})$	$\mathsf{P}(C, \mathcal{O})$
Support	$\#\,\mathsf{Inst}(C \sqcap D, \mathcal{O})$	$\mathsf{P}(C \sqcap D, \mathcal{O})$
Contradiction	$\#\,\mathsf{Inst}(C \sqcap \neg D, \mathcal{O})$	$\mathsf{P}(C \sqcap \neg D, \mathcal{O})$
Assumption	$\#\,\mathsf{Inst}(C, \mathcal{O}) \cap \mathsf{UnKn}(D, \mathcal{O})$	$\ldots$
Confidence		$\mathsf{P}(C \sqcap D, \mathcal{O})/P(C, \mathcal{O})$
Lift		$\mathsf{P}(C \sqcap D, \mathcal{O})/P(C, \mathcal{O})P(D, \mathcal{O})$
$\ldots$		

where $\mathsf{P}(X, \mathcal{O}) = \#\,\mathsf{Ind}(X, \mathcal{O})/\#\,\mathsf{Ind}(\top, \mathcal{O})$

4. Statistically Sound Hypotheses: Example

	A	B	C1	C2	…
Ind1	X	X	X	X	…
…	…	…	…	…	…
Ind180	X	X	X	X	…
Ind181	X	?	X	?	…
…	…	…	…	…	…
Ind200	X	?	X	?	…
Ind201	?	?	?	?	…
…	…	…	…	…	…
Ind400	?	?	?	?	…

	relativized	$A \sqsubseteq B$	$B \sqsubseteq C1$	$B \sqsubseteq C2$
Coverage	P$(C, \mathcal{O})$	200/400	180/400	180/400
Support	P$(C \sqcap D, \mathcal{O})$	180/400	180/400	180/400
Assumption	…	20/400	0	0
Confidence	P$(C \sqcap D, \mathcal{O})/P(C, \mathcal{O})$	180/200	180/180	180/180
Lift	P$(C \sqcap D, \mathcal{O})/P(C, \mathcal{O})P(D, \mathcal{O})$	400/200	400/200	400/180

4. Statistically Sound Hypotheses: Example

	A	B	C1	C2	…
Ind1	X	X	X	X	…
…	…	…	…	…	…
Ind180	X	X	X	X	…
Ind181	X	?	X	?	…
…	…	…	…	…	…
Ind200	X	?	X	?	…
Ind201	?	?	?	?	…
…	…	…	…	…	…
Ind400	?	?	?	?	…

	relativized	$A \sqsubseteq B$	$B \sqsubseteq C1$	$B \sqsubseteq C2$
Coverage	$\mathrm{P}(C, \mathcal{O})$	200/400	180/400	180/400
Support	$\mathrm{P}(C \sqcap D, \mathcal{O})$	180/400	180/400	180/400
Assumption	…	20/400	0	0
Confidence	$\mathrm{P}(C \sqcap D, \mathcal{O}) / P(C, \mathcal{O})$	180/200	180/180	180/180
Lift	$\mathrm{P}(C \sqcap D, \mathcal{O}) / P(C, \mathcal{O}) P(D, \mathcal{O})$	400/200	400/200	400/180

4. Statistically Sound Hypotheses: Example

	A	B	C1	C2	…
Ind1	X	X	X	X	…
…	…	…	…	…	…
Ind180	X	X	X	X	…
Ind181	X	?	X	?	…
…	…	…	…	…	…
Ind200	X	?	X	?	…
Ind201	?	?	?	?	…
…	…	…	…	…	…
Ind400	?	?	?	?	…

	relativized	$A \sqsubseteq B$	$B \sqsubseteq C1$	$B \sqsubseteq C2$
Coverage	P$(C, \mathcal{O})$	200/400	180/400	180/400
Support	P$(C \sqcap D, \mathcal{O})$	180/400	180/400	180/400
Assumption	…	20/400	0	0
Confidence	P$(C \sqcap D, \mathcal{O})/P(C, \mathcal{O})$	180/200	180/180	180/180
Lift	P$(C \sqcap D, \mathcal{O})/P(C, \mathcal{O})P(D, \mathcal{O})$	400/200	400/200	400/180

4. Statistically Sound Hypotheses: Example

	A	B	C1	C2	…
Ind1	X	X	X	X	…
…	…	…	…	…	…
Ind180	X	X	X	X	…
Ind181	X	?	X	?	…
…	…	…	…	…	…
Ind200	X	?	X	?	…
Ind201	?	?	?	?	…
…	…	…	…	…	…
Ind400	?	?	?	?	…

	relativized	$A \sqsubseteq B$	$B \sqsubseteq C1$	$B \sqsubseteq C2$
Coverage	P$(C, \mathcal{O})$	200/400	180/400	180/400
Support	P$(C \sqcap D, \mathcal{O})$	180/400	180/400	180/400
Assumption	…	20/400	0	0
Confidence	P$(C \sqcap D, \mathcal{O}) / P(C, \mathcal{O})$	180/200	180/180	180/180
Lift	P$(C \sqcap D, \mathcal{O}) / P(C, \mathcal{O}) P(D, \mathcal{O})$	400/200	400/200	400/180

4. Statistically Sound Hypotheses: Example

	A	B	C1	C2	…
Ind1	X	X	X	X	…
…	…	…	…	…	…
Ind180	X	X	X	X	…
Ind181	X	?	X	?	…
…	…	…	…	…	…
Ind200	X	?	X	?	…
Ind201	?	?	?	?	…
…	…	…	…	…	…
Ind400	?	?	?	?	…

	relativized	$A \sqsubseteq B$	$B \sqsubseteq C1$	$B \sqsubseteq C2$
Coverage	P$(C, \mathcal{O})$	0.5	0.45	0.45
Support	P$(C \sqcap D, \mathcal{O})$	0.45	0.45	0.45
Assumption	…	0.05	0	0
Confidence	P$(C \sqcap D, \mathcal{O})/P(C, \mathcal{O})$	0.45	1	1
Lift	P$(C \sqcap D, \mathcal{O})/P(C, \mathcal{O})P(D, \mathcal{O})$	2	2	2.22

4. Statistically Sound Hypotheses: Axioms

Oooops!

- make sure we treat GCIs as axioms and not as rules
  - contrapositive!
- so: turn each GCI $X \sqsubseteq Y$ into equivalent $X \sqcup \neg Y \sqsubseteq Y \sqcup \neg X$
  read C below as 'the resulting LHS'…
  read D below as 'the resulting RHS'…

	main	relativized
Coverage	$\#\,\mathsf{Inst}(C, \mathcal{O})$	$\mathsf{P}(C, \mathcal{O})$
Support	$\#\,\mathsf{Inst}(C \sqcap D, \mathcal{O})$	$\mathsf{P}(C \sqcap D, \mathcal{O})$
Contradiction	$\#\,\mathsf{Inst}(C \sqcap \neg D, \mathcal{O})$	$\mathsf{P}(C \sqcap \neg D, \mathcal{O})$
Assumption	$\#\,\mathsf{Inst}(C, \mathcal{O}) \cap \mathsf{UnKn}(D, \mathcal{O})$	…
Confidence		$\mathsf{P}(C \sqcap D, \mathcal{O})/P(C, \mathcal{O})$
Lift		$\mathsf{P}(C \sqcap D, \mathcal{O})/P(C, \mathcal{O})P(D, \mathcal{O})$
…		

4. Statistically Sound Hypotheses: Axioms

Oooops!

- make sure we treat GCIs as axioms and not as rules
  - contrapositive!
- so: turn each GCI $X \sqsubseteq Y$ into equivalen... $\neg X$
  read C below as 'th...
  read D...

Axiom measures are semantically faithful,
i.e., $\text{Ass}(A \sqsubseteq B, \mathcal{O}) = \text{Ass}(\neg B \sqsubseteq \neg A, \mathcal{O})$

		relativized
Coverag...	$\#\,\text{Inst}(C, \mathcal{O})$	$\text{P}(C, \mathcal{O})$
Support	$\#\,\text{Inst}(C \sqcap D, \mathcal{O})$	$\text{P}(C \sqcap D, \mathcal{O})$
Contradiction	$\#\,\text{Inst}(C \sqcap \neg D, \mathcal{O})$	$\text{P}(C \sqcap \neg D, \mathcal{O})$
Assumption	$\#\,\text{Inst}(C, \mathcal{O}) \cap \text{UnKn}(D, \mathcal{O})$	…
Confidence		$\text{P}(C \sqcap D, \mathcal{O})/P(C, \mathcal{O})$
Lift		$\text{P}(C \sqcap D, \mathcal{O})/P(C, \mathcal{O})P(D, \mathcal{O})$
…		

4. Statistically Sound Hypotheses: Axioms

Oooops!

- make sure we treat GCIs as axioms and not as rules
  - contrapositive!
- so: turn each GCI $X \sqsubseteq Y$ into ...
  read C bel...

Axiom measures are **not** semantically faithful, e.g.,

Support$(A \sqsubseteq B, \mathcal{O}) \neq$ Support$(\top \sqsubseteq \neg A \sqcup B, \mathcal{O})$

		relativized
Cove...	...$, \mathcal{O})$	P$(C, \mathcal{O})$
Support	$\#$ Inst$(C \sqcap D, \mathcal{O})$	P$(C \sqcap D, \mathcal{O})$
Contradiction	$\#$ Inst$(C \sqcap \neg D, \mathcal{O})$	P$(C \sqcap \neg D, \mathcal{O})$
Assumption	$\#$ Inst$(C, \mathcal{O}) \cap$ UnKn$(D, \mathcal{O})$	…
Confidence		P$(C \sqcap D, \mathcal{O})/P(C, \mathcal{O})$
Lift		P$(C \sqcap D, \mathcal{O})/P(C, \mathcal{O})P(D, \mathcal{O})$
…		

4. Stat. Sound Hypotheses: Sets of Axioms

Goal: mine small sets of (short) axioms

- more readable
  - close to what people write
- *synergy* between axioms should lead to better quality
- how to measure their qualities?

4. Stat. Sound Hypotheses: Sets of Axioms

Goal: learn small sets of (short) axioms

- more readable
  - close to what people write
- *synergy* between axioms should lead to better quality
- how to measure their qualities?
  - …easy:
    1. rewrite set into single axiom as usual
    2. measure resulting axiom

4. Stat. Sound Hypotheses: Sets of Axioms

$$\text{H1} = \{A \sqsubseteq B, B \sqsubseteq C1\}$$
$$\equiv \{\top \sqsubseteq (\neg A \sqcup B) \sqcap (\neg B \sqcup C1)\}$$

	A	B	C1	C2	...
Ind1	X	X	X	X	...
...	...	...	...	...	...
Ind180	X	X	X	X	...
Ind181	X	?	X	?	...
...	...	...	...	...	...
Ind200	X	?	X	?	...
Ind201	?	?	?	?	...
...	...	...	...	...	...
Ind400	?	?	?	?	...

	$A \sqsubseteq B$	$B \sqsubseteq C1$	$B \sqsubseteq C2$	H1
Coverage	0.5	0.45	0.45	
Support	0.45	0.45	0.45	
Assumption	0.05	0	0	
Confidence	0.45	1	1	
Lift	2	2	2.22	

4. Stat. Sound Hypotheses: Sets of Axioms

$$
\begin{aligned}
\text{H1} &= \{A \sqsubseteq B, B \sqsubseteq C1\} \\
&\equiv \{\top \sqsubseteq (\neg A \sqcup B) \sqcap (\neg B \sqcup C1)\}
\end{aligned}
$$

	A	B	C1	C2	...
Ind1	X	X	X	X	...
...	...	...	...	...	...
Ind180	X	X	X	X	...
Ind181	X	?	X	?	...
...	...	...	...	...	...
Ind200	X	?	X	?	...
Ind201	?	?	?	?	...
...	...	...	...	...	...
Ind400	?	?	?	?	...

	$A \sqsubseteq B$	$B \sqsubseteq C1$	$B \sqsubseteq C2$	H1	
Coverage	0.5	0.45	0.45	1	always!
Support	0.45	0.45	0.45	0.45	min
Assumption	0.05	0	0	0.55	?
Confidence	0.45	1	1	0.45	support!
Lift	2	2	2.22	1	always!

4. Stat. Sound Hypotheses: Sets of Axioms

Goal: learn small sets of (short) axioms

- more readable
  - close to what people write
- *synergy* between axioms should lead to better quality
- how to measure their qualities?
  - sum/average quality of their axioms!

4. Stat. Sound Hypotheses: Sets of Axioms

H1 $= \{A \sqsubseteq B, B \sqsubseteq C1\}$

	A	B	C1	C2	…
Ind1	X	X	X	X	…
…	…	…	…	…	…
Ind180	X	X	X	X	…
Ind181	X	?	X	?	…
…	…	…	…	…	…
Ind200	X	?	X	?	…
Ind201	?	?	?	?	…
…	…	…	…	…	…
Ind400	?	?	?	?	…

	$A \sqsubseteq B$	$B \sqsubseteq C1$		H1	
Coverage	0.5	0.45		0.475?	
Support	0.45	0.45		0.45	
Assumption	0.05	0		0.05	
Confidence	0.45	1		?	
Lift	2	2		?	

4. Stat. Sound Hypotheses: Sets of Axioms

H1 $= \{A \sqsubseteq B, B \sqsubseteq C1\}$

	A	B	C1	C2	…
Ind1	X	X	X	X	…
…	…	…	…	…	…
Ind180	X	X	X	X	…
Ind181	X	?	X	?	…
…	…	…	…	…	…
Ind200	X	?	X	?	…
Ind201	?	?	?	?	…
…	…	…	…	…	…
Ind400	?	?	?	?	…

	$A \sqsubseteq B$	$B \sqsubseteq C1$		H1	
Coverage	0.5	0.45		0.475?	
Support	0.45	0.45		0.45	
Assumption	0.05	0		0.05	
Confidence	0.45	1		?	
Lift	2	2		?	

4. Stat. Sound Hypotheses: Sets of Axioms

H1 $= \{A \sqsubseteq B, B \sqsubseteq C1\}$

H2 $= \{A \sqsubseteq B, B \sqsubseteq C2\}$

	A	B	C1	C2	…
Ind1	X	X	X	X	…
…	…	…	…	…	…
Ind180	X	X	X	X	…
Ind181	X	?	X	?	…
…	…	…	…	…	…
Ind200	X	?	X	?	…
Ind201	?	?	?	?	…
…	…	…	…	…	…
Ind400	?	?	?	?	…

	$A \sqsubseteq B$	$B \sqsubseteq C1$	$B \sqsubseteq C2$	H1	H2
Coverage	0.5	0.45	0.45	0.475?	0.475?
Support	0.45	0.45	0.45	0.45	0.45
Assumption	0.05	0	0	0.05	0.05
Confidence	0.45	1	1	?	?
Lift	2	2	2.22	?	?

4. Stat. Sound Hypotheses: Sets of Axioms

H1 $= \{A \sqsubseteq B, B \sqsubseteq C1\}$

H2 $= \{A \sqsubseteq B, B \sqsubseteq C2\}$

	A	B	C1	C2	…
Ind1	X	X	X	X	…
…	…	…	…	…	…
Ind180	X	X	X	X	…
Ind181	X	?	X	?	…
…	…	…	…	…	…
Ind200	X	?	X	?	…
Ind201	?	?	?	?	…
…	…	…	…	…	…
Ind400	?	?	?	?	…

	$A \sqsubseteq B$	$B \sqsubseteq C1$	$B \sqsubseteq C2$	H1	H2
Coverage	0.5	0.45	0.45	0.475?	0.475?
Support	0.45	0.45	0.45	0.45	0.45
Assumption	0.05	0	0	0.05	0.05
Confidence	0.45	1	1	?	?
Lift	2	2	2.22	?	?

4. Stat. Sound Hypotheses: Sets of Axioms

Goal: learn small sets of (short) axioms

- more readable
  - close to what people write
- *synergy* between axioms should lead to better quality
- how to measure their qualities?
  - observe that a good hypothesis
    - allows us to *shrink* our ABox since it
    - captures *recurring patterns*
    - *(minimum description length induction)*

4. Stat. Sound Hypotheses: Sets of Axioms

Goal: learn small sets of (short) axioms

- more readable
  - close to what people write
- *synergy* between axioms should lead to better quality
- how to measure their qualities?
  - observe that a good hypothesis
    - allows us to *shrink* our ABox since it
    - captures *recurring patterns*
  - use this shrinkage factor to measure a hypothesis'
    - fitness - support by data
    - braveness - number of assumptions

Capturing shrinkage…for fitness

- Fix a finite set of
  - concepts $\mathbb{C}$, closed under negation
  - roles $\mathbb{R}$

Capturing shrinkage…for fitness

- Fix a finite set of
  - concepts $\mathbb{C}$, closed under negation
  - roles $\mathbb{R}$
- Define a *projection*:

$$\begin{aligned}\pi(\mathcal{O},\mathbb{C},\mathbb{R}) = \; & \{C(a) \mid \mathcal{O} \models C(a) \wedge C \in \mathbb{C}\} \cup \\ & \{R(a,b) \mid \mathcal{O} \models C(a) \wedge R \in \mathbb{R}\}\end{aligned}$$

Capturing shrinkage…for fitness

- Fix a finite set of
  - concepts $\mathbb{C}$, closed under negation
  - roles $\mathbb{R}$
- Define a *projection:*

$$\begin{aligned}\pi(\mathcal{O}, \mathbb{C}, \mathbb{R}) = \; & \{C(a) \mid \mathcal{O} \models C(a) \wedge C \in \mathbb{C}\} \cup \\ & \{R(a,b) \mid \mathcal{O} \models C(a) \wedge R \in \mathbb{R}\}\end{aligned}$$

- For an ABox, define its *description length:*

$$\mathrm{dLen}(\mathcal{A}, \mathcal{O}) = \min\{\ell(\mathcal{A}') \mid \mathcal{A}' \cup \mathcal{O} \equiv \mathcal{A} \cup \mathcal{O}\}$$

Capturing shrinkage…for fitness

- Fix a finite set of
  - concepts $\mathbb{C}$, closed under negation
  - roles $\mathbb{R}$
- Define a *projection*:

$$\begin{aligned}\pi(\mathcal{O}, \mathbb{C}, \mathbb{R}) = \; & \{C(a) \mid \mathcal{O} \models C(a) \land C \in \mathbb{C}\} \cup \\ & \{R(a,b) \mid \mathcal{O} \models C(a) \land R \in \mathbb{R}\}\end{aligned}$$

- For an ABox, define its *description length:*

$$\text{dLen}(\mathcal{A}, \mathcal{O}) = \min\{\ell(\mathcal{A}') \mid \mathcal{A}' \cup \mathcal{O} \equiv \mathcal{A} \cup \mathcal{O}\}$$

- Define the *fitness* of a hypothesis *H:*

$$\begin{aligned}\text{fitn}(H, \mathcal{O}, \mathbb{C}, \mathbb{R}) = \; & \text{dLen}(\pi(\mathcal{O}, \mathbb{C}, \mathbb{R}), \mathcal{T}) - \\ & \text{dLen}(\pi(\mathcal{O}, \mathbb{C}, \mathbb{R}), \mathcal{T} \cup H)\end{aligned}$$

Capturing shrinkage…for braveness

- Fix a finite set of
  - concepts $\mathbb{C}$, closed under negation
  - roles $\mathbb{R}$
- Define a *projection*:

$$\begin{aligned}\pi(\mathcal{O}, \mathbb{C}, \mathbb{R}) = \; & \{C(a) \mid \mathcal{O} \models C(a) \wedge C \in \mathbb{C}\} \cup \\ & \{R(a,b) \mid \mathcal{O} \models C(a) \wedge R \in \mathbb{R}\}\end{aligned}$$

