Friday

Lemmas
Model based explanations

Schedule for today

- Motivation for lemmas
- Towards a complexity model
- Model interaction and computation

Motivation for lemmas

Remember: proofs are hard

- Full proofs have mind numbing detail
- You need to know the proof theory
  - Which proof theory?
    - Proofs may be difficult to compute
- Way more proofs than justifications
- Proofs are not repair oriented
- Are proofs always (or ever) needed?
- Excellent tools needed!

MANCHESTER 1824

The University of Manchester

What do we want?

- Explanation should yield/provoke understanding (What sort?)
- Some tasks
    1. Debugging an unsatisfiable class
    2. Explaining an entailment to someone
    3. Verifying a possible entailment
        - (Debugging the reasoner)
- Other goals: formalism mastery, understanding reasoners

What else?

- Don't make things worse
  - Ever (if possible); definitely not often
- Integrate with existing
  - Tools
  - Practice
- Good cost/benefit
- Beware confounding factors and wishful thinking

Operationalization

- Measure understanding via task performance
  - Objective(ish) metrics
  - Clear role for subjective factors
  - Isolates benefits
- Doesn't capture overall effects
  - Or long terms ones

Problems with Justs?

Three kinds

1. Finding them (solved by services)
2. Their number (presentation issue?)
3. Understanding them individually

   Not always necessary to successful repair
   ~> domain knowledge!

Open Question

➡ Are there justifications that are too hard?

Example

$$\mathcal{J} = \{$$

$$\begin{aligned}
\text{Person} &\sqsubseteq \neg\text{Movie} \\
\text{RRated} &\sqsubseteq \text{CatMovie} \\
\text{CatMovie} &\sqsubseteq \text{Movie} \\
\text{RRated} &\equiv \exists\, \text{hasScript}.\text{ThrillerScript} \\
&\quad \sqcup\ \forall\, \text{hasViolenceLevel}.\text{High} \\
\exists\, \text{hasViolenceLevel} &\sqsubseteq \text{Movie}\ \}
\end{aligned}$$

$$\models \text{Person} \sqsubseteq \bot$$

RRated: *Movie viewers under 17 (18) to be accompanied by adult*
CatMovie: *Categorised movie*

Example

$\mathcal{J} = \{$ InverseProperties(hasPet, isPetOf)

isPetOf(Rex, Mick)

Domain(hasPet, Person)

Male(Mick)

reads(Mick, DailyMirror)

drives(Mick, Q123ABC)

Van(Q123ABC)

$\mathsf{Van} \sqsubseteq \mathsf{Vehicle}$

WhiteThing(Q123ABC)

$\mathsf{Driver} \equiv \mathsf{Person} \sqcap \exists \mathsf{drives}.\mathsf{Vehicle}$

$\mathsf{Driver} \sqsubseteq \mathsf{Adult}$

$\mathsf{Man} \equiv \mathsf{Adult} \sqcap \mathsf{Male} \sqcap \mathsf{Person}$

$\mathsf{WhiteVanMan} \equiv \mathsf{Man} \sqcap \exists \mathsf{drives}.(\mathsf{Van} \sqcap \mathsf{WhiteThing})$

$\mathsf{WhiteVanMan} \sqsubseteq \forall \mathsf{reads}.\mathsf{Tabloid}$

$\mathsf{Tabloid} \sqsubseteq \mathsf{Newspaper} \}$ $\models \mathsf{Tabloid(DailyMirror)}$

Justification Hardness

- Not obvious for key user base
  - Description logics are restricted
  - People use them in constrained ways
  - People "know their ontologies"
  - Justifications support experimentation
- Even if a just is hard, a proof might not help
  - E.g., destroys relation to repair

A hard example?

$$\mathcal{J} = \{ \quad A \sqsubseteq B$$
$$C \sqsubseteq \exists R.E$$
$$B \sqsubseteq \exists R.D$$
$$\exists R.E \sqsubseteq A$$
$$\exists R.D \sqsubseteq F \ \} \qquad \models C \sqsubseteq F$$

A hard example?

$$\mathcal{J} = \{ \quad C \sqsubseteq \exists R.E$$
$$\exists R.E \sqsubseteq A$$
$$A \sqsubseteq B$$
$$B \sqsubseteq \exists R.D$$
$$\exists R.D \sqsubseteq F \ \} \qquad \models C \sqsubseteq F$$

A hard example?

$$\mathcal{J} = \{ \quad C \sqsubseteq \exists R.E$$
$$\exists R.E \sqsubseteq A$$
$$A \sqsubset B$$
$$B \sqsubseteq \exists R.D$$
$$\exists R.D \sqsubseteq F \ \} \qquad \models C \sqsubseteq F$$

Changing the order of the axioms helps.

MANCHESTER 1824

The University of Manchester

A hard example?

$$
\begin{aligned}
\mathcal{J} = \{ \quad \mathsf{Person} &\sqsubseteq \neg\mathsf{Movie} \\
\mathsf{RRated} &\sqsubseteq \mathsf{CatMovie} \\
\mathsf{CatMovie} &\sqsubseteq \mathsf{Movie} \\
\mathsf{RRated} &\equiv \exists\, \mathsf{hasScript}.\mathsf{ThrillerScript} \\
&\quad \sqcup\ \forall\, \mathsf{hasViolenceLevel}.\mathsf{High} \\
\exists\, \mathsf{hasViolenceLevel} &\sqsubseteq \mathsf{Movie} \quad \}
\end{aligned}
$$

$$\models \mathsf{Person} \sqsubseteq \bot$$

MANCHESTER 1824

The University of Manchester

A hard example?

$$\mathcal{J} = \{$$

$$\begin{aligned}
\mathsf{Person} &\sqsubseteq \neg\mathsf{Movie} \\
\mathsf{RRated} &\sqsubseteq \mathsf{CatMovie} \\
\mathsf{CatMovie} &\sqsubseteq \mathsf{Movie} \\
\mathsf{RRated} &\equiv \exists\, \mathsf{hasScript}.\mathsf{ThrillerScript} \\
&\quad \sqcup\ \forall\, \mathsf{hasViolenceLevel}.\mathsf{High} \\
\exists\, \mathsf{hasViolenceLevel} &\sqsubseteq \mathsf{Movie}\ \}
\end{aligned}$$

$$\models \mathsf{Person} \sqsubseteq \bot$$

Changing the order of the axioms doesn't seem to make this easier.