Capturing shrinkage…for braveness

- Fix a finite set of
  - concepts $\mathbb{C}$, closed under negation
  - roles $\mathbb{R}$
- Define a *projection*:

$$\pi(\mathcal{O},\mathbb{C},\mathbb{R}) = \{C(a) \mid \mathcal{O} \models C(a) \wedge C \in \mathbb{C}\} \cup \{R(a,b) \mid \mathcal{O} \models C(a) \wedge R \in \mathbb{R}\}$$

- Define a hypothesis' *assumptions*:

$$\mathrm{Ass}(\mathcal{O},H,\mathbb{C},\mathbb{R}) = \pi(\mathcal{O}\cup H,\mathbb{C},\mathbb{R}) \setminus \pi(\mathcal{O},\mathbb{C},\mathbb{R})$$

Capturing shrinkage…for braveness

- Fix a finite set of
  - concepts $\mathbb{C}$, closed under negation
  - roles $\mathbb{R}$
- Define a *projection*:

$$\begin{aligned}\pi(\mathcal{O},\mathbb{C},\mathbb{R}) = \ & \{C(a) \mid \mathcal{O} \models C(a) \wedge C \in \mathbb{C}\} \cup \\ & \{R(a,b) \mid \mathcal{O} \models C(a) \wedge R \in \mathbb{R}\}\end{aligned}$$

- Define a hypothesis' *assumptions*:

$$\mathsf{Ass}(\mathcal{O}, H, \mathbb{C}, \mathbb{R}) = \pi(\mathcal{O} \cup H, \mathbb{C}, \mathbb{R}) \setminus \pi(\mathcal{O}, \mathbb{C}, \mathbb{R})$$

- Define the *braveness* of a hypothesis *H:*

$$\mathsf{brave}(H, \mathcal{O}, \mathbb{C}, \mathbb{R}) = \mathsf{dLen}(\mathsf{Ass}(\mathcal{O}, H, \mathbb{C}, \mathbb{R}), \mathcal{O})$$

Capturing shrinkage…for braveness

- Fix a finite set of
  - concepts $\mathbb{C}$, closed under negation
  - roles $\mathbb{R}$
- Define a *projection*:

$$\pi(\mathcal{O}, \mathbb{C}, \mathbb{R}) \ldots$$

Axiom set measures **are** semantically faithful, i.e.,

$$H \equiv H' \Rightarrow \begin{array}{rcl} \mathsf{fitn}(H, \mathcal{O}, \mathbb{C}, \mathbb{R}) & = & \mathsf{fitn}(H', \mathcal{O}, \mathbb{C}, \mathbb{R}) \\ \mathsf{brave}(H, \mathcal{O}, \mathbb{C}, \mathbb{R}) & = & \mathsf{brave}(H', \mathcal{O}, \mathbb{C}, \mathbb{R}) \end{array}$$

$$\ldots(\mathcal{O} \cup H, \mathbb{C}, \mathbb{R}) \setminus \pi(\mathcal{O}, \mathbb{C}, \mathbb{R})$$

- Define the *braveness* of a hypothesis *H:*

$$\mathsf{brave}(H, \mathcal{O}, \mathbb{C}, \mathbb{R}) = \mathsf{dLen}(\mathsf{Ass}(\mathcal{O}, H, \mathbb{C}, \mathbb{R}), \mathcal{O})$$

4. Stat. Sound Hypotheses: Sets of Axioms

H1 $= \{A \sqsubseteq B, B \sqsubseteq C1\}$

	A	B	C1	C2
Ind1	X	X	X	X
...	...	...	...	...
Ind180	X	X	X	X
Ind181	X	?	X	?
...	...	...	...	...
Ind200	X	?	X	?
Ind201	?	?	?	?
...	...	...	...	...
Ind400	?	?	?	?

$$\mathsf{fitn}(H1, \mathcal{A}, \ldots) =$$
$$\mathsf{dLen}(\pi(\mathcal{A}, \ldots), \emptyset) - \mathsf{dLen}(\pi(\mathcal{A}, \ldots), H1) = 760 - 380 = 380$$

$$\mathsf{brave}(H1, \mathcal{A}, \ldots) = \mathsf{dLen}(\mathsf{Ass}(\mathcal{A}, H1, \ldots), \mathcal{A}) = 20$$

4. Stat. Sound Hypotheses: Sets of Axioms

H1 $= \{A \sqsubseteq B, B \sqsubseteq C1\}$

	A	B	C1	C2
Ind1	X	X	X	X
...	...	...	...	...
Ind180	X	X	X	X
Ind181	X	?	X	?
...	...	...	...	...
Ind200	X	?	X	?
Ind201	?	?	?	?
...	...	...	...	...
Ind400	?	?	?	?

$$\mathsf{fitn}(H1, \mathcal{A}, \ldots) =$$
$$\mathsf{dLen}(\pi(\mathcal{A}, \ldots), \emptyset) - \mathsf{dLen}(\pi(\mathcal{A}, \ldots), H1) = 760 - 380 = 380$$

$$\mathsf{brave}(H1, \mathcal{A}, \ldots) = \mathsf{dLen}(\mathsf{Ass}(\mathcal{A}, H1, \ldots), \mathcal{A}) = 20$$

4. Stat. Sound Hypotheses: Sets of Axioms

H1 $= \{A \sqsubseteq B, B \sqsubseteq C1\}$

H2 $= \{A \sqsubseteq B, B \sqsubseteq C2\}$

	A	B	C1	C2
Ind1	X	X	X	X
...	...	...	...	...
Ind180	X	X	X	X
Ind181	X	?	X	?
...	...	...	...	...
Ind200	X	?	X	?
Ind201	?	?	?	?
...	...	...	...	...
Ind400	?	?	?	?

$$\mathsf{fitn}(H1, \mathcal{A}, \ldots) =$$
$$\mathsf{dLen}(\pi(\mathcal{A}, \ldots), \emptyset) - \mathsf{dLen}(\pi(\mathcal{A}, \ldots), H1) = 760 - 380 = 380$$

$$\mathsf{fitn}(H2, \mathcal{A}, \ldots) =$$
$$\mathsf{dLen}(\pi(\mathcal{A}, \ldots), \emptyset) - \mathsf{dLen}(\pi(\mathcal{A}, \ldots), H2) = 760 - 400 = 360$$

$$\mathsf{brave}(H1, \mathcal{A}, \ldots) = \mathsf{dLen}(\mathsf{Ass}(\mathcal{A}, H1, \ldots), \mathcal{A}) = 20$$

$$\mathsf{brave}(H2, \mathcal{A}, \ldots) = \mathsf{dLen}(\mathsf{Ass}(\mathcal{A}, H2, \ldots), \mathcal{A}) = 40$$

4. Stat. Sound Hypotheses: Sets of Axioms

H1 $= \{A \sqsubseteq B, B \sqsubseteq C1\}$

H2 $= \{A \sqsubseteq B, B \sqsubseteq C2\}$

H1 >> H2

	A	B	C1	C2
Ind1	X	X	X	X
...	...	...	...	...
Ind180	X	X	X	X
Ind181	X	?	X	?
...	...	...	...	...
Ind200	X	?	X	?
Ind201	?	?	?	?
...	...	...	...	...
Ind400	?	?	?	?

$$\mathsf{fitn}(H1, \mathcal{A}, \ldots) =$$
$$\mathsf{dLen}(\pi(\mathcal{A}, \ldots), \emptyset) - \mathsf{dLen}(\pi(\mathcal{A}, \ldots), H1) = 760 - 380 = 380$$

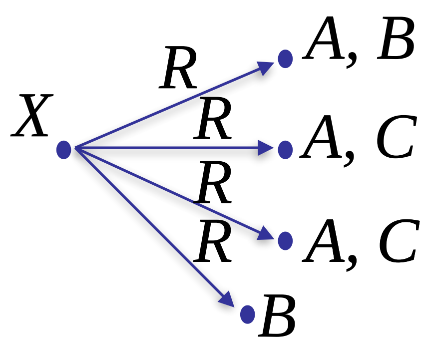
$$\mathsf{fitn}(H2, \mathcal{A}, \ldots) =$$
$$\mathsf{dLen}(\pi(\mathcal{A}, \ldots), \emptyset) - \mathsf{dLen}(\pi(\mathcal{A}, \ldots), H2) = 760 - 400 = 360$$

$$\mathsf{brave}(H1, \mathcal{A}, \ldots) = \mathsf{dLen}(\mathsf{Ass}(\mathcal{A}, H1, \ldots), \mathcal{A}) = 20$$

$$\mathsf{brave}(H2, \mathcal{A}, \ldots) = \mathsf{dLen}(\mathsf{Ass}(\mathcal{A}, H2, \ldots), \mathcal{A}) = 40$$

4. Stat. Sound Hypotheses: Sets of Axioms

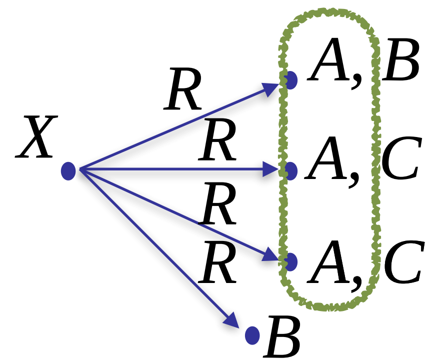
Example: empty TBox, ABox $\mathcal{A}$



$$
\begin{aligned}
\mathsf{fitn}(\{X \sqsubseteq \forall R.A\}, \mathcal{A}, \ldots) &= \mathsf{dLen}(\pi(\mathcal{A}, \ldots), \emptyset) - \\
&\ \mathsf{dLen}(\pi(\mathcal{A}, \ldots), \{X \sqsubseteq \forall R.A\}) \\
&= 12 - 9 \\
&= 3 \\
\mathsf{brave}(\{X \sqsubseteq \forall R.A\}, \mathcal{A}, \ldots) &= \mathsf{dLen}(\mathsf{Ass}(\mathcal{A}, \{X \sqsubseteq \forall R.A\}, \ldots), \mathcal{A}) \\
&= 1
\end{aligned}
$$

4. Stat. Sound Hypotheses: Sets of Axioms

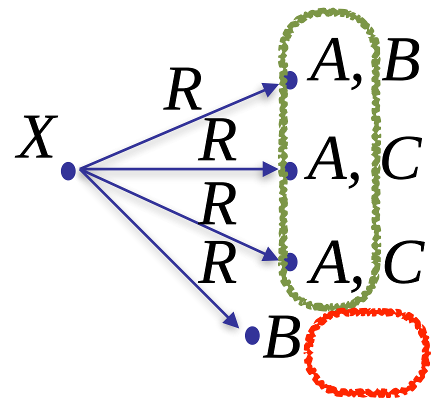
Example: empty TBox, ABox $\mathcal{A}$



$$
\begin{aligned}
\mathsf{fitn}(\{X \sqsubseteq \forall R.A\}, \mathcal{A}, \ldots) &= \mathsf{dLen}(\pi(\mathcal{A}, \ldots), \emptyset) - \\
&\ \mathsf{dLen}(\pi(\mathcal{A}, \ldots), \{X \sqsubseteq \forall R.A\}) \\
&= 12 - 9 \\
&= 3 \\
\mathsf{brave}(\{X \sqsubseteq \forall R.A\}, \mathcal{A}, \ldots) &= \mathsf{dLen}(\mathsf{Ass}(\mathcal{A}, \{X \sqsubseteq \forall R.A\}, \ldots), \mathcal{A}) \\
&= 1
\end{aligned}
$$

4. Stat. Sound Hypotheses: Sets of Axioms

Example: empty TBox, ABox $\mathcal{A}$



$$
\begin{aligned}
\mathsf{fitn}(\{X \sqsubseteq \forall R.A\}, \mathcal{A}, \ldots) &= \mathsf{dLen}(\pi(\mathcal{A}, \ldots), \emptyset) - \\
&\ \mathsf{dLen}(\pi(\mathcal{A}, \ldots), \{X \sqsubseteq \forall R.A\}) \\
&= 12 - 9 \\
&= 3 \\
\mathsf{brave}(\{X \sqsubseteq \forall R.A\}, \mathcal{A}, \ldots) &= \mathsf{dLen}(\mathsf{Ass}(\mathcal{A}, \{X \sqsubseteq \forall R.A\}, \ldots), \mathcal{A}) \\
&= 1
\end{aligned}
$$

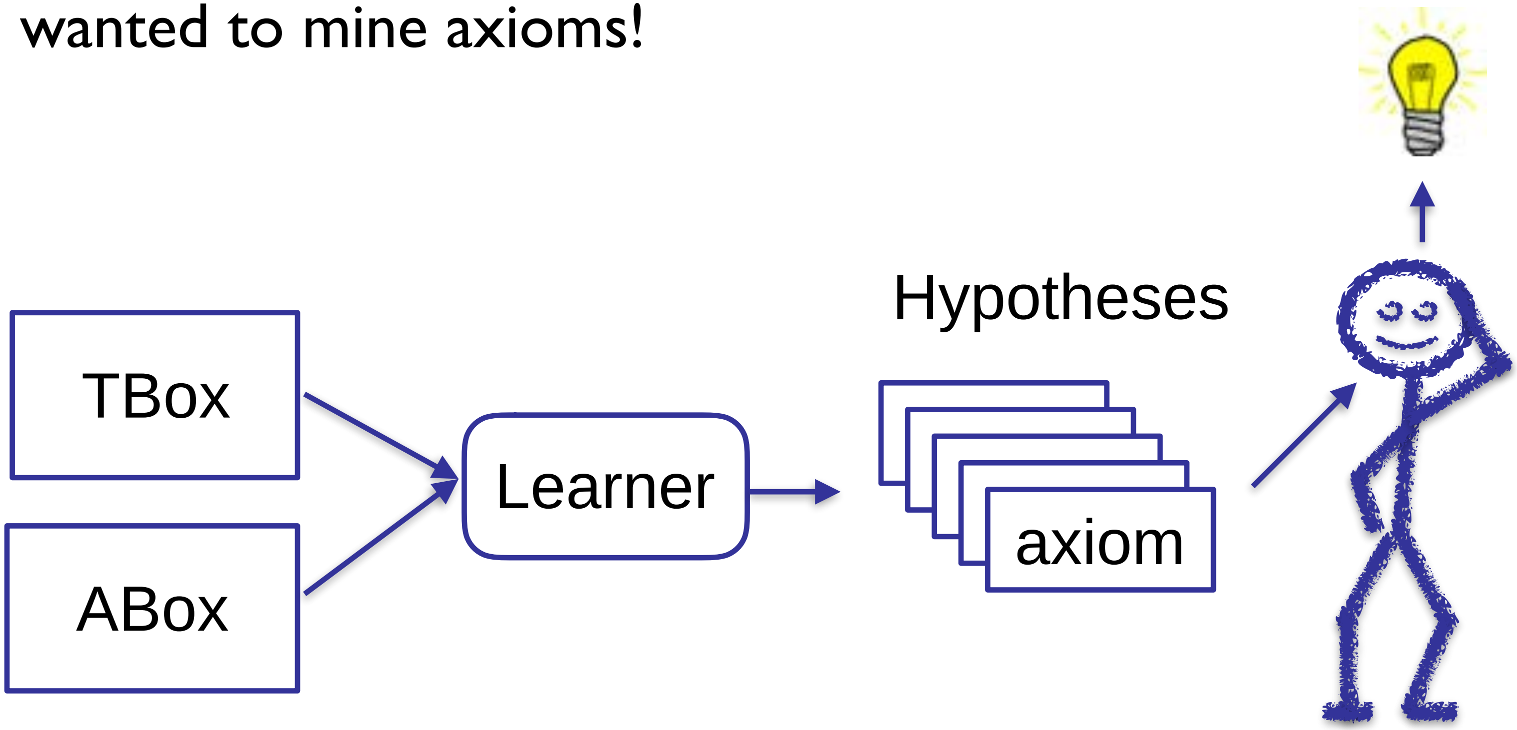
MANCHESTER
1824

The University
of Manchester

phew…

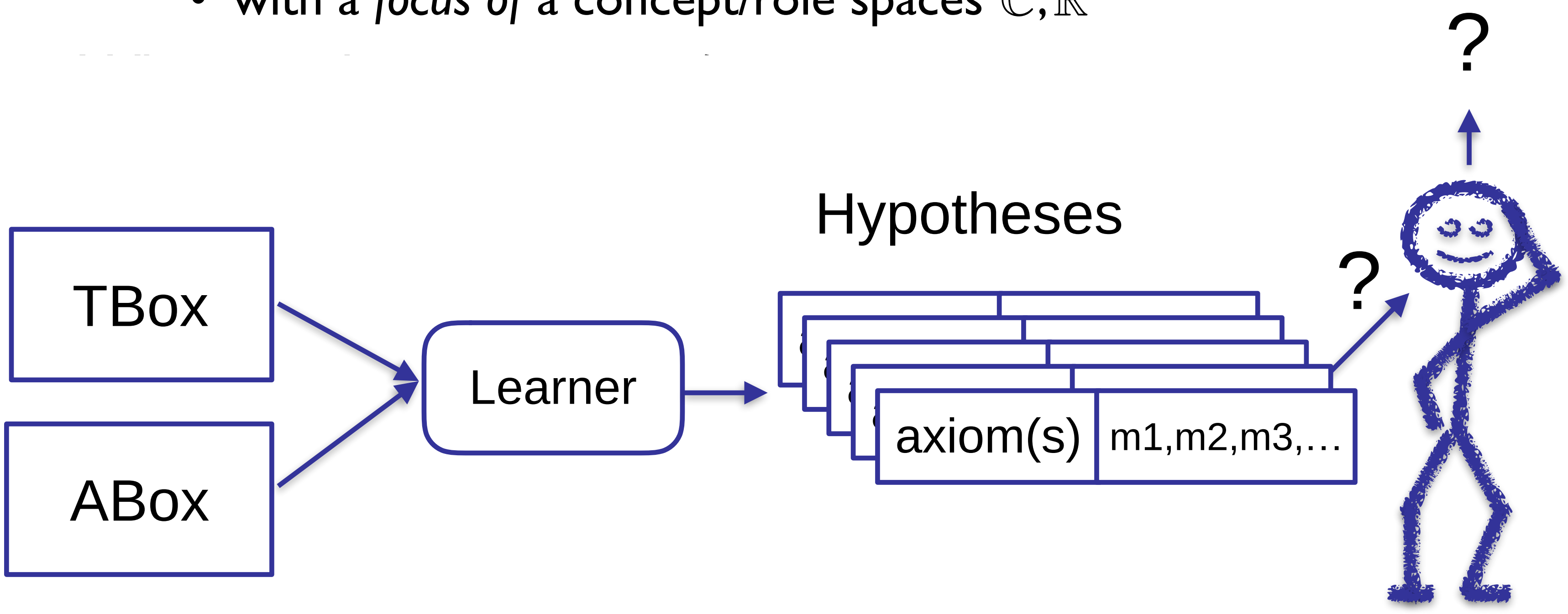
Remember:

we wanted to mine axioms!



So, what have we got?

- (Sets of) axioms as Hypotheses
- Loads of measures to capture
  1. axiom hypothesis' coverage, support, assumption, lift, …
  2. set of axioms hypothesis fitness, braveness
     - with a *focus of* a concept/role spaces $\mathbb{C}, \mathbb{R}$



So, what have we got?

- (Sets of) axioms as Hypotheses
- Loads of measures to capture
  1. axiom hypothesis' coverage, support, assumption, lift, …
  2. set of axioms hypothesis fitness, braveness
     - with a *focus of* a concept/role spaces $\mathbb{C}, \mathbb{R}$
- What are their properties?
  – semantically faithful:

$$\mathcal{O} \models H \Rightarrow \mathsf{Ass}(\mathcal{O}, H, \mathbb{C}, \mathbb{R}) = 0$$

$$H \equiv H' \Rightarrow \mathsf{fitn}(\mathcal{O}, H, \mathbb{C}, \mathbb{R}) = \mathsf{fitn}(\mathcal{O}, H', \mathbb{C}, \mathbb{R})$$

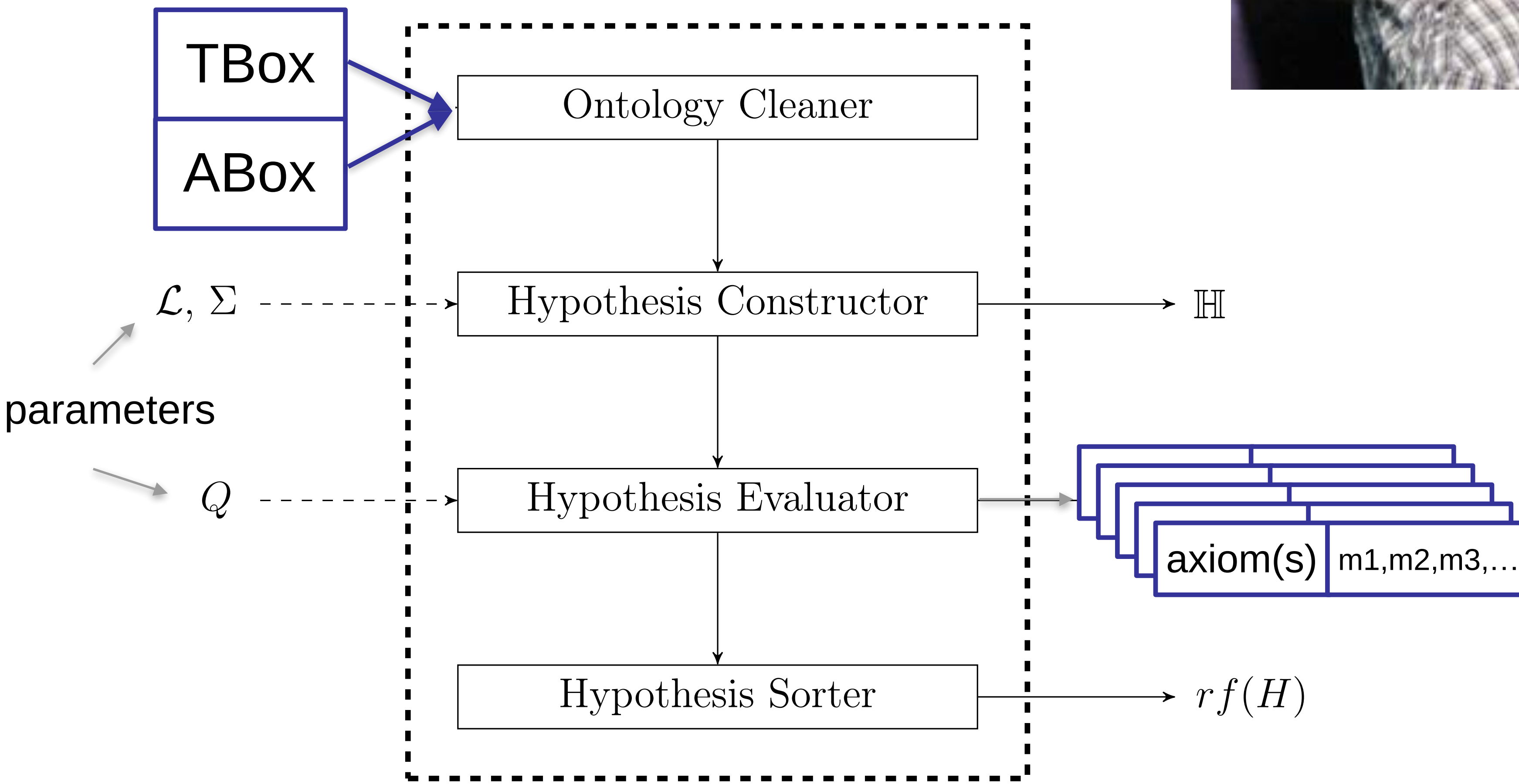
$$\ldots$$

- Can we compute these measure?
  – easy for (1), tricky for (2): $\mathsf{dLen}(\mathcal{A}, \mathcal{O}) = \min\{\ell(\mathcal{A}') \mid \mathcal{A}' \cup \mathcal{O} \equiv \mathcal{A} \cup \mathcal{O}\}$

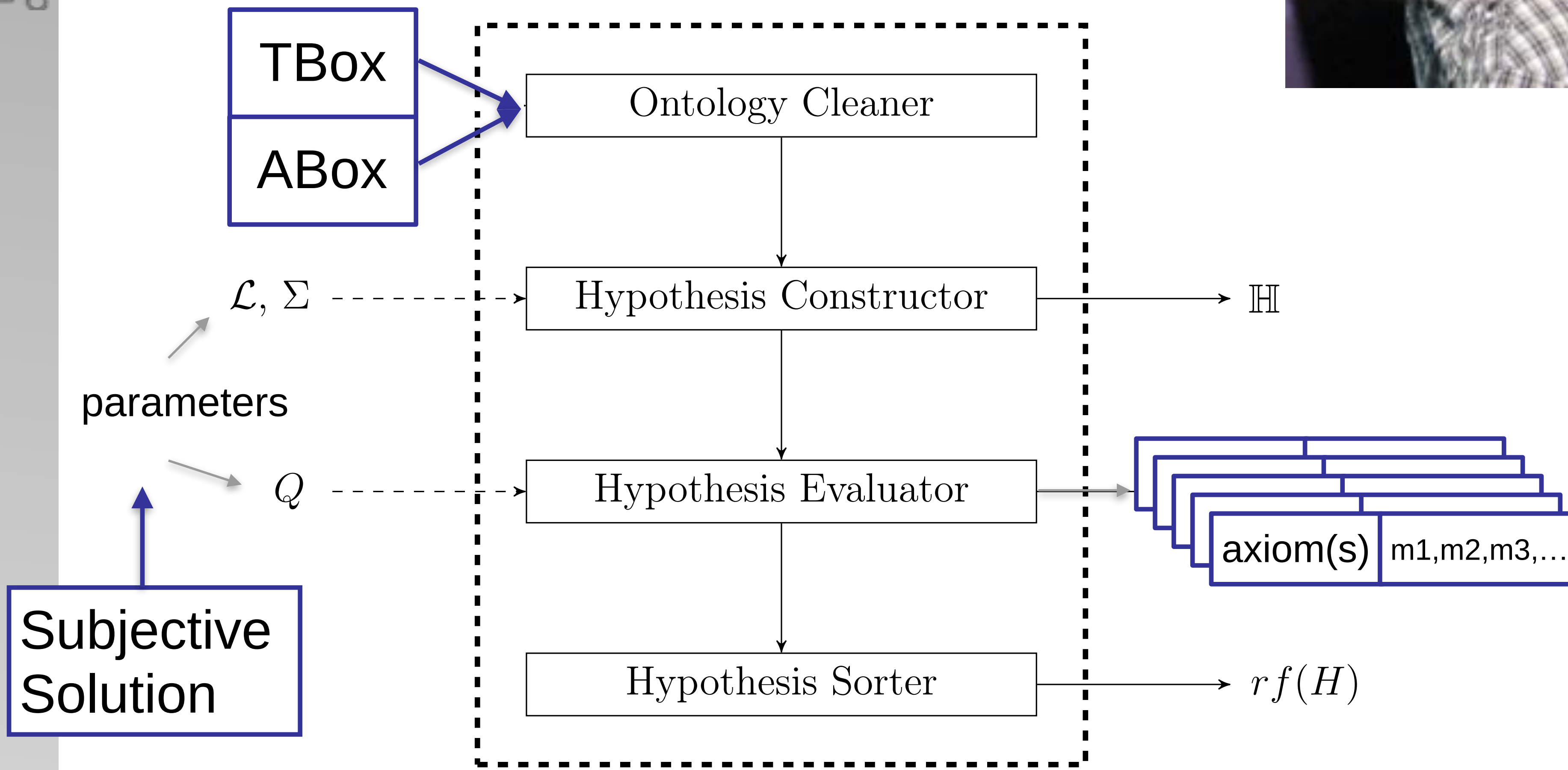
So, what have we got? (2)

- If we can compute measure, how feasible is this?
- If "feasible",
  - do these measures correlate?
  - how independent are they?
- For which DLs & inputs can we create & evaluate hypotheses?
- Which measures indicate *interesting* hypothesis?
- What is the shape for *interesting* hypothesis?
  - are longer/bigger hypotheses better?
- What do we do with them?
  - how do we guide users through these?

Slava implements: DL Miner



Slava implements: DL Miner



DL Miner: Hypothesis Constructor

Easy:

- construct all concepts $C1, C2, \ldots$
  - finitely many thanks to language bias $\mathcal{L}$
- check for each $Ci \sqsubseteq Cj$ whether it's logically ok:
  - $\mathcal{O} \cup \{Ci \sqsubseteq Cj\} \not\models \top \sqsubseteq \bot$
  - $\mathcal{O} \not\models Ci \sqsubseteq Cj$

  if yes, add it to $\mathbb{H}$
- remove redundant hypotheses from $\mathbb{H}$

DL Miner: Hypothesis Constructor

Easy:

- construct all concepts $C1, C2$
  - finite
- check fo
  - $\mathcal{O} \cup \{$
  - $\mathcal{O} \not\models C$

  if yes, add
- remove re

Bonkers!

Even for $\mathcal{EL}$,

100 concept/role names

4 max length of concepts Ci

~100,000,000 concepts Ci

~100,000,000^2 GCIs to test

DL Miner: Hypothesis Constructor

Easy:

- construct all concepts $C1, C2$
  - finite
- check f
  - $\mathcal{O} \cup$
  - $\mathcal{O} \not\models$

  if yes, add
- remove

Bonkers!

Even for $\mathcal{EL}$,

n concept/role names

k max length of concepts Ci

n^k concepts Ci

n^{2k} GCIs to test

DL Miner: Hypothesis Constructor

Use a *refinement operator* to build Ci *informed by* ABox

- used in concept learning, conceptual blending

- Given a logic $\mathcal{L}$, define a *refinement operator* as
  - a function $\rho : \mathsf{Conc}(\mathcal{L}) \mapsto \mathcal{P}(\mathsf{Conc}(\mathcal{L}))$ such that, for each $C \in \mathcal{L}, C' \in \rho(C) : C' \sqsubseteq C$
- A refinement operator is
  - *proper* if, for all $C \in \mathcal{L}, C' \in \rho(C) : C' \not\equiv C$
  - *complete* if, for all $C, C' \in \mathcal{L}$: if $C' \sqsubsetneq C$ then there is some $n, C'' \equiv C$ with $C' \in \rho^n(C'')$
  - *suitable* if, for all $C \in \mathcal{L}$ there is $n, C' \in \rho^n(\top)$: $C' \equiv C$ and $\ell(C') \leqslant \ell(C)$

DL Miner: Hypothesis Constructor

Use a *refinement operator* to build Ci *informed by* ABox

– used in concept learning, conceptual blending

- Given a logic $\mathcal{L}$, define a *refinement operator* as
  - a function $\rho : \mathrm{Conc}(\mathcal{L}) \mapsto \mathcal{P}(C$
    for each C
- A r
  - $(C) : C' \not\equiv C$
  - c if, for all $C, C' \in \mathcal{L}$: if $C' \sqsubsetneq C$
    then there is some $n, C'' \equiv C$
    with $C' \in \rho^n(C'')$
  - *suitable* if, for all $C \in \mathcal{L}$ there is $n, C' \in \rho^n(\top)$: $C' \equiv C$ and $\ell(C') \leqslant \ell(C)$

Great: there are known refinement operators (proper, complete, suitable,…) for $\mathcal{ALC}$ [LehmHitzler2010]

DL Miner: Concept Constructor

Algorithm 8 DL-APRIORI $(\mathcal{O}, \Sigma, \mathcal{DL}, \ell_{max}, p_{min})$

inputs
  $\mathcal{O} := \mathcal{T} \cup \mathcal{A}$: an ontology
  Σ: a finite set of terms such that $\top \in \Sigma$
  $\mathcal{DL}$: a DL for concepts
  ℓ_{max}: a maximal length of a concept such that $1 \leq \ell_{max} < \infty$
  p_{min}: a minimal concept support such that $0 < p_{min} \leq |in(\mathcal{O})|$
outputs
  $\mathbb{C}$: the set of suitable concepts
do
  $\mathbb{C} \leftarrow \emptyset$ % initialise the final set of suitable concepts
  $\mathbb{D} \leftarrow \{\top\}$ % initialise the set of concepts yet to be specialised
  $\rho \leftarrow getOperator(\mathcal{DL})$ % initialise a suitable operator ρ for $\mathcal{DL}$
  while $\mathbb{D} \neq \emptyset$ **do**
    $C \leftarrow pick(\mathbb{D})$ % pick a concept C to be specialised
    $\mathbb{D} \leftarrow \mathbb{D} \setminus \{C\}$ % remove C from the concepts to be specialised
    $\mathbb{C} \leftarrow \mathbb{C} \cup \{C\}$ % add C to the final set
    $\rho_C \leftarrow specialise(C, \rho, \Sigma, \ell_{max})$ % specialise C using ρ
    $\mathbb{D}_C \leftarrow \{D \in urc(\rho_C) \mid \nexists D' \in \mathbb{C} \cup \mathbb{D} : D' \equiv D\}$ % discard variations
    $\mathbb{D} \leftarrow \mathbb{D} \cup \{D \in \mathbb{D}_C \mid p(D, \mathcal{O}) \geq p_{min}\}$ % add suitable specialisations
  end while
  return $\mathbb{C}$

DL Miner: Concept Constructor

Algorithm 8 DL-APRIORI $(\mathcal{O}, \Sigma, \mathcal{DL}, \ell_{max}, p_{min})$

inputs
- $\mathcal{O} := \mathcal{T} \cup \mathcal{A}$: an ontology
- Σ: a finite set of terms such that $\top \in \Sigma$
- $\mathcal{DL}$: a DL for concepts
- ℓ_{max}: a maximal length of a concept such that $1 \leq \ell_{max} < \infty$
- p_{min}: a minimal concept support such that $0 < p_{min} \leq |in(\mathcal{O})|$

outputs
- $\mathbb{C}$: the set of suitable concepts

do
- $\mathbb{C} \leftarrow \emptyset$ % initialise the final set of suitable concepts
- $\mathbb{D} \leftarrow \{\top\}$ % initialise the set of concepts yet to be specialised
- $\rho \leftarrow getOperator(\mathcal{DL})$ % initialise a suitable operator ρ for $\mathcal{DL}$
- **while** $\mathbb{D} \neq \emptyset$ **do**
  - $C \leftarrow pick(\mathbb{D})$ % pick a concept C to be specialised
  - $\mathbb{D} \leftarrow \mathbb{D} \backslash \{C\}$ % remove C from the concepts to be specialised
  - $\mathbb{C} \leftarrow \mathbb{C} \cup \{C\}$ % add C to the final set
  - $\rho_C \leftarrow specialise(C, \rho, \Sigma, \ell_{max})$ % specialise C using ρ
  - $\mathbb{D}_C \leftarrow \{D \in urc(\rho_C) \mid \nexists D' \in \mathbb{C} \cup \mathbb{D} : D' \equiv D\}$ % discard variations
  - $\mathbb{D} \leftarrow \mathbb{D} \cup \{D \in \mathbb{D}_C \mid p(D, \mathcal{O}) \geq p_{min}\}$ % add suitable specialisations
- **end while**
- **return** $\mathbb{C}$

specialise concepts only if they have $\geqslant \ell_{\max}$ instances!

DL Miner: Concept Constructor

Don't even construct most of the n^k concepts Ci

specialise concepts only if they have $\geqslant \ell_{\max}$ instances!

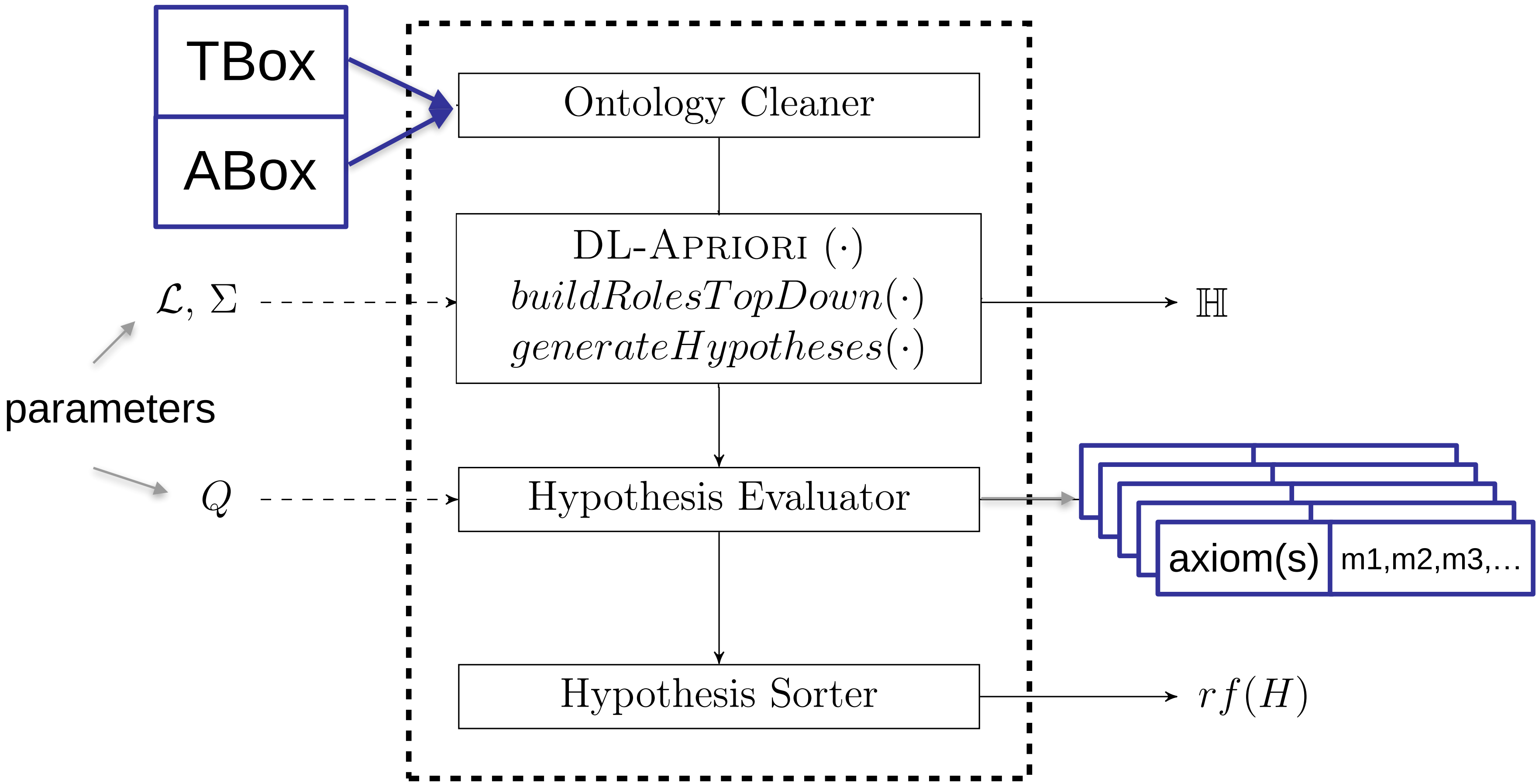
Algorithm 8 DL-Apriori $(\mathcal{O}, \Sigma, \mathcal{DL}, \ell_{max}, p_{min})$

inputs
- $\mathcal{O} := \mathcal{T} \cup \mathcal{A}$: an ontology
- Σ: a finite set of terms such that $\top \in \Sigma$
- $\mathcal{DL}$: a DL for concepts
- ℓ_{max}: a maximal length of a concept such that $1 \leq \ell_{max} < \infty$
- p_{min}: a minimal concept support such that $0 < p_{min} \leq |in(\mathcal{O})|$