MANCHESTER 1824

The University of Manchester

How to Proceed?

Study users!

- Determine if justifications are hard
- Determine what makes them hard
- Distinguish inherent hardness from "newbie" hardness

Towards a complexity model

MANCHESTER 1824

The University of Manchester

Is the Problem Real?

User Study

Are some justifications difficult or impossible to understand?

How to people fare with justifications they can understand?

How do people read through justifications?

What makes justifications difficult to understand?

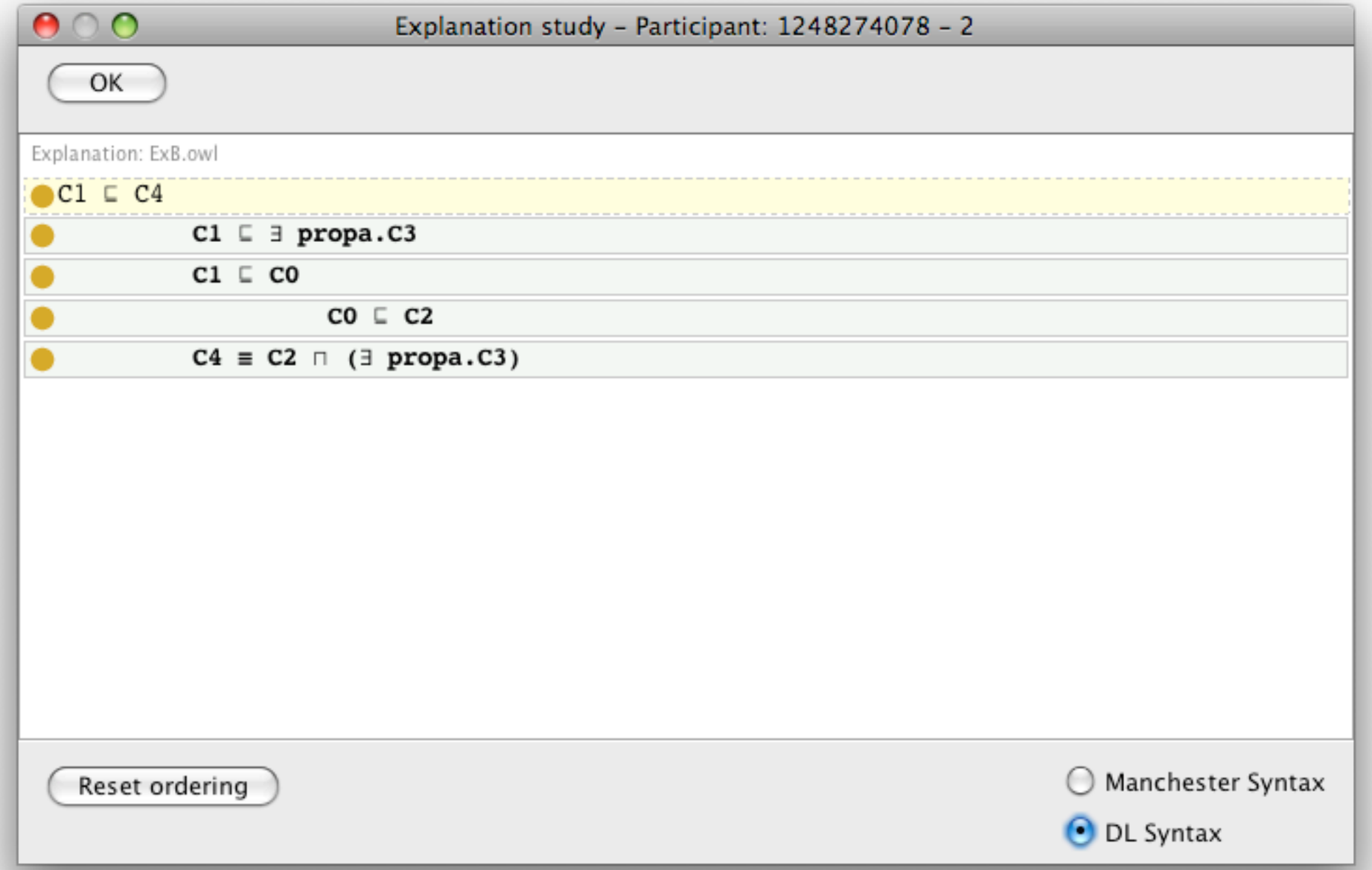
MANCHESTER 1824

The University of Manchester

User Study—Participants

- Number: 16
  - Staff+students from CS department
  - Divided into two rounds, 12+4
- Experience: ≤6 months ... ≥4 years
- Background: Users of ontology editors
  (e.g., Protégé)