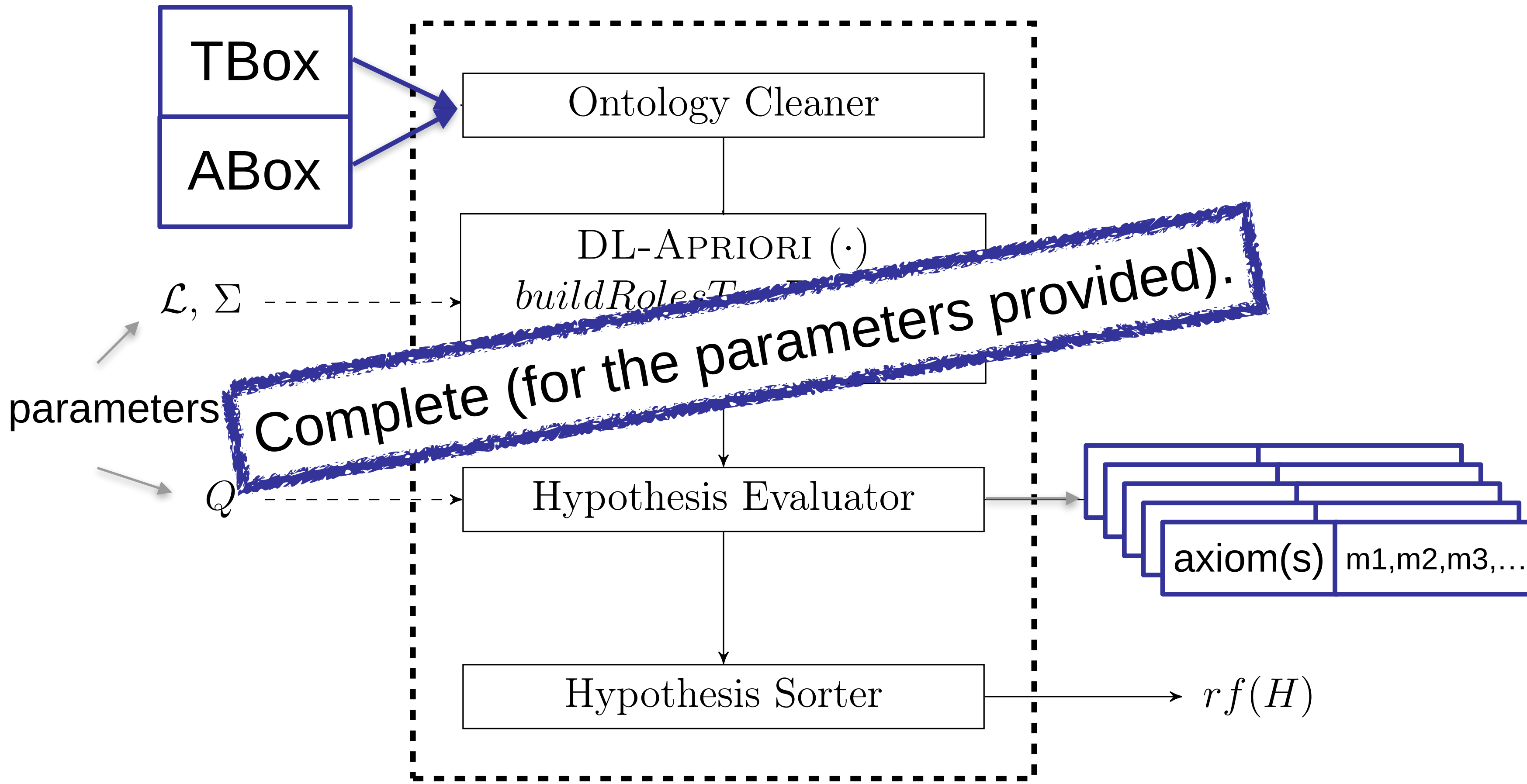
outputs
- $\mathbb{C}$: the set of suitable concepts

do
- $\mathbb{C} \leftarrow \emptyset$ % initialise the final set of suitable concepts
- $\mathbb{D} \leftarrow \{\top\}$ % initialise the set of concepts yet to be specialised
- $\rho \leftarrow getOperator(\mathcal{DL})$ % initialise a suitable operator ρ for $\mathcal{DL}$
- **while** $\mathbb{D} \neq \emptyset$ **do**
  - $C \leftarrow pick(\mathbb{D})$ % pick a concept C to be specialised
  - $\mathbb{D} \leftarrow \mathbb{D} \backslash \{C\}$ % remove C from the concepts to be specialised
  - $\mathbb{C} \leftarrow \mathbb{C} \cup \{C\}$ % add C to the final set
  - $\rho_C \leftarrow specialise(C, \rho, \Sigma, \ell_{max})$ % specialise C using ρ
  - $\mathbb{D}_C \leftarrow \{D \in urc(\rho_C) \mid \nexists D' \in \mathbb{C} \cup \mathbb{D} : D' \equiv D\}$ % discard variations
  - $\mathbb{D} \leftarrow \mathbb{D} \cup \{D \in \mathbb{D}_C \mid p(D, \mathcal{O}) \geq p_{min}\}$ % add suitable specialisations
- **end while**
- **return** $\mathbb{C}$

Slava implements: DL Miner



Slava implements: DL Miner



DL Miner: Hypothesis Evaluator

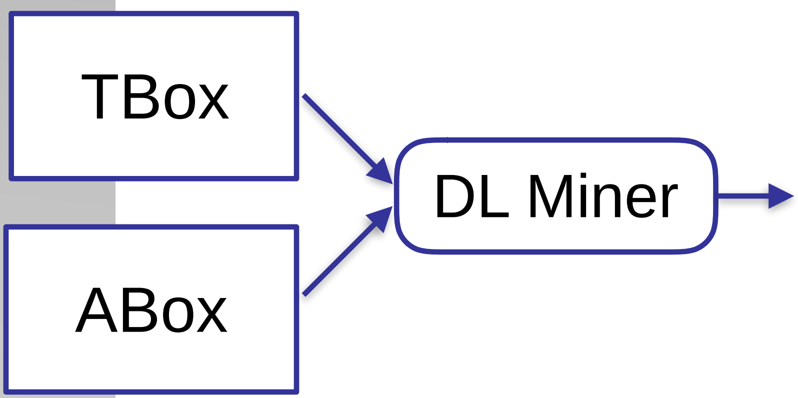
- Relatively straightforward for axiom measures
  - hard test case for instance retrieval
- Hard for set-of-axiom measures (fitness & braveness)
  - due to $\mathrm{dLen}(\mathcal{A}, \mathcal{O}) = \min\{\ell(\mathcal{A}') \mid \mathcal{A}' \cup \mathcal{O} \equiv \mathcal{A} \cup \mathcal{O}\}$
  - DL Miner implements an approximation that
    - identifies redundant assertions in ABox
      $\mathrm{dLen}^*(\mathcal{A}, \mathcal{O}) = \ell(\mathcal{A}) - \ell(\mathrm{Redundt}(\mathcal{A}, \mathcal{O}))$
    - does consider 1-step interactions between individuals
    - ignores 'longer' interactions
    - underestimates fitness, overestimates braveness
  - great test case for incremental reasoning: Pellet!

DL Miner: Hypothesis Sorter

- Last step in DL Miner's workflow
- Easy:
  - throw away all hypotheses that are dominated by another one
  - i.e., compute the *Pareto front* wrt the measures provided

DL Miner: Example

Given a Kinship Ontology,[1] it mines 536 Hs with confidence above 0.9, e.g.



$$\begin{aligned}
Woman \sqcap \exists hasChild.\top &\sqsubseteq Mother \\
Man \sqcap \exists hasChild.\top &\sqsubseteq Father \\
\exists hasChild.\top &\sqsubseteq \exists marriedTo.\top \\
\exists marriedTo.\top &\sqsubseteq \exists hasChild.\top \\
\exists marriedTo.Woman &\sqsubseteq Man \\
\exists marriedTo.Mother &\sqsubseteq Father \\
Father &\sqsubseteq \exists marriedTo.(\exists hasChild.\top) \\
Mother &\sqsubseteq \exists marriedTo.(\exists hasChild.\top) \\
\exists hasChild.\top &\sqsubseteq Mother \sqcup Father \\
\exists hasChild.\top &\sqsubseteq Man \sqcup Woman \\
\exists hasChild.\top &\sqsubseteq Father \sqcup Woman
\end{aligned}$$

1. adapted from
UCI Machine Learning Repository

DL Miner: Example

Given a Kinship Ontology,[1] it mines 536 Hs with confidence above 0.9, e.g.

$$Woman \sqcap \exists hasChild.\top \sqsubseteq Mother$$

$$Man \sqcap \exists hasChild.\top \sqsubseteq Father$$

$$\exists ha \ldots marriedTo.\top$$

$$\ldots asChild.\top$$

$$\ldots an \sqsubseteq Man$$

$$\exists marriedTo.Mother \sqsubseteq Father$$

$$Father \sqsubseteq \exists marriedTo.(\exists hasChild.\top)$$

$$Mother \sqsubseteq \exists marriedTo.(\exists hasChild.\top)$$

$$\exists hasChild.\top \sqsubseteq Mother \sqcup Father$$

$$\exists hasChild.\top \sqsubseteq Man \sqcup Woman$$

$$\exists hasChild.\top \sqsubseteq Father \sqcup Woman$$

TBox

ABox

DL

Great - it works really well on a toy ontology!

1. adapted from UCI Machine Learning Repository

Still: many open questions

- If we can compute measure, how feasible is this?
- If “feasible”,
  - do these measures correlate?
  - how independent are they?
- For which DLs & inputs can we create & evaluate hypotheses?
- Which measures indicate *interesting* hypothesis?
- What is the shape of *interesting* hypothesis?
  - are longer/bigger hypotheses better?
- What do we do with them?
  - how do we guide users through these?

Design, run, analyse experiments

Design, run, analyse experiments

- A corpus - or two:
  1. handpicked corpus from related work: 16 ontologies
  2. principled one:
     - All BioPortal ontologies with >= 100 individuals and >= 100 RAs 21 ontologies

Design, run, analyse experiments

- A corpus - or two:
  1. handpicked corpus from related work: 16 ontologies
  2. principled one:
     - All BioPortal ontologies with >= 100 individuals and >= 100 RAs 21 ontologies
- Settings for hypothesis parameters:
  - $\mathcal{L}$ is $\mathcal{SHI}$
    - RIAs with inverse, composition
  - minsupport = 10
  - max concept length in GCIs = 4

Design, run, analyse experiments

- A corpus - or two:
  1. handpicked corpus from related work: 16 ontologies
  2. principled one:
     - All BioPortal ontologies with >= 100 individuals and >= 100 RAs 21 ontologies
- Settings for hypothesis parameters:
  - $\mathcal{L}$ is $\mathcal{SHI}$
    - RIAs with inverse, composition
  - minsupport = 10
  - max concept length in GCIs = 4
- generate & evaluate up to 500 hypotheses per ontology

Design, run, analyse experiments

- What kind of axioms do people write?
  - re. readability of hypotheses:
  - what kind of axioms should we roughly aim for?

Use of DL constructors in Bioportal - Taxonomies

DL constructor	C	$\exists R.C$	$C \sqcap D$	$\forall R.C$	$C \sqcup D$	$\neg C$
Axioms, %	99.73	67.82	1.15	0.46	0.09	0.01

Length & role depth of axioms in Bioportal - Taxonomies

	mean	**mode**	**5%**	**25%**	**50%**	**75%**	**95%**	**99%**	**99.9%**
length	2.63	3	2	2	3	3	3	3	5
depth	0.69	1	0	0	1	1	1	1	3

Design, run, analyse experiments

- What kind of axioms do people write?
  - re. readability of hypotheses:
  - what kind of axioms should we roughly aim for?

Use of DL constructors in Bioportal - T[illegible]

DL constr[illegible]	[illegible]	D	$\neg C$
Axioms [illegible]	[illegible]	[illegible]	0.01

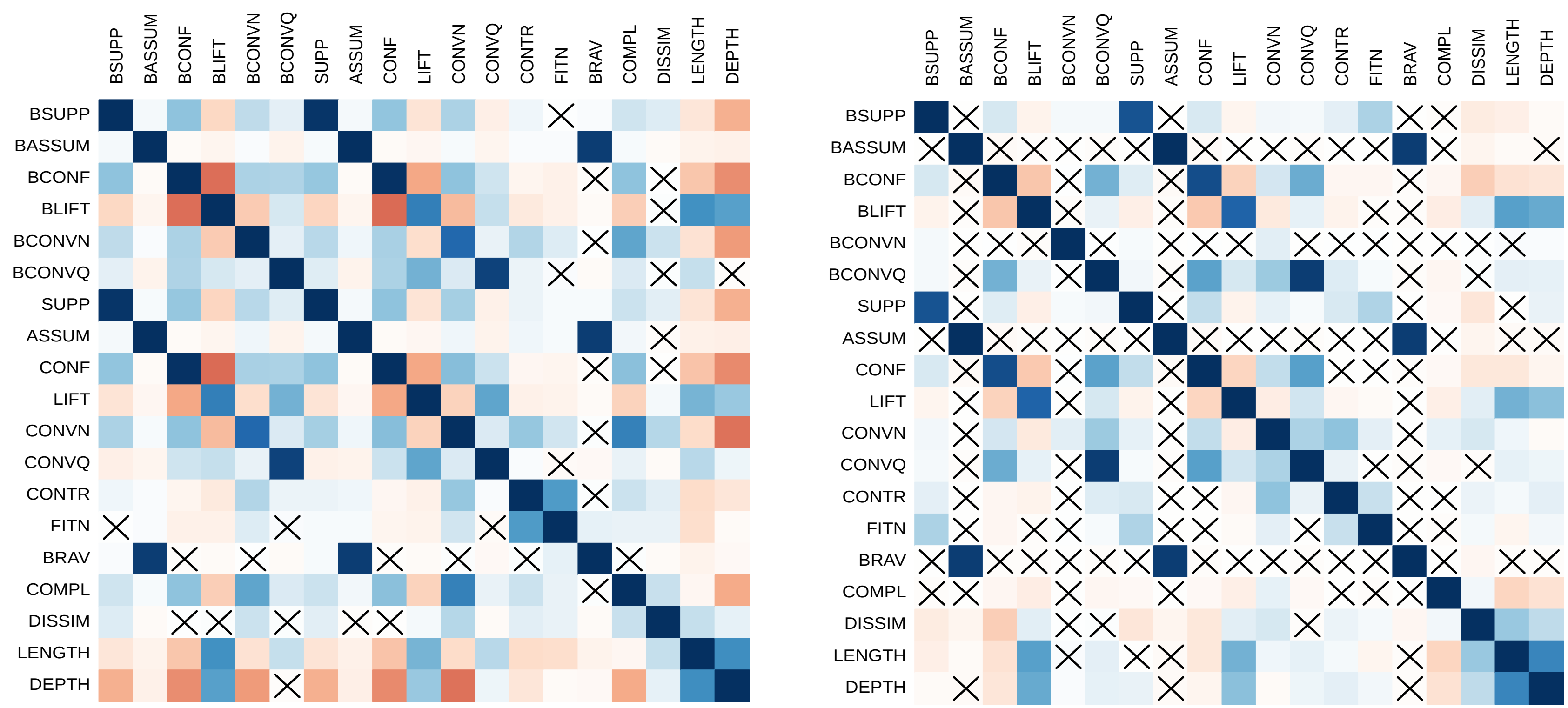
Restricting length of concepts in axioms to 4 (axioms to 8) is fine!

Length & role [illegible] xioms in Bioportal - Taxonomies

	mean	mode	5%	25%	50%	75%	95%	99%	99.9%
length	2.63	3	2	2	3	3	3	3	5
depth	0.69	1	0	0	1	1	1	1	3

Design, run, analyse experiments

How do the measures correlate?

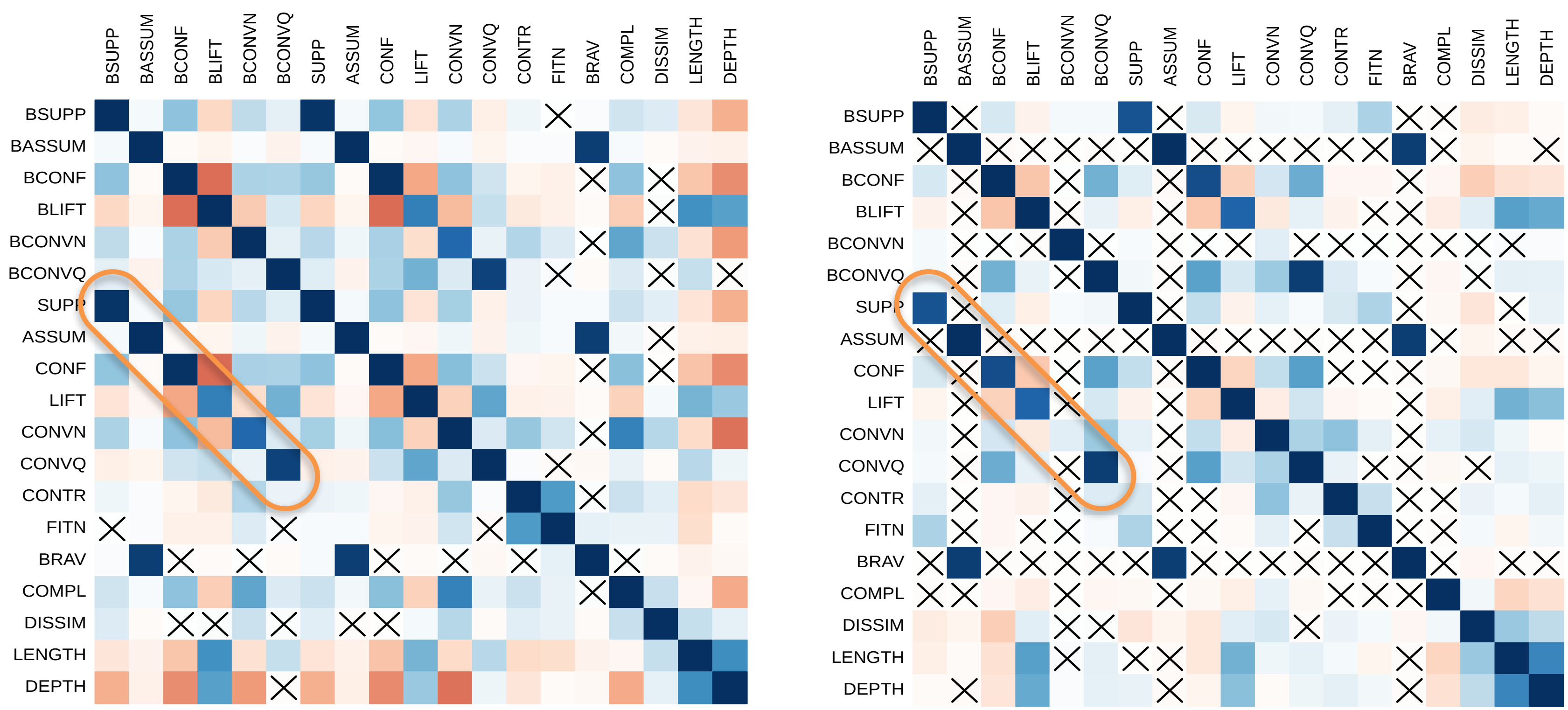


(a) Handpicked corpus

(b) Principled corpus

Design, run, analyse experiments

How do the measures correlate?