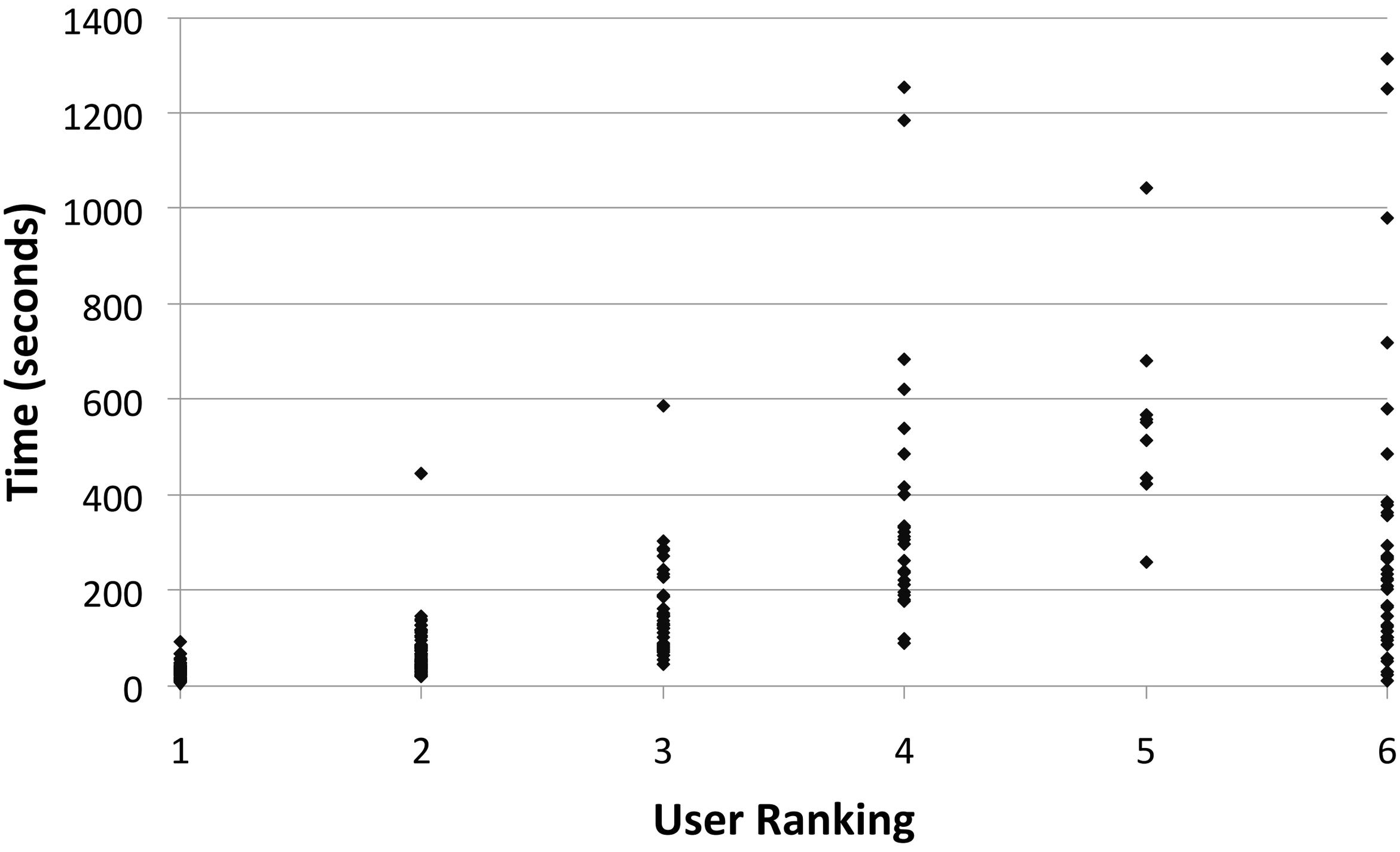
Procedure


Explanation study - Participant: 1248274078 - 2
OK
Explanation: ExB.owl
C1 ⊑ C4
C1 ⊑ ∃ propa.C3
C1 ⊑ C0
C0 ⊑ C2
C4 ≡ C2 ⊓ (∃ propa.C3)
Reset ordering
Manchester Syntax
DL Syntax

MANCHESTER 1824

ne University
Manchester

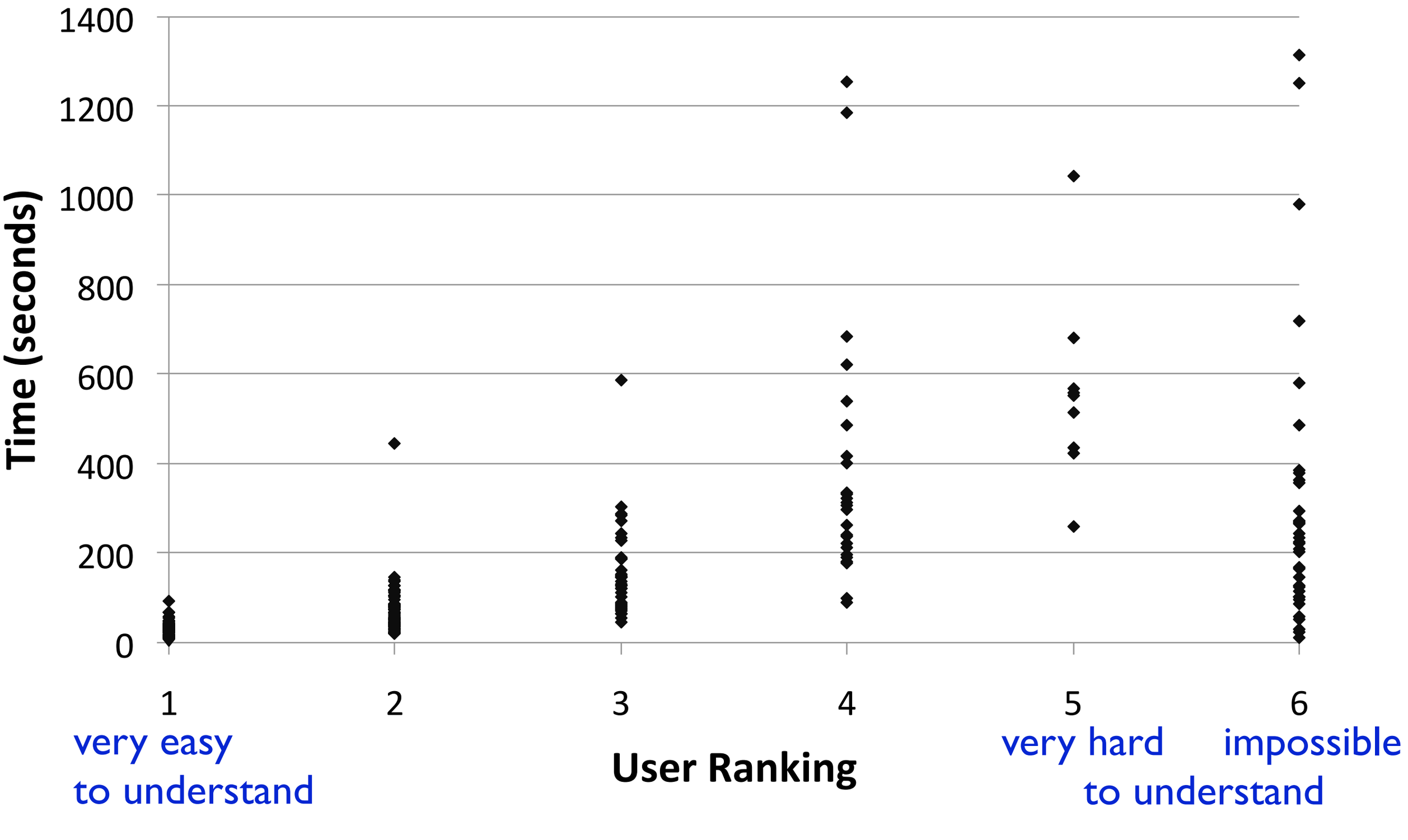
Participant Ranking versus Time



MANCHESTER 1824

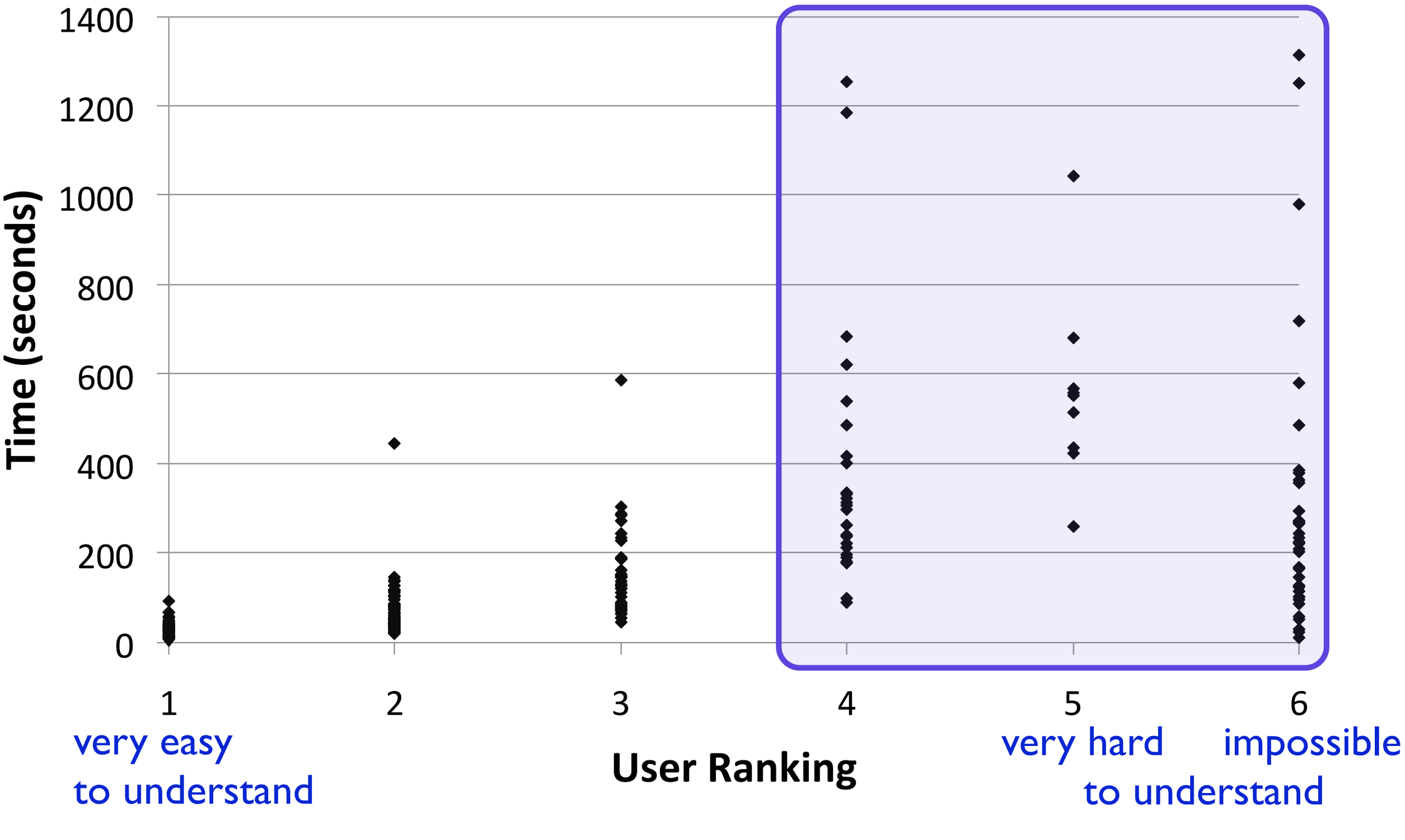
ne University
Manchester

Participant Ranking versus Time



MANCHESTER 1824

Participant Ranking versus Time



MANCHESTER 1824

The University of Manchester

Observations

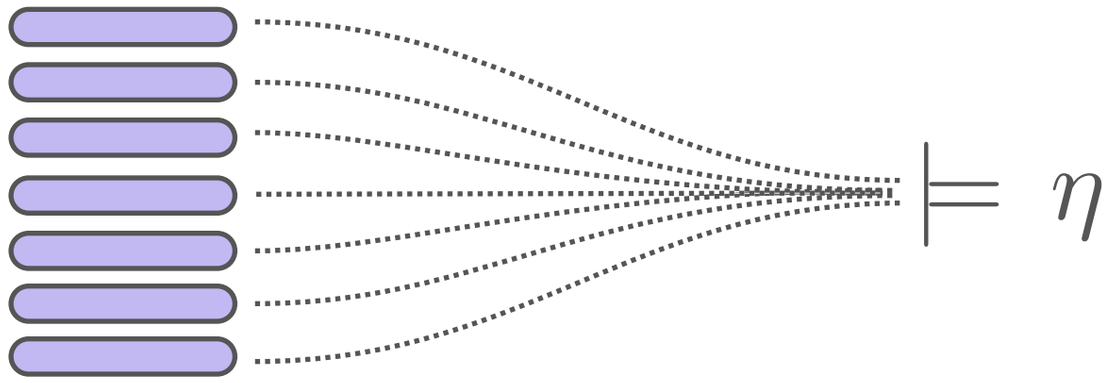
People arrive at intermediate conclusions as they read justifications

With difficult justifications people don't spot intermediate conclusions

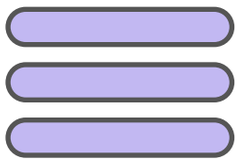
People can comfortably work with justifications:
everyone could understand simple justifications

~> Can we design a service that makes these steps explicit?