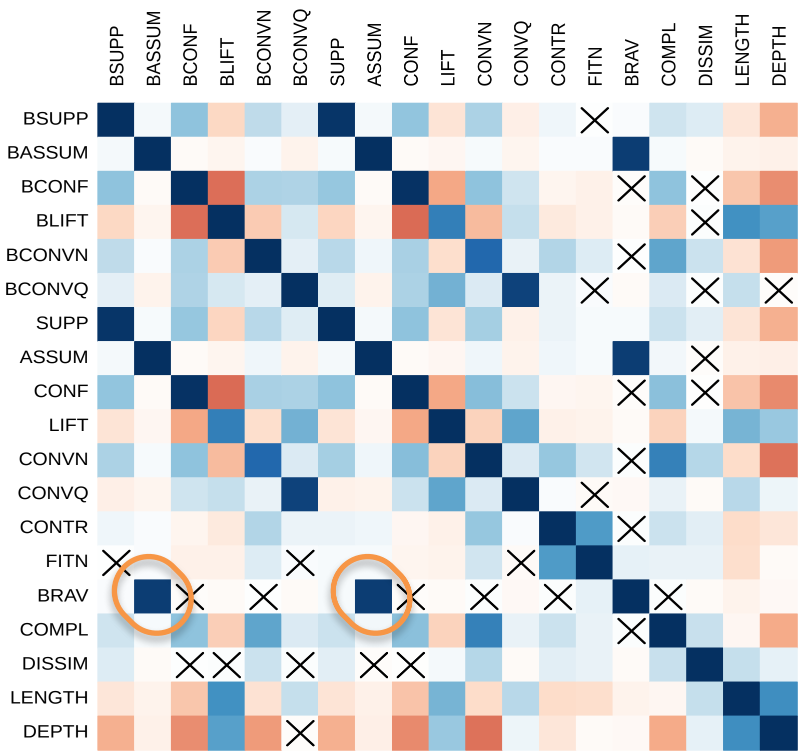


(a) Handpicked corpus

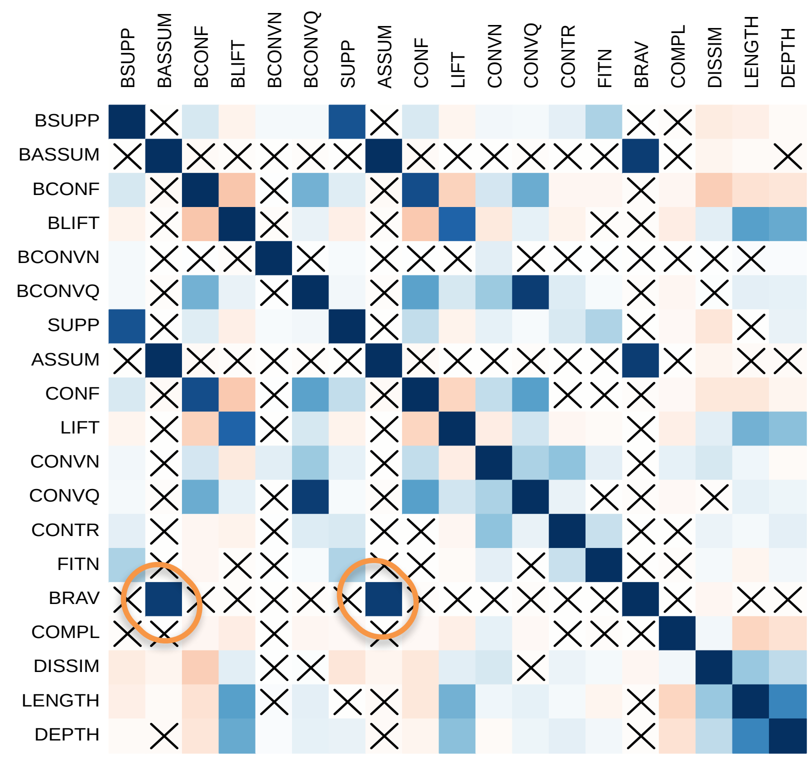
(b) Principled corpus

Design, run, analyse experiments

How do the measures correlate?



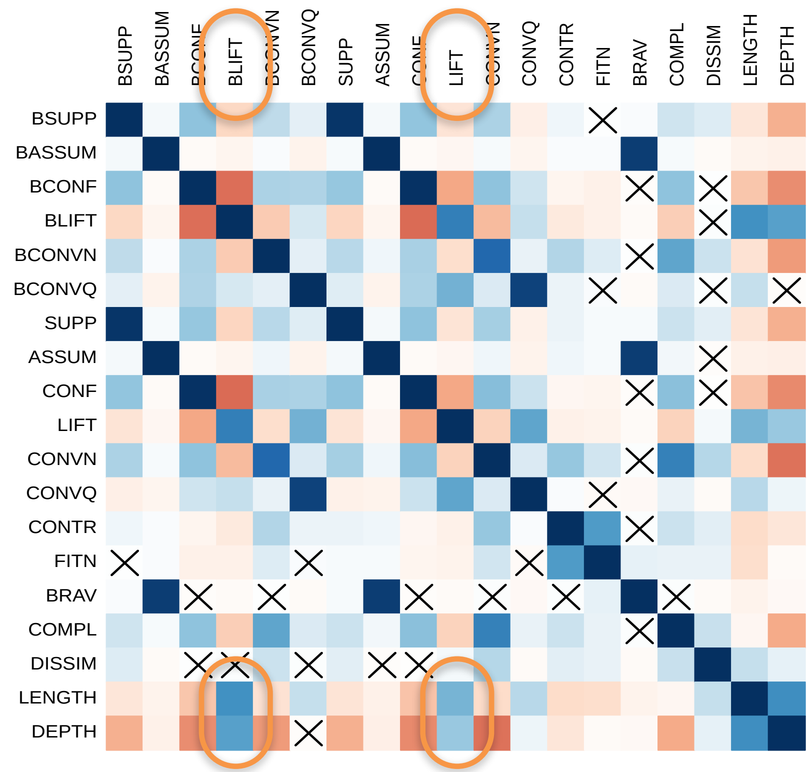
(a) Handpicked corpus



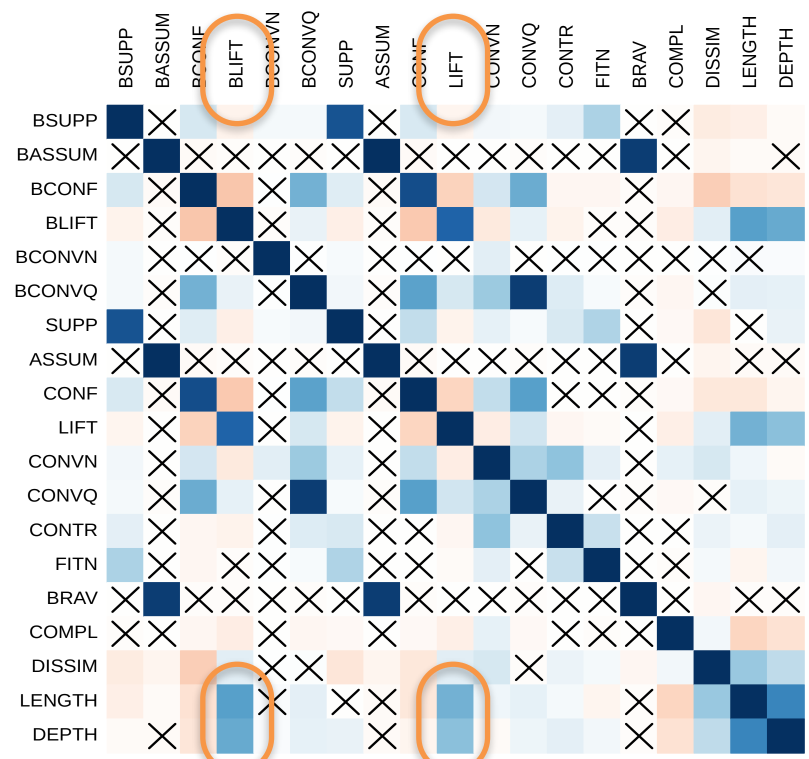
(b) Principled corpus

Design, run, analyse experiments

How do the measures correlate?



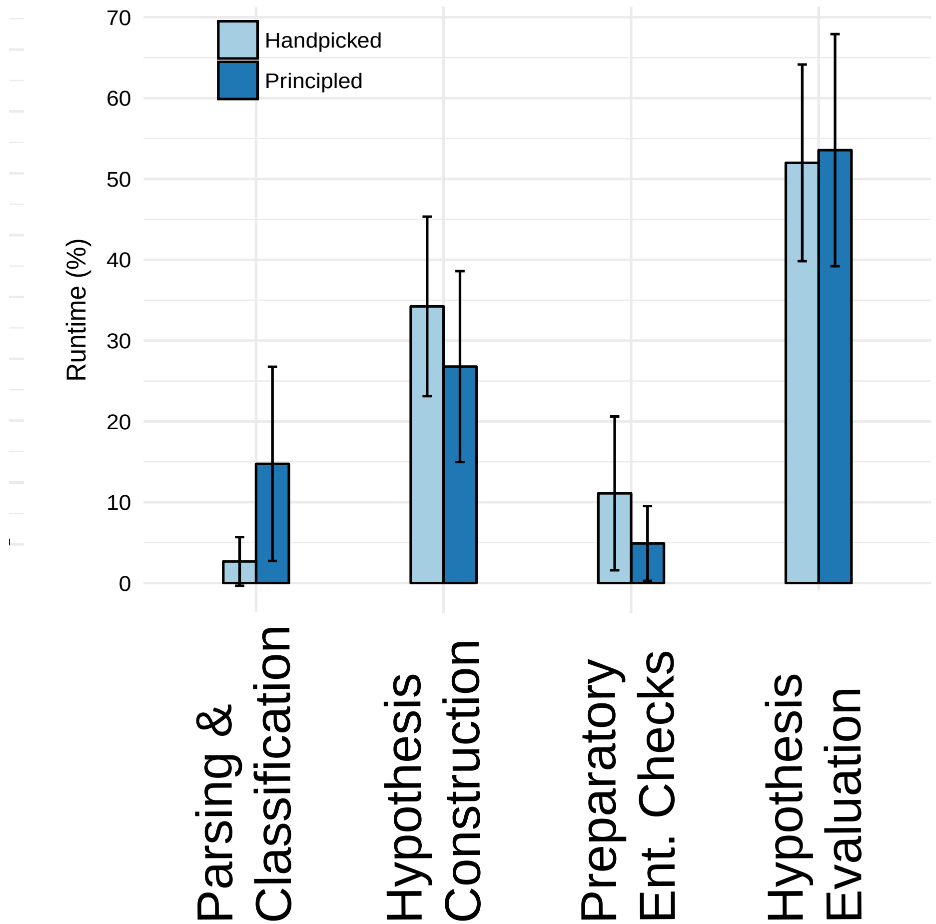
(a) Handpicked corpus



(b) Principled corpus

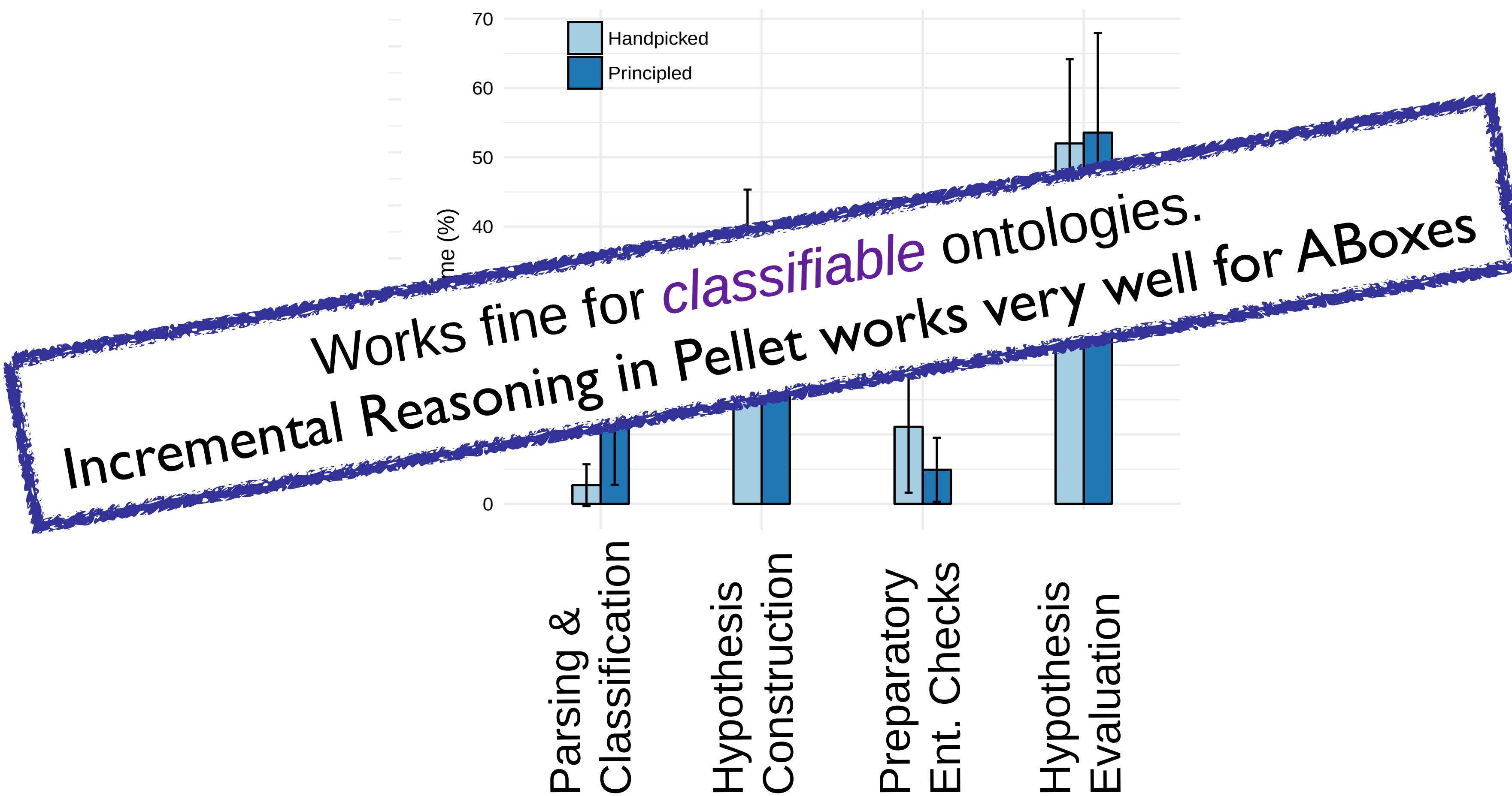
Design, run, analyse experiments

How feasible is hypothesis mining?



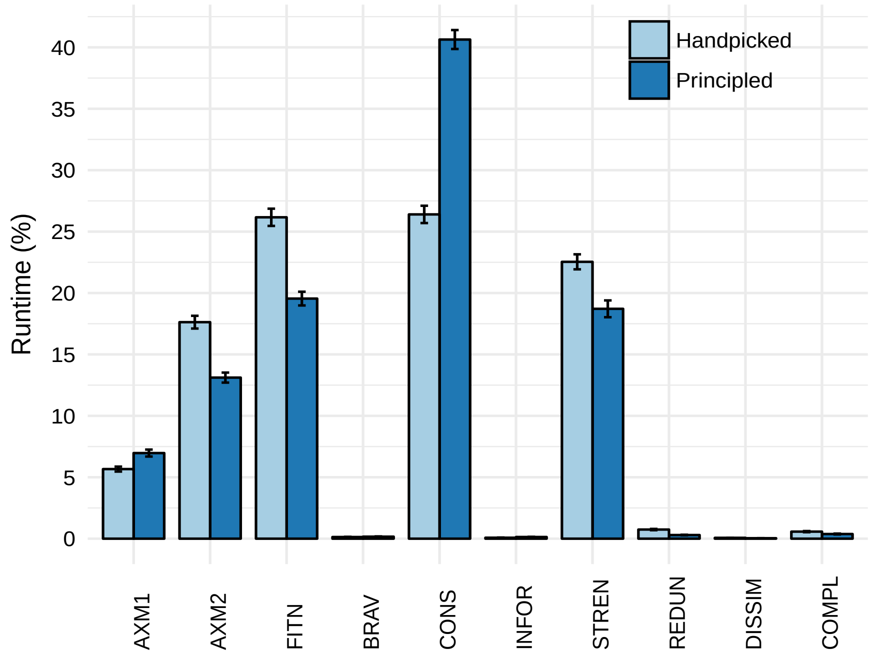
Design, run, analyse experiments

How feasible is hypothesis mining?



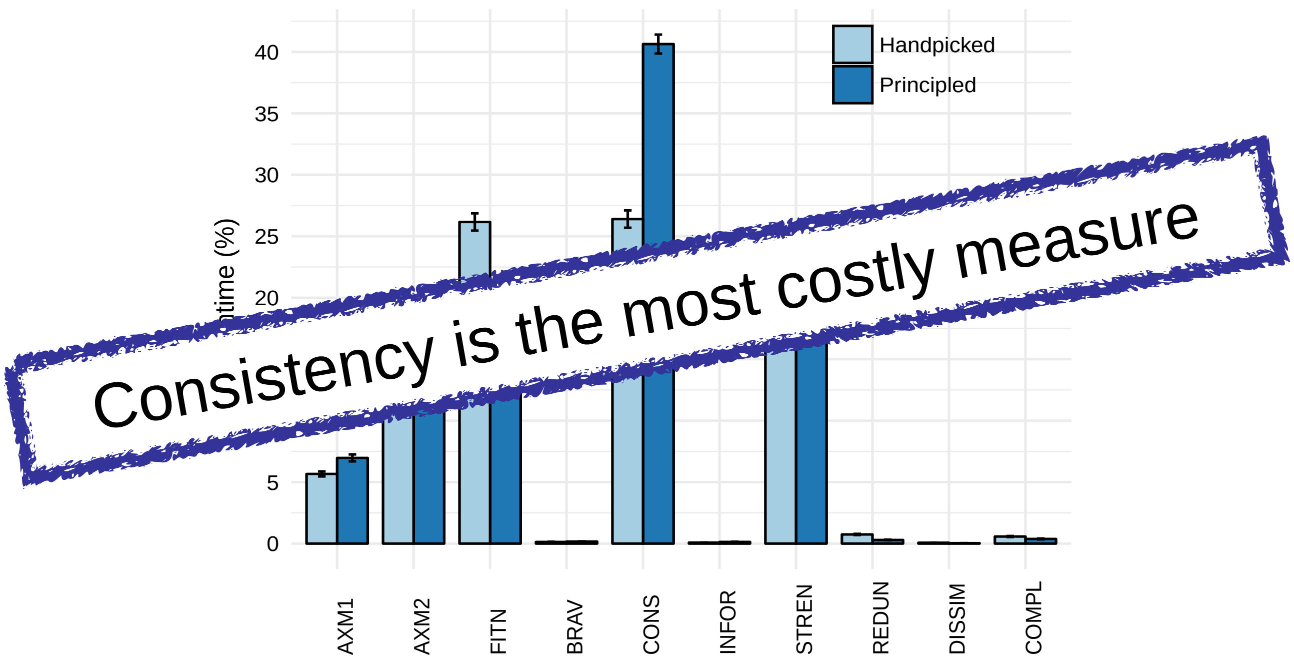
Design, run, analyse experiments

How costly are the different measures?

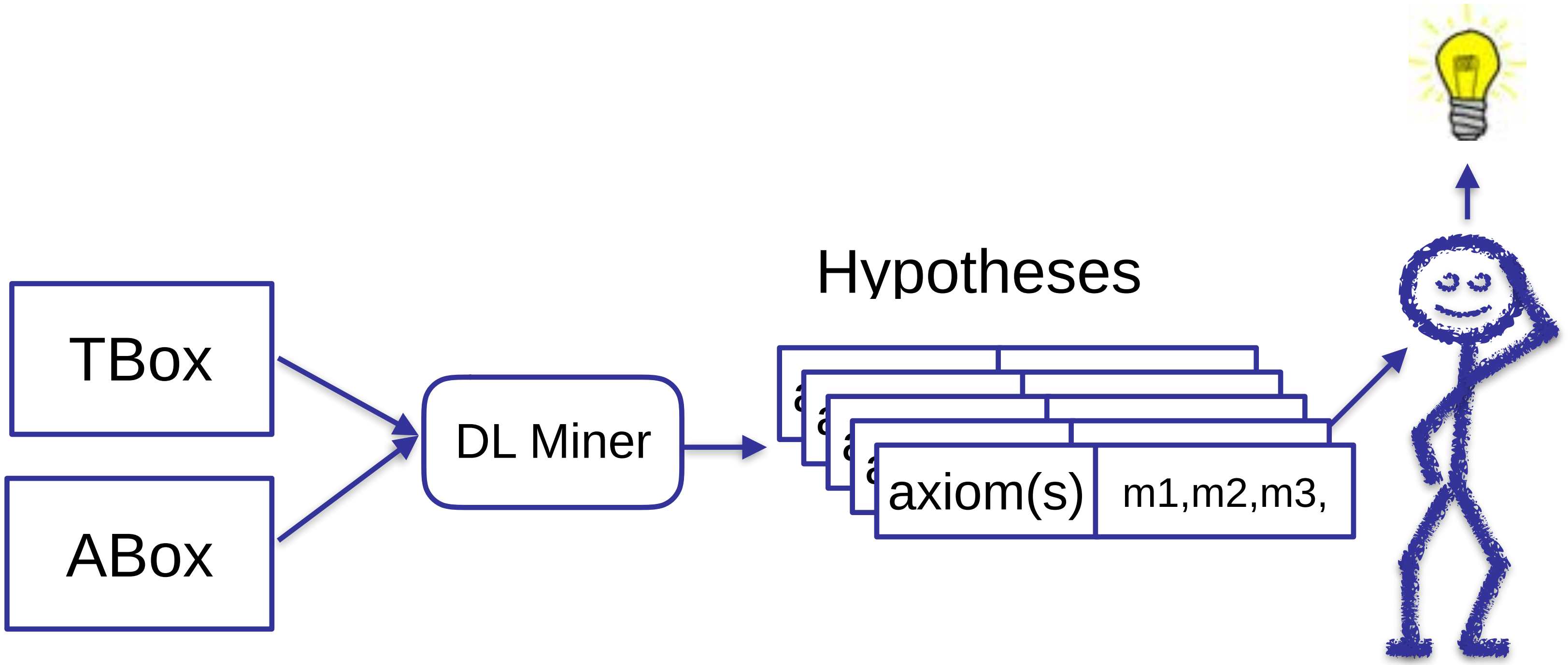


Design, run, analyse experiments

How costly are the different measures?



But - what about the semantic mining?



So, what have we got? (new version)

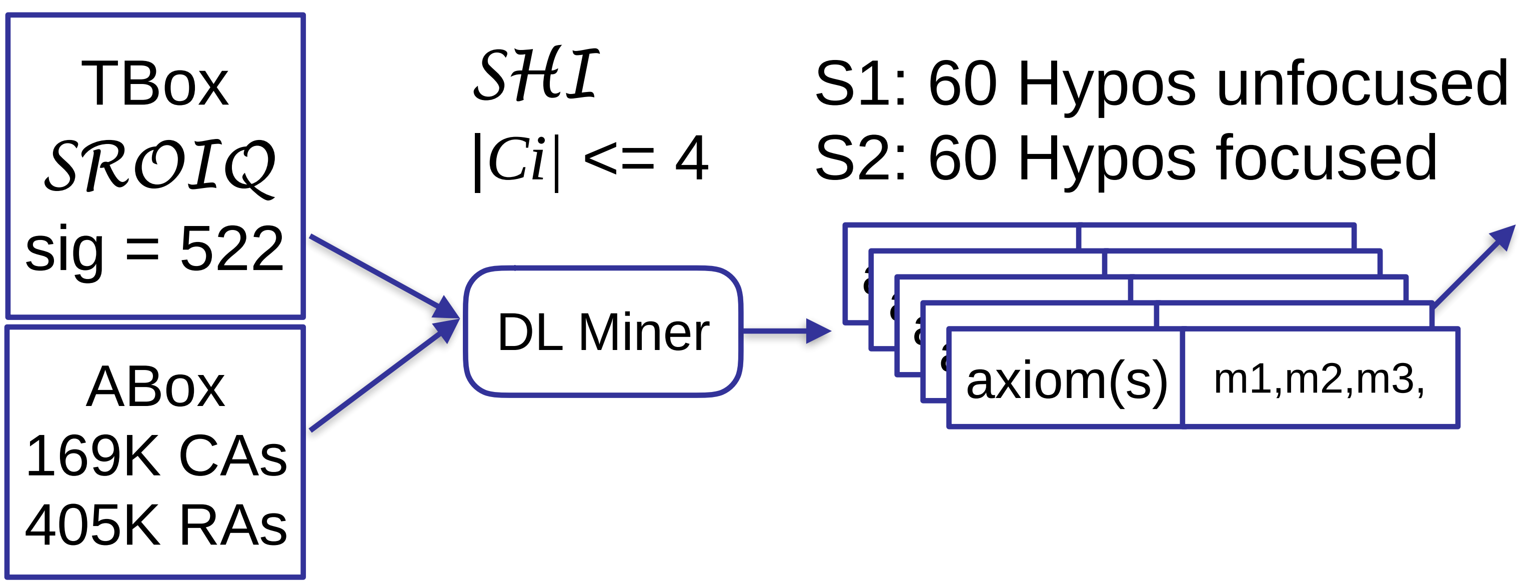
- ✓ Loads of measures to capture aspects of hypotheses
  - mostly independent
  - some superfluous on positive data (unsurprisingly)
- ✓ Hypothesis generation & evaluation is feasible
  - provided our ontology is classifiable
  - provided our search space isn't too massive
    - …focus!
- Which measures indicate *interesting* hypothesis?
- What is the shape for *interesting* hypothesis?
  - are longer/bigger hypotheses better?
- What do we do with them?
  - how do we guide users through these?

Design, run, analyse survey

Can we learn hypotheses are

- usefull/interesting?

…and how does this correlate with measures?



Design, run, analyse survey

Can we learn hypotheses are

- usefull/interesting?

…and how does this correlate with measures?

30 high-confidence
30 low-confidence

TBox
$\mathcal{SROIQ}$
sig = 522

ABox
169K CAs
405K RAs

$\mathcal{SHI}$
$|Ci|$ <= 4

DL Miner

S1: 60 Hypos unfocused
S2: 60 Hypos focused

axiom(s)	m1,m2,m3,

Design, run, analyse survey

Can we learn hypotheses are

- usefull/interesting?

…and how does this correlate with measures?

30 high-confidence
30 low-confidence

Valid?
Interesting?

TBox
$\mathcal{SROIQ}$
sig = 522

ABox
169K CAs
405K RAs

$\mathcal{SHI}$
$|Ci|$ <= 4

DL Miner

S1: 60 Hypos unfocused
S2: 60 Hypos focused

axiom(s)	m1,m2,m3,

Design, run, analyse survey

How good/valid are the mined hypotheses?

	Validity	Interestingness				
		0	*1*	*2*	*3*	*4*
Survey 1 (unfocused)	*Wrong*	6	11	30	-	-
	Don't know	-	1	-	2	4
	Correct	-	-	-	6	-

Design, run, analyse survey

How good/valid are the mined hypotheses?

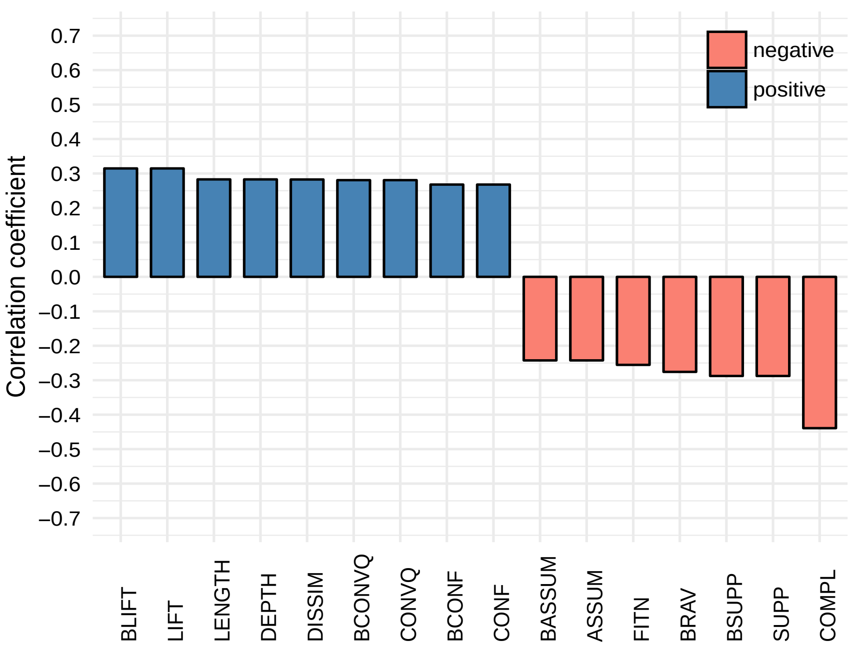
	Validity	Interestingness				
		0	*1*	*2*	*3*	*4*
Survey 1 (unfocused)	*Wrong*	6	11	30	-	-
	Don't know	-	1	-	2	4
	Correct	-	-	-	6	-
Survey 2 (focused)	*Wrong*	1	-	1	-	5
	Don't know	-	-	-	-	49
	Correct	-	-	-	-	4

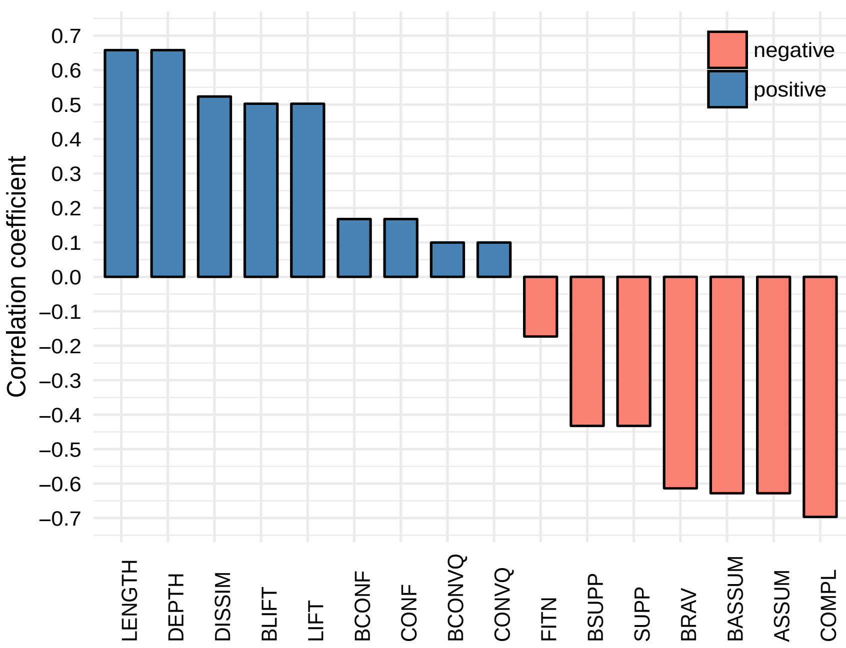
Design, run, analyse survey

How does validity/interestingness correlate with our metrics?

Design, run, analyse survey

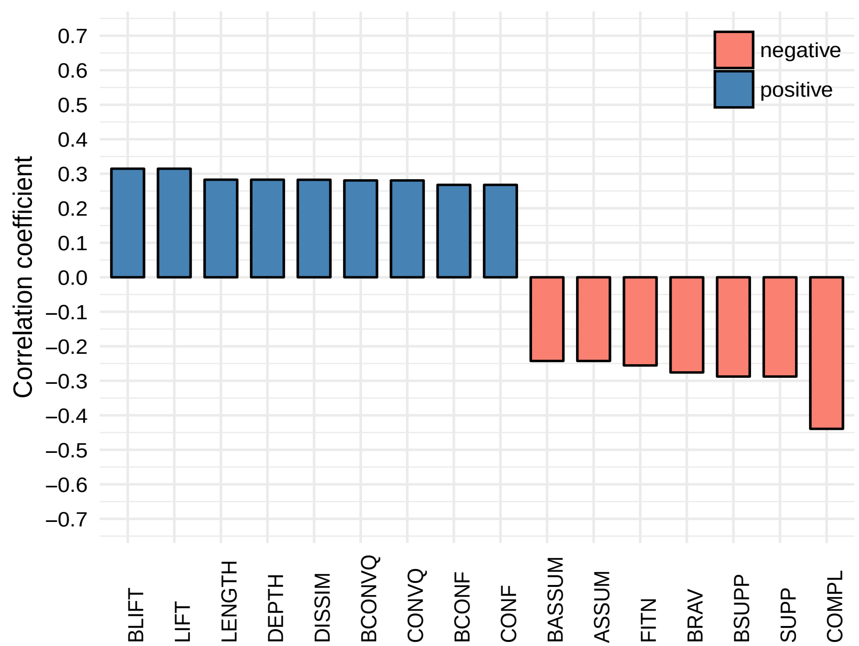
How does validity/interestingness correlate with our metrics?

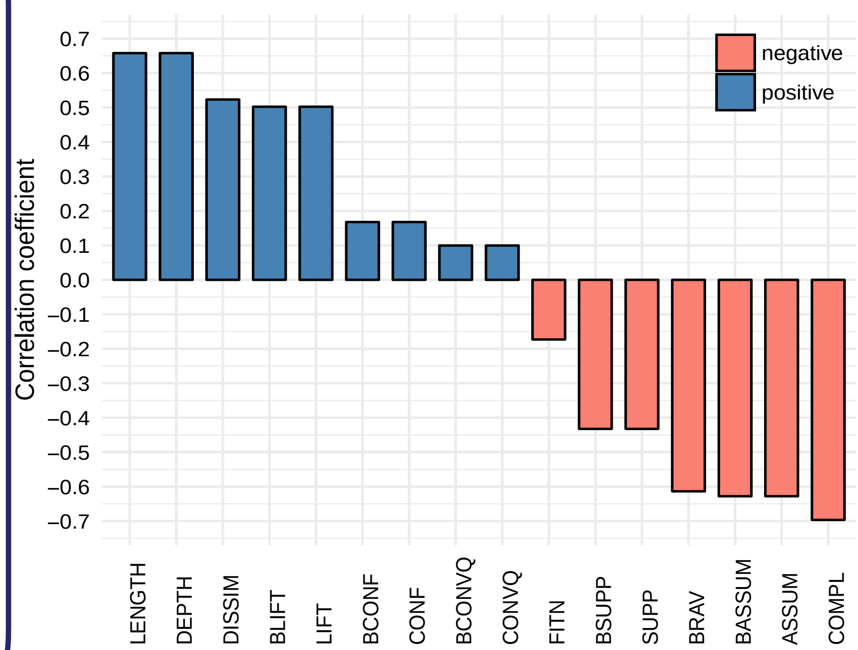




Design, run, analyse survey

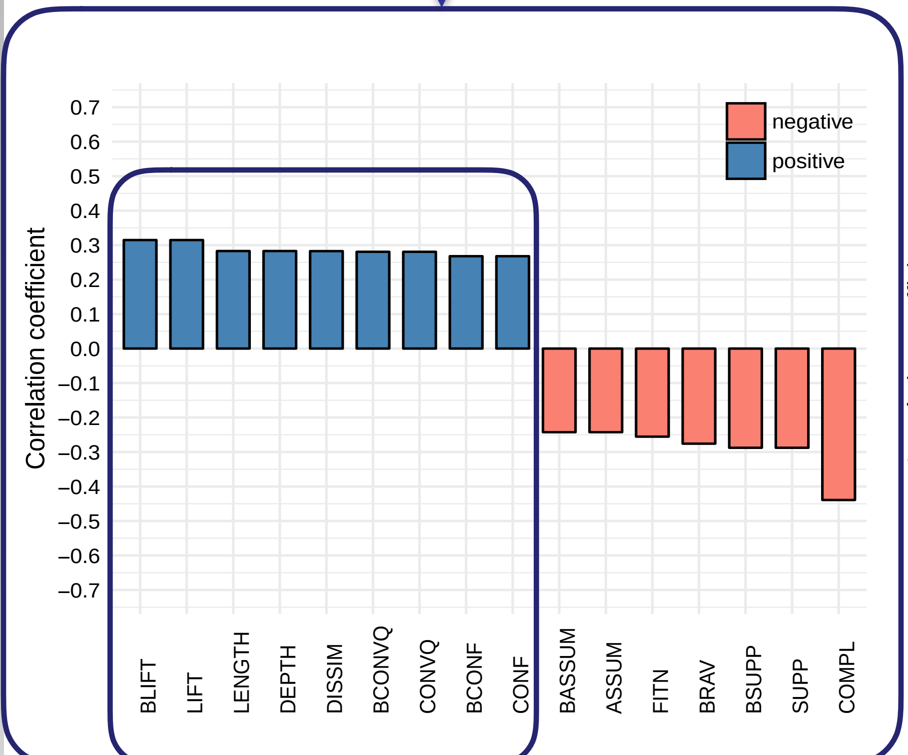
How does validity/interestingness correlate with our metrics?