Overall Goal

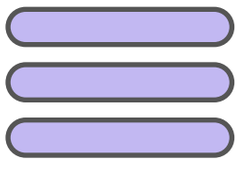


Overall Goal



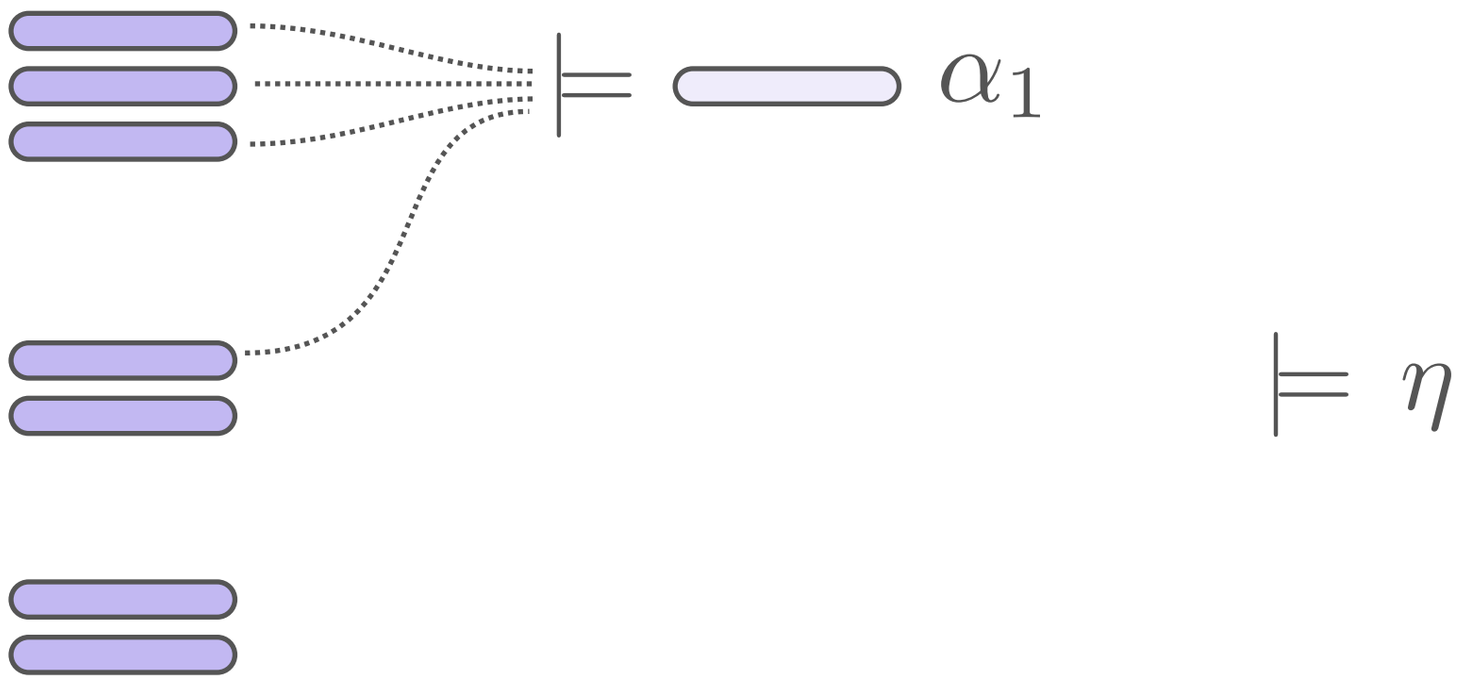
$\models \eta$