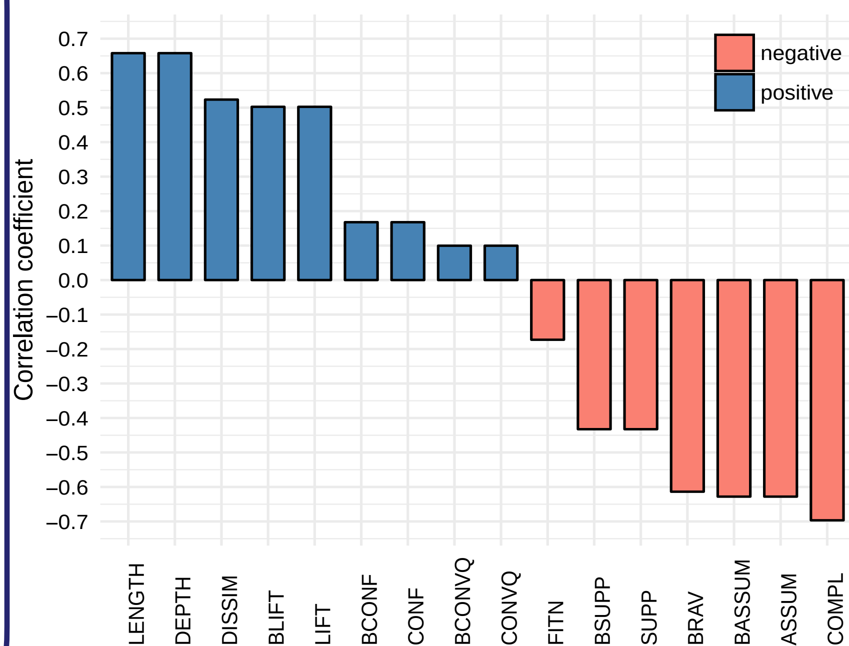




Design, run, analyse survey

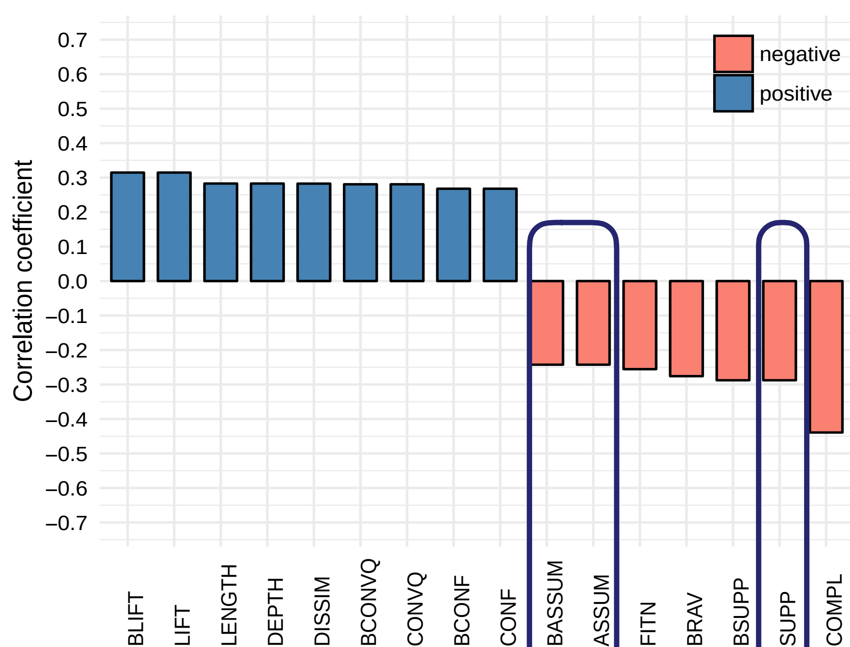
How does validity/interestingness correlate with our metrics?

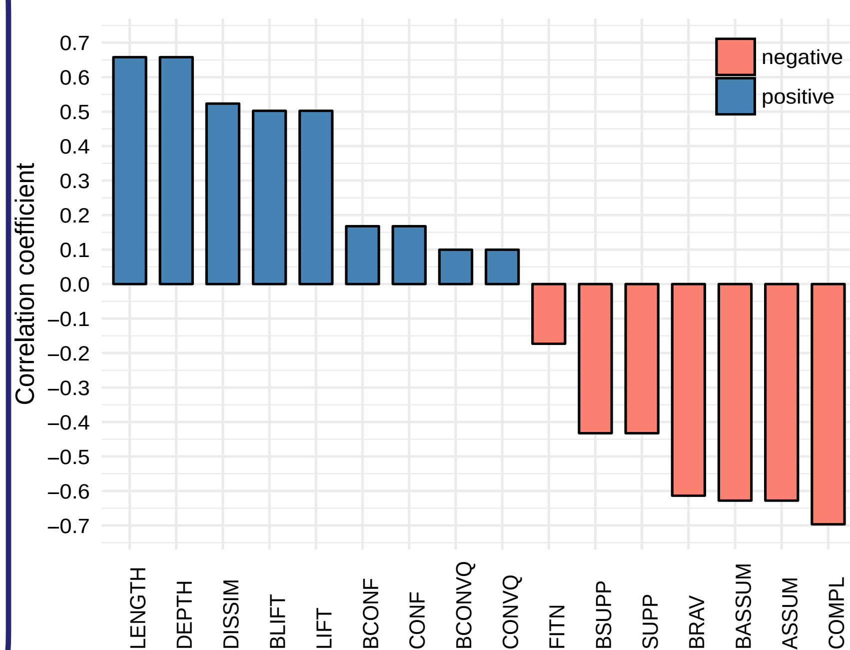




Design, run, analyse survey

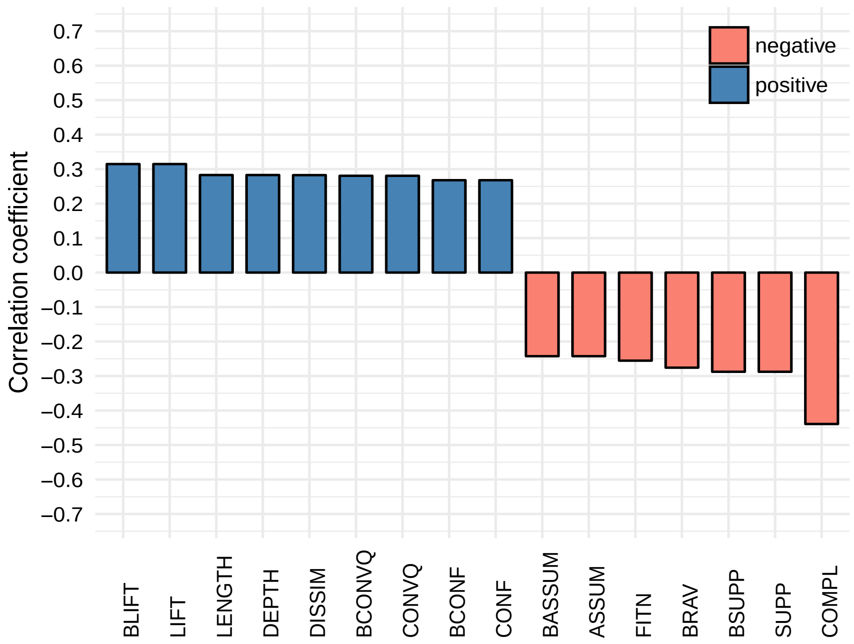
How does validity/interestingness correlate with our metrics?

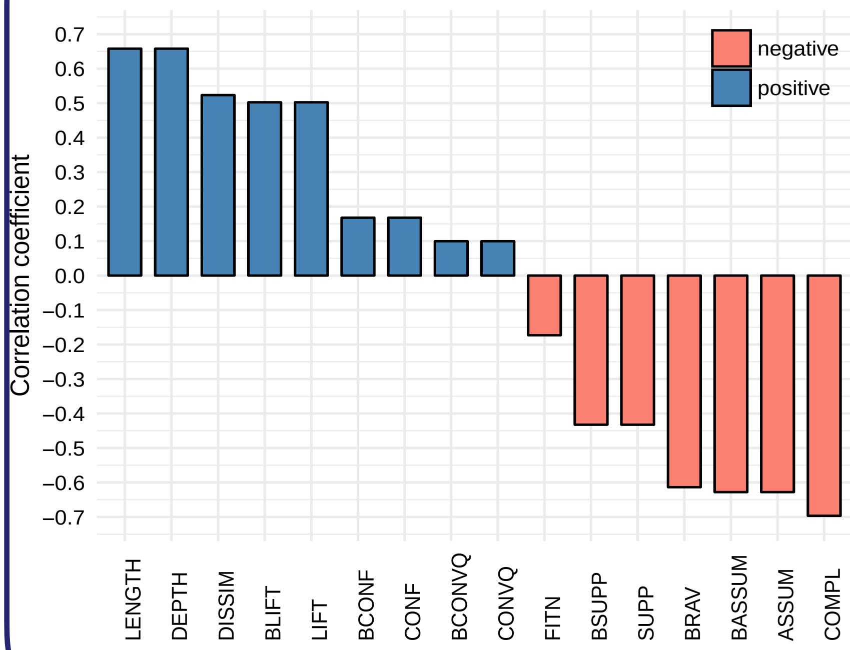




Design, run, analyse survey

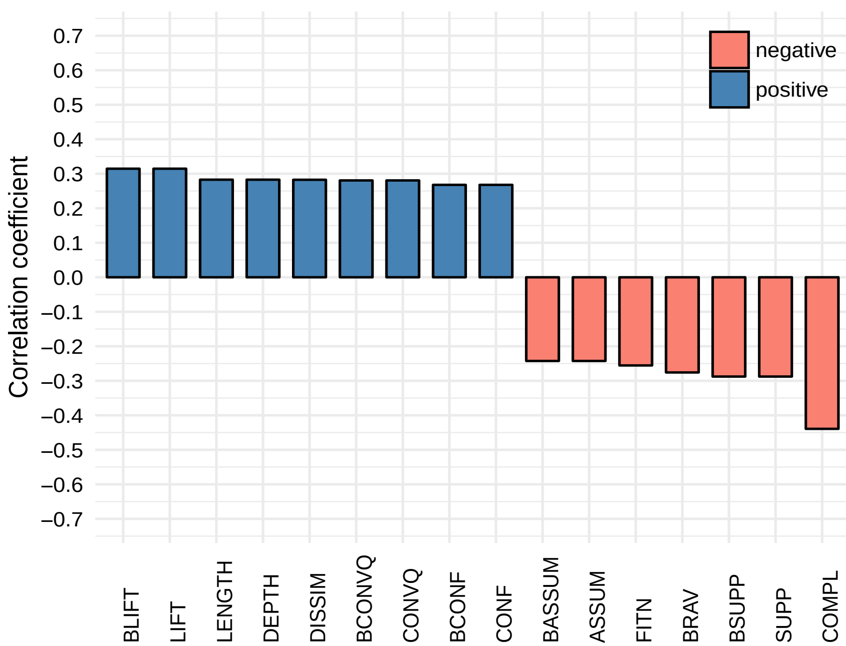
How does validity/interestingness correlate with our metrics?

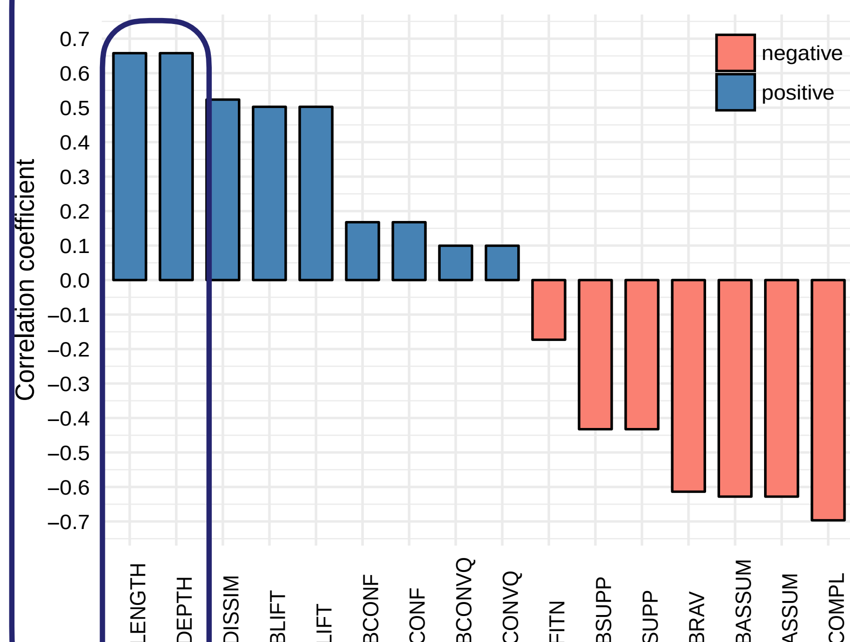




Design, run, analyse survey

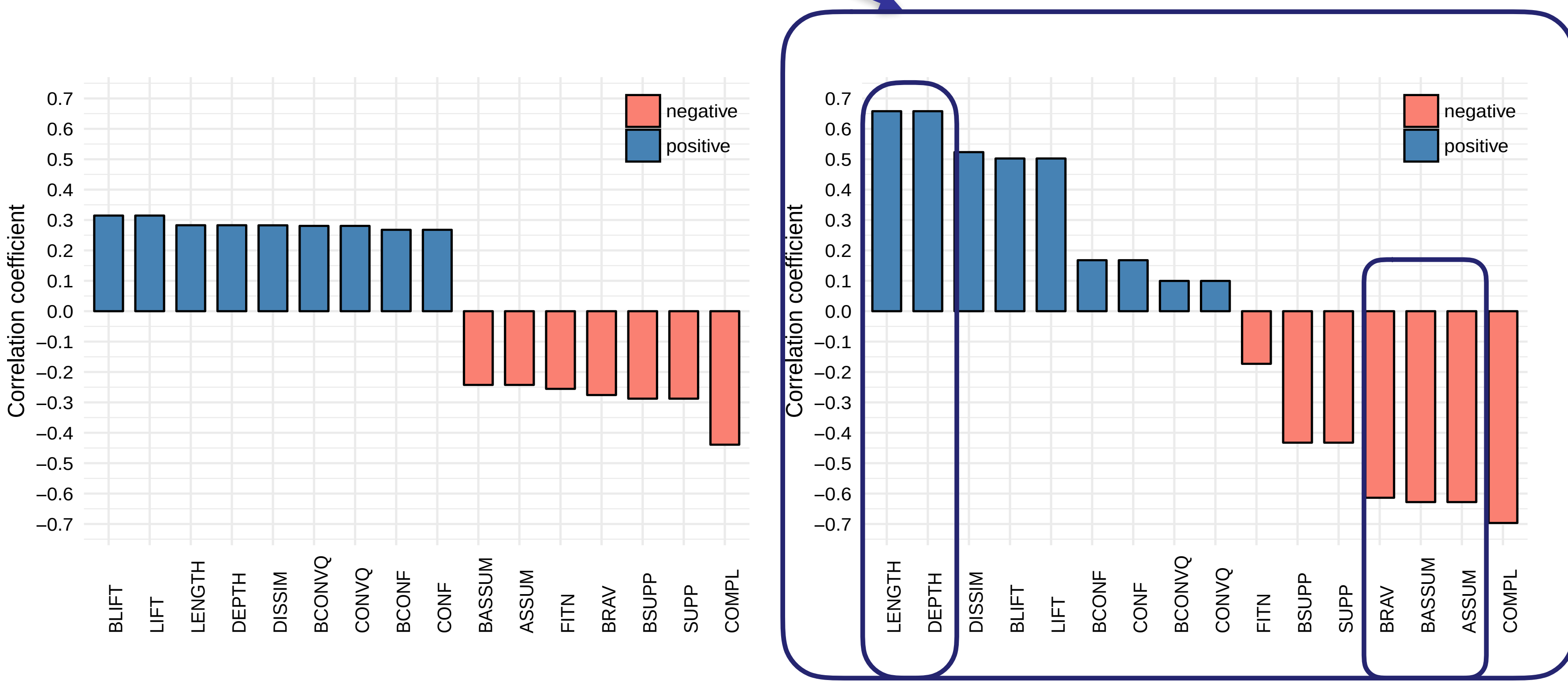
How does validity/interestingness correlate with our metrics?





Design, run, analyse survey

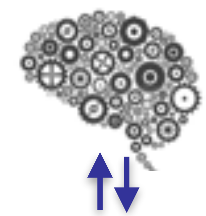
How does validity/interestingness correlate with our metrics?

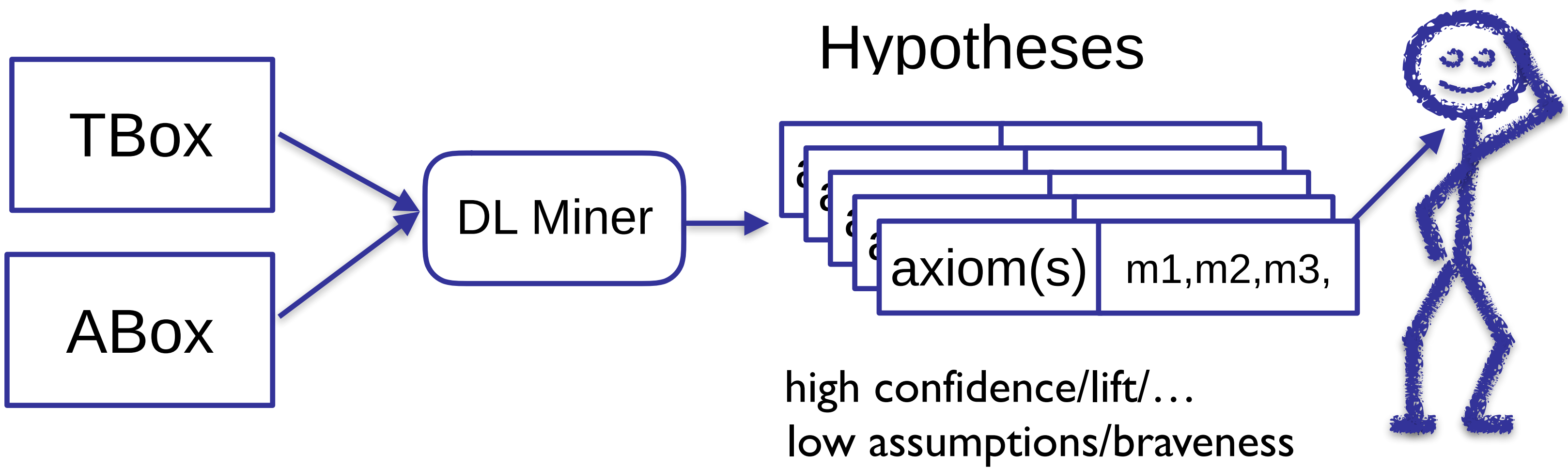


What we learned: 3 kinds of hypotheses!

1. An interesting hypothesis can give

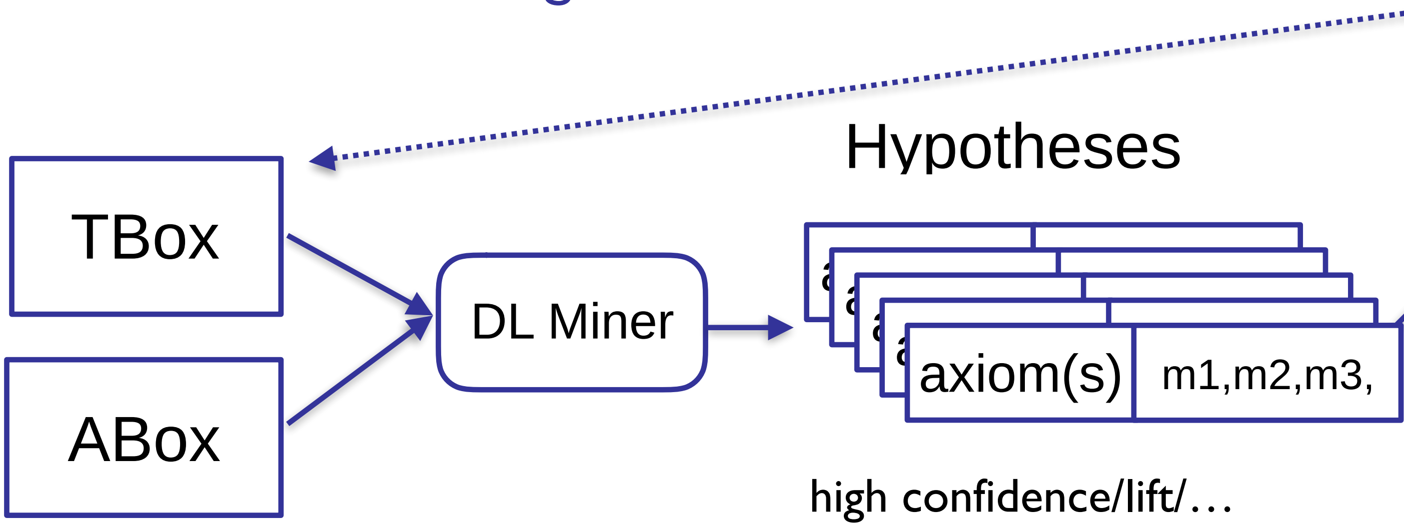
   new insights into domain

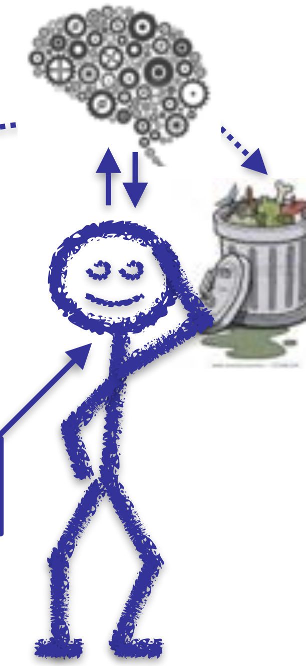




What we learned: 3 kinds of hypotheses!