Overall Goal

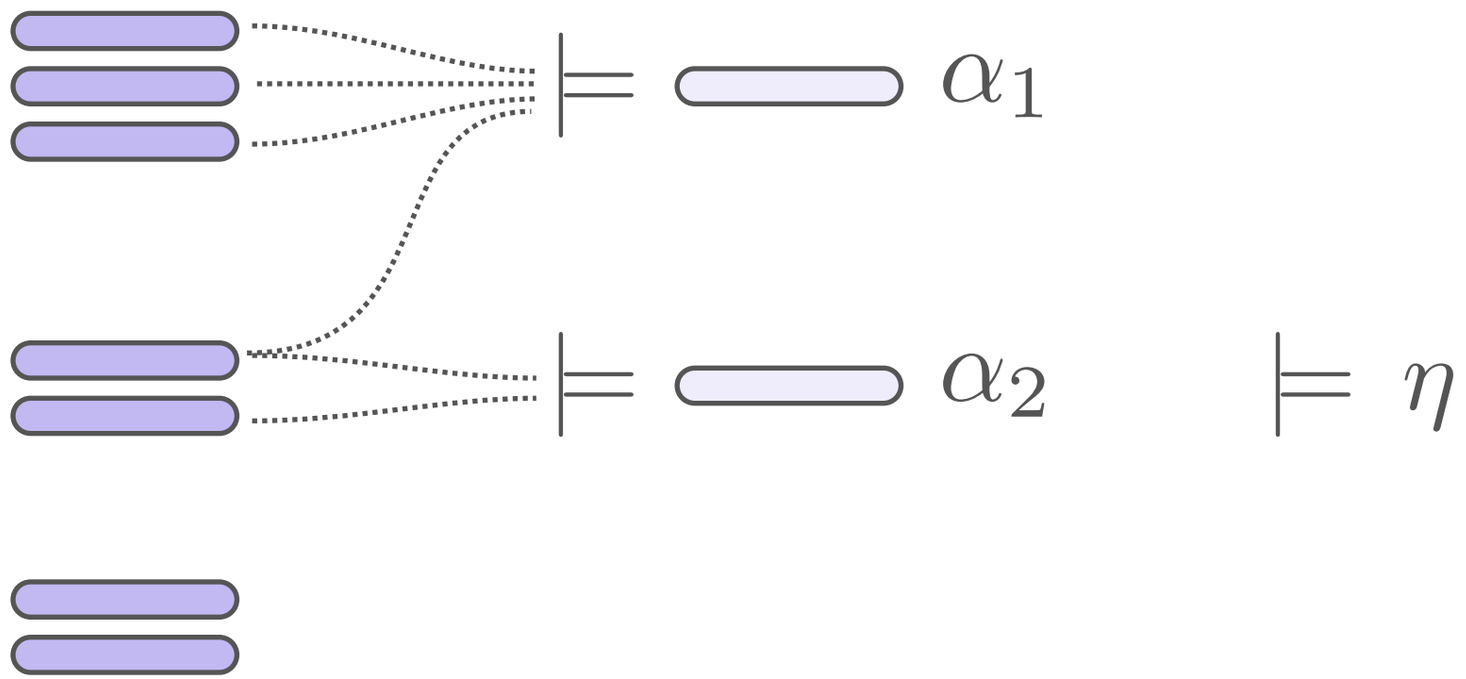


$\models \eta$

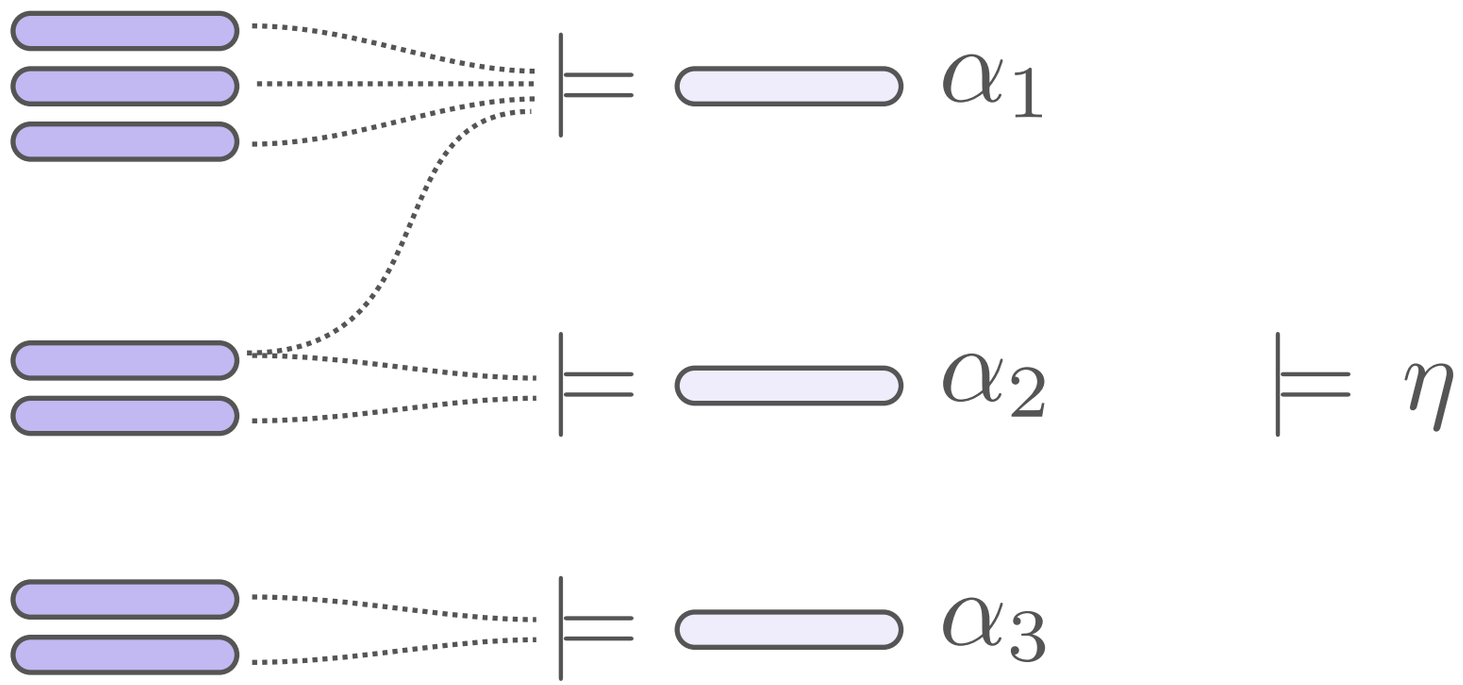
Overall Goal



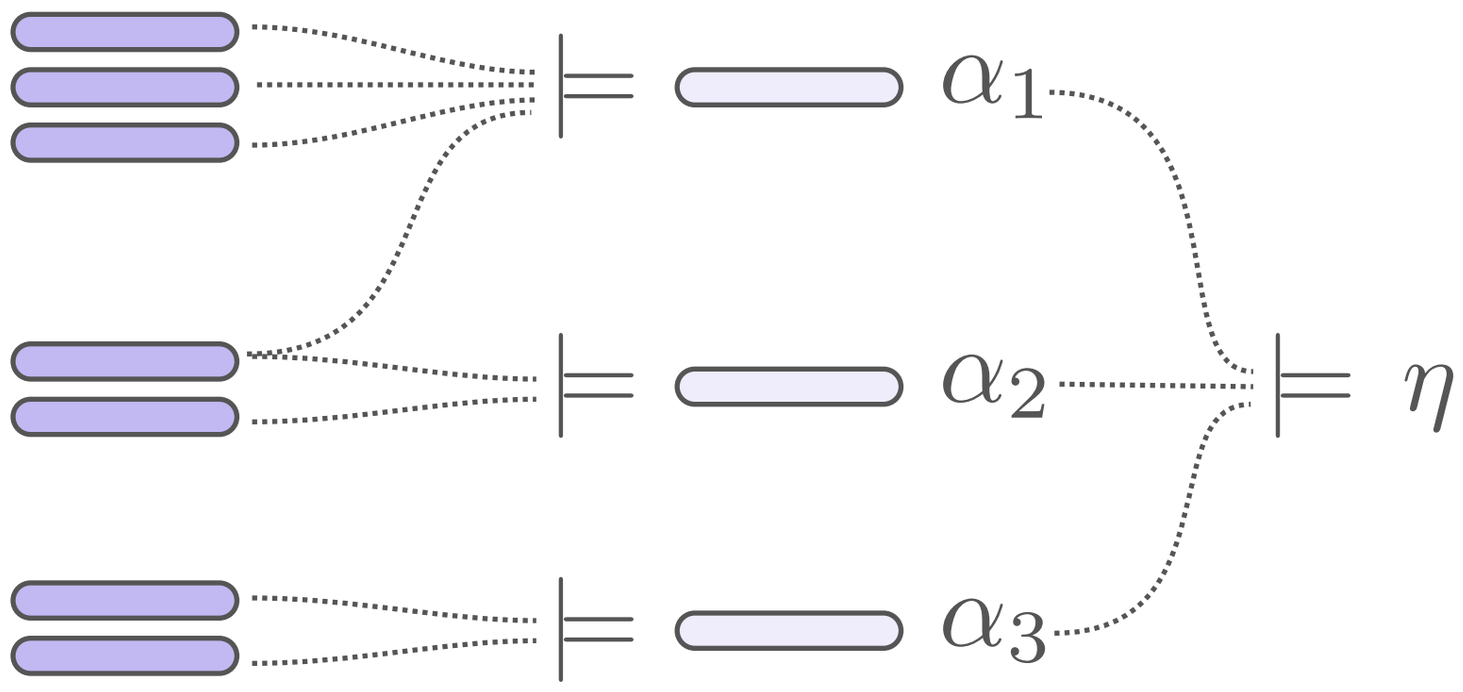
Overall Goal



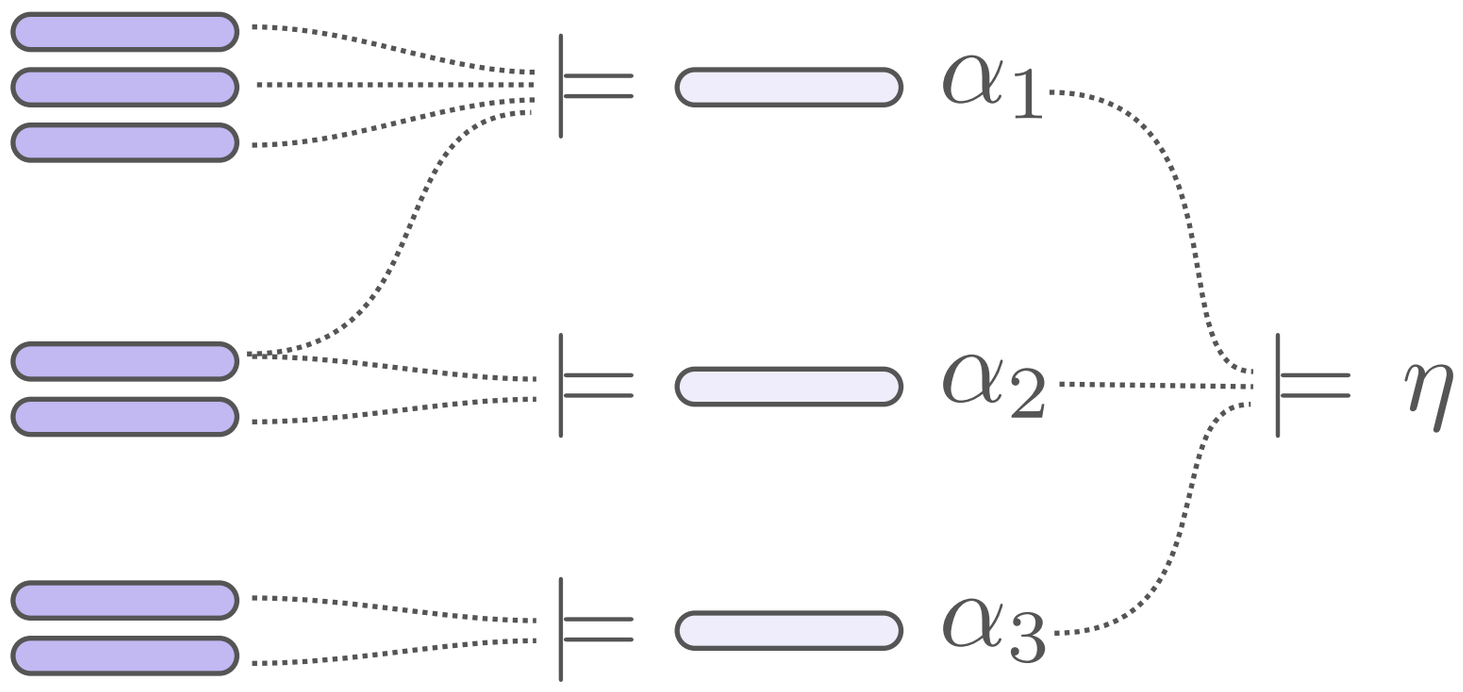
Overall Goal