1. An interesting hypothesis can give
   new insights into domain



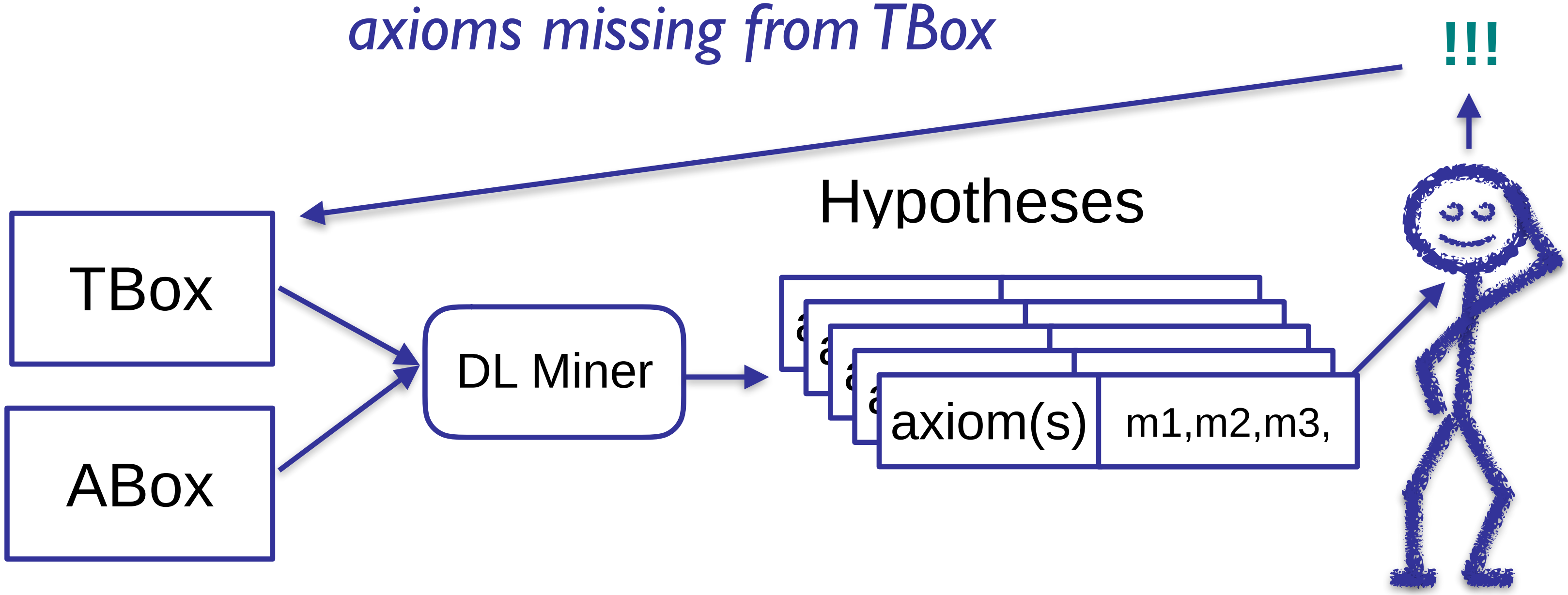


high confidence/lift/…
low assumptions/braveness

Semantic Mining

What we learned: 3 kinds of hypotheses!

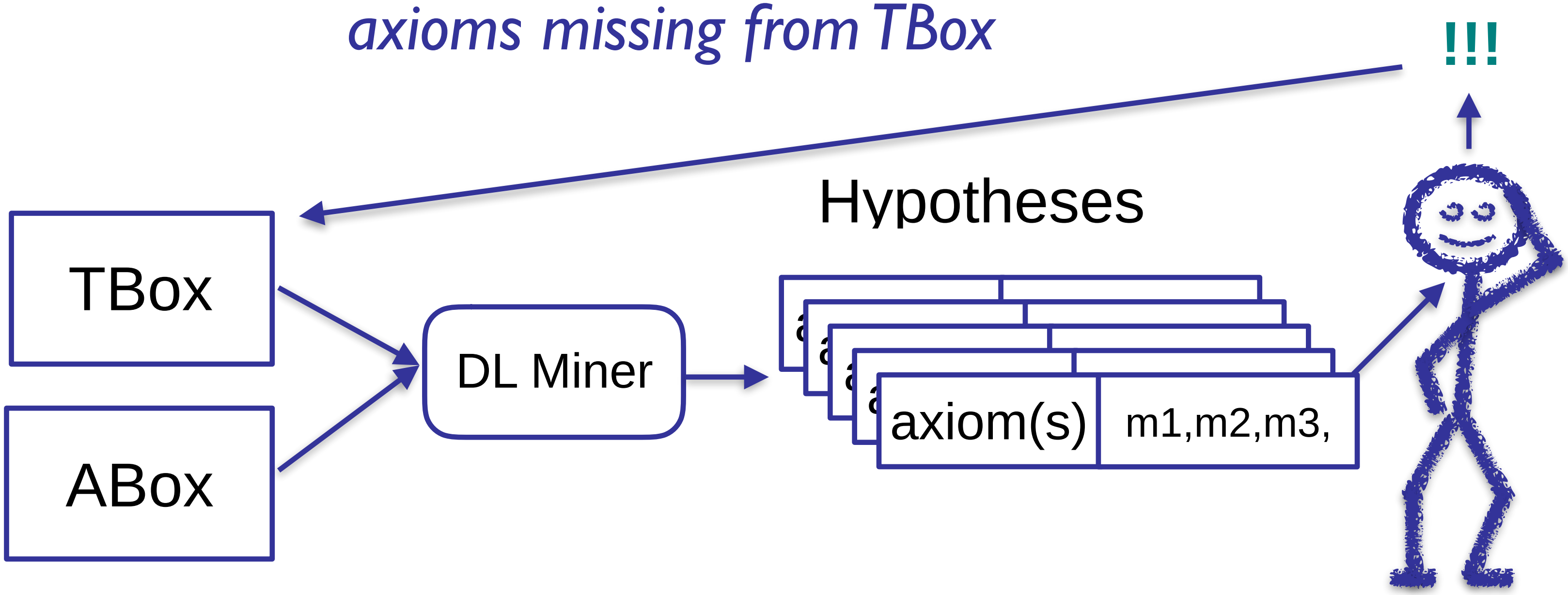
2. An interesting hypothesis can reveal
axioms missing from TBox



What we learned: 3 kinds of hypotheses!

2. An interesting hypothesis can reveal

axioms missing from TBox

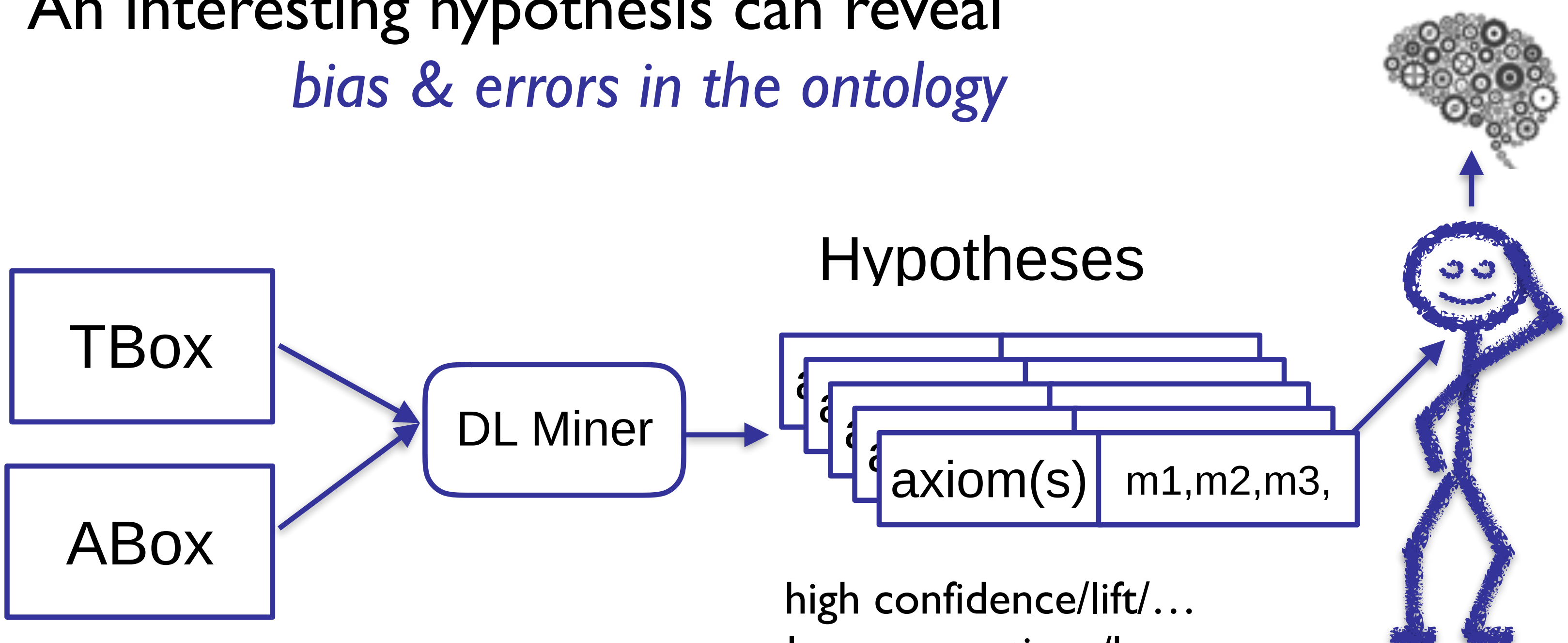


TBox completion

ontology learning from data

What we learned: 3 kinds of hypotheses!

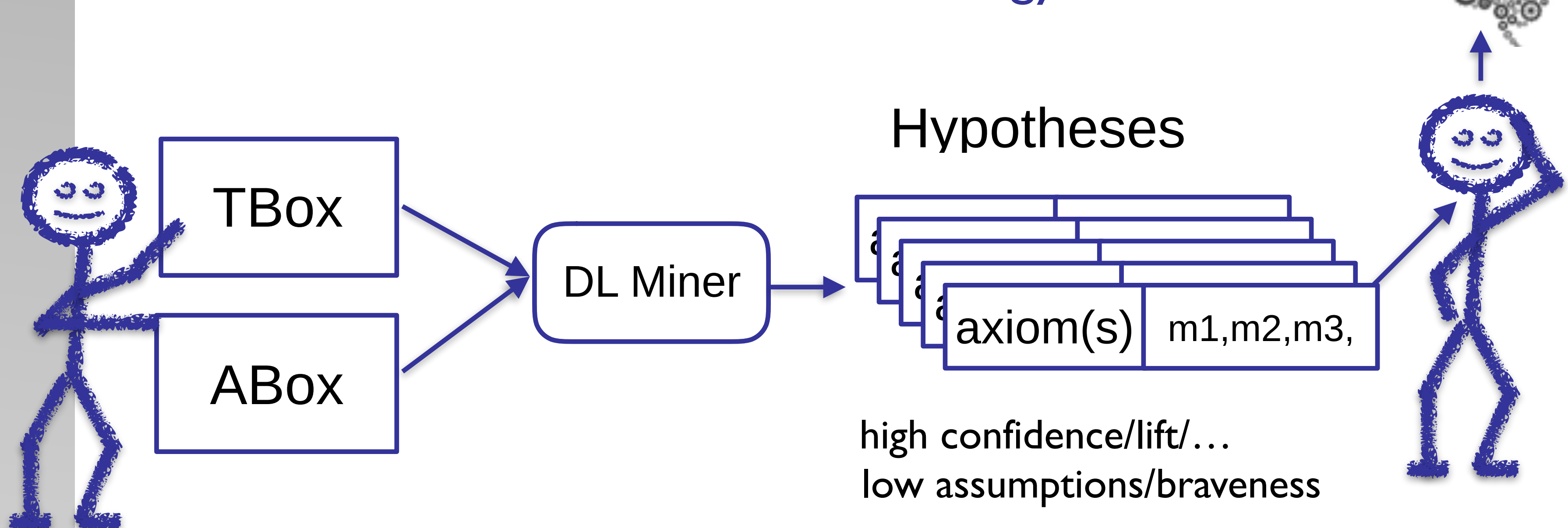
3. An interesting hypothesis can reveal *bias & errors in the ontology*



high confidence/lift/…
low assumptions/braveness

What we learned: 3 kinds of hypotheses!

3. An interesting hypothesis can reveal
bias & errors in the ontology

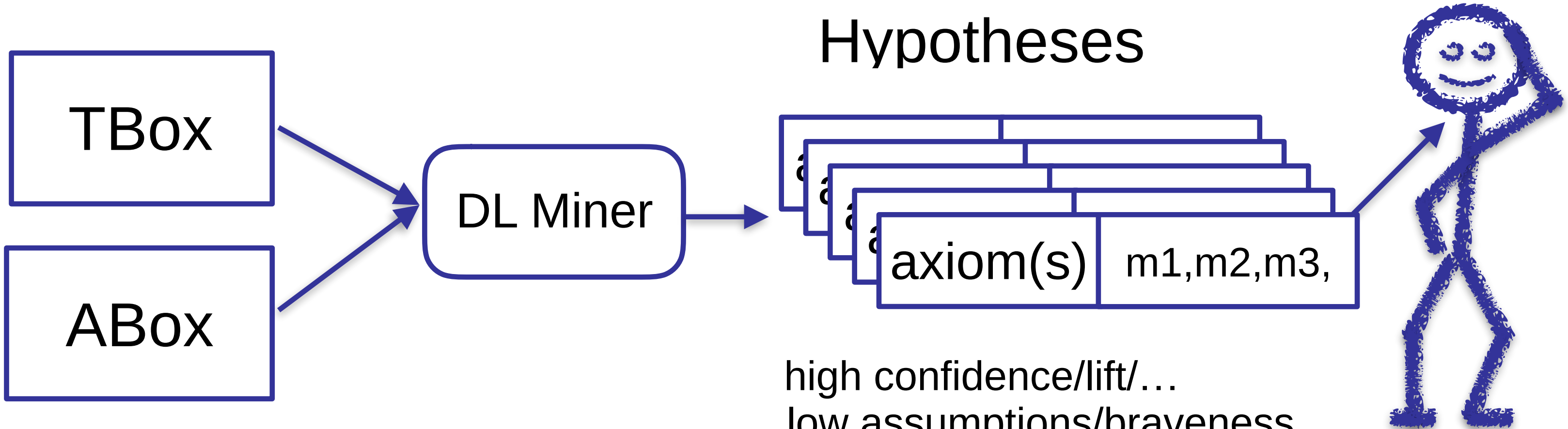


high confidence/lift/…
low assumptions/braveness

Semantic Data Analysis

3 kinds of hypotheses - can we predict?

No - they look alike

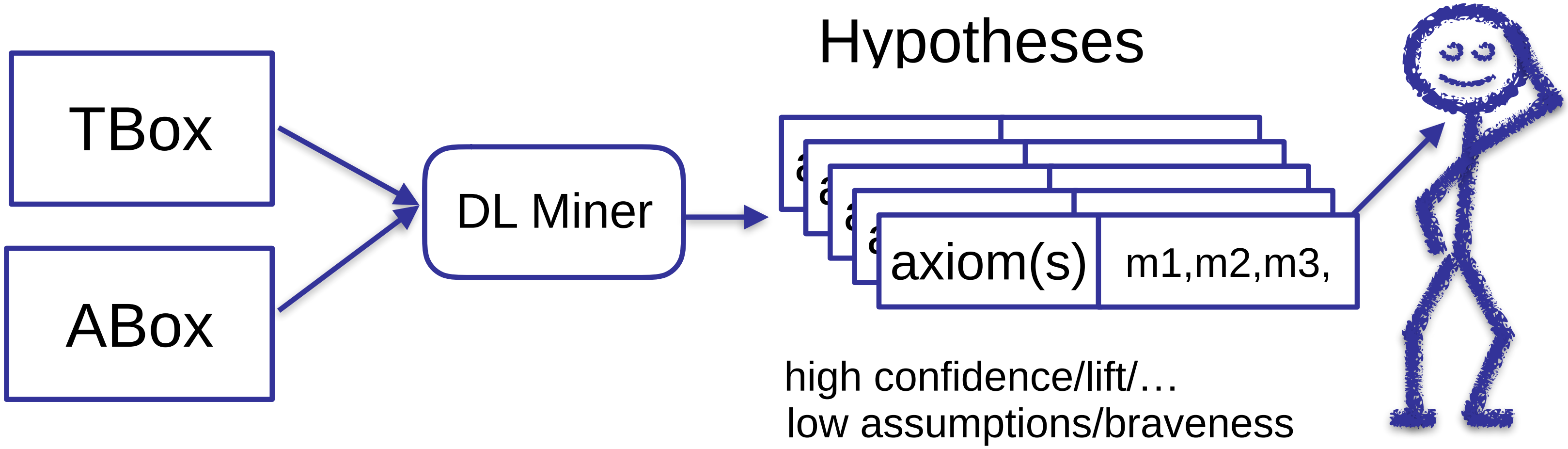


high confidence/lift/…
low assumptions/braveness

3 kinds of hypotheses - can we predict?

No - they look alike

Perhaps - with different ABoxes/other sources



Summary & Outlook

- Mining rich axioms from ontologies is possible
  - gives us more than we thought
  - expressive axioms are better!
- Fine test case for incremental/ABox reasoning
- More surveys
  - to better understand relevance of metrics
  - but we've got the shape now
- Redundancy in general is tricky & costly
  - stripping superfluous parts from concepts, (sets of) axioms
- We need even better refinement operators:
  - for more expressive DLs
  - redundancy-free
  - ontology-aware

Subjective ontology-based problems

- are great fun
  - design of experiments & surveys
  - but also rather complex: *sooo* many design choices
- specifying & implementing good parameters is tricky
  - metrics make "ontology mining" subjective
  - requires understanding of logic & reasoners & …
- are plentiful/numerous
  - abduction
  - similarity
  - good explanations/proofs for entailments justifications
  - good counter-models for non-entailments
  - good repair of inconsistent/incoherent ontologies
  - …

Special Thanks
to
Slava Sazonau

be described and so that computers can reason about this knowledge. DLs have recently gained increased importance since they form the logical basis of widely used ontology languages, in particular the web ontology language OWL.

Written by four renowned experts, this is the first textbook on Description Logic. It is suitable for self-study by graduates and as the basis for a university course. Starting from a basic DL, the book introduces the reader to their syntax, semantics, reasoning problems and model theory, and discusses the computational complexity of these reasoning problems and algorithms to solve them. It then explores a variety of reasoning techniques, knowledge-based applications and tools, and describes the relationship between DLs and OWL.

Department of Computer Science at the University of Oxford.

Carsten Lutz is a professor in the Department of Computer Science at the University of Bremen.

Uli Sattler is a professor in the Information Management Group within the School of Computer Science at the University of Manchester.

Cover illustration: The Description Logic logo. Courtesy of Enrico Franconi.

Designed by Zoe Naylor.

ocks,

An Introduction to Description Logic

CAMBRIDGE
UNIVERSITY PRESS
www.cambridge.org

ISBN 978-0-521-87361-1

9 780521 873611

CAMBRIDGE

An Introduction to Description Logic

Franz Baader
Ian Horrocks
Carsten Lutz
Uli Sattler