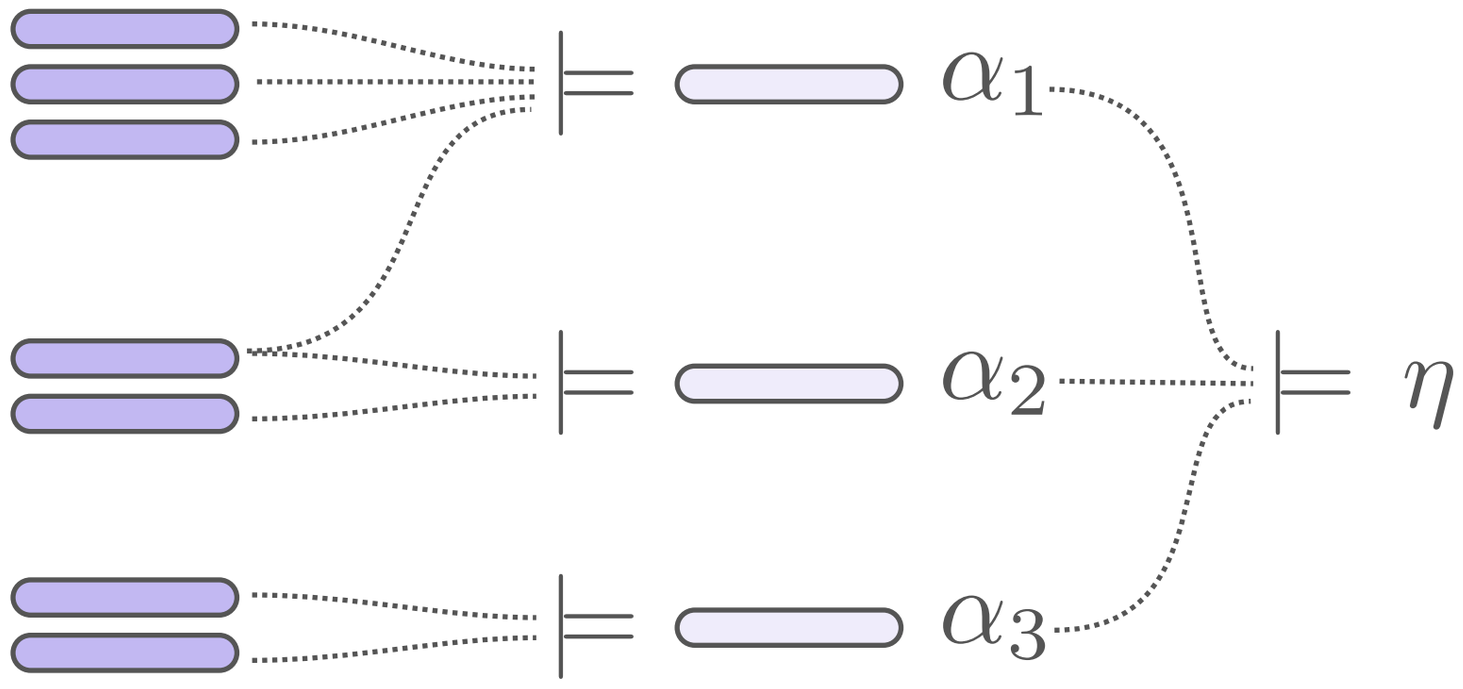
Overall Goal



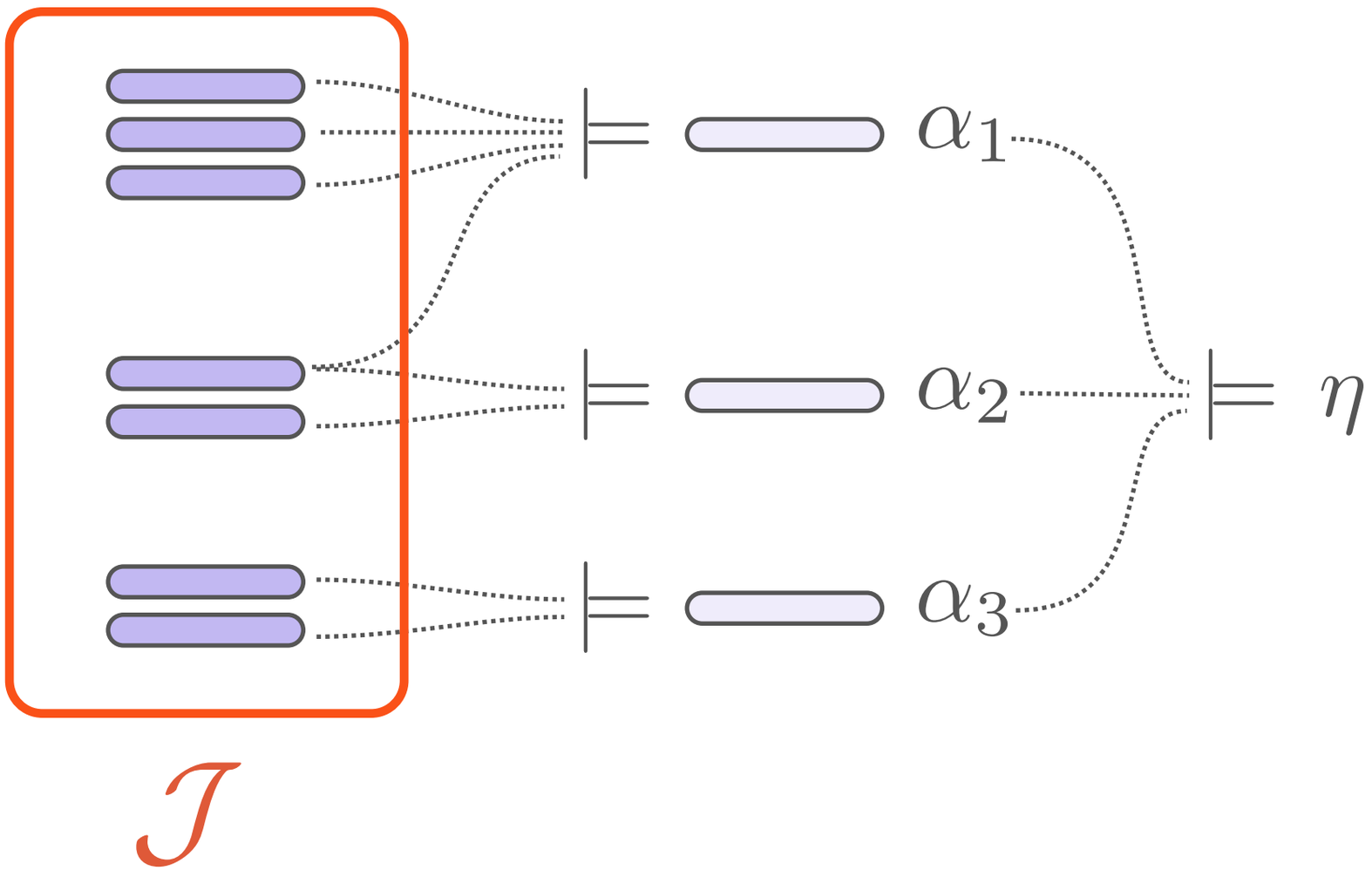
Overall Goal



Lemmas for Justifications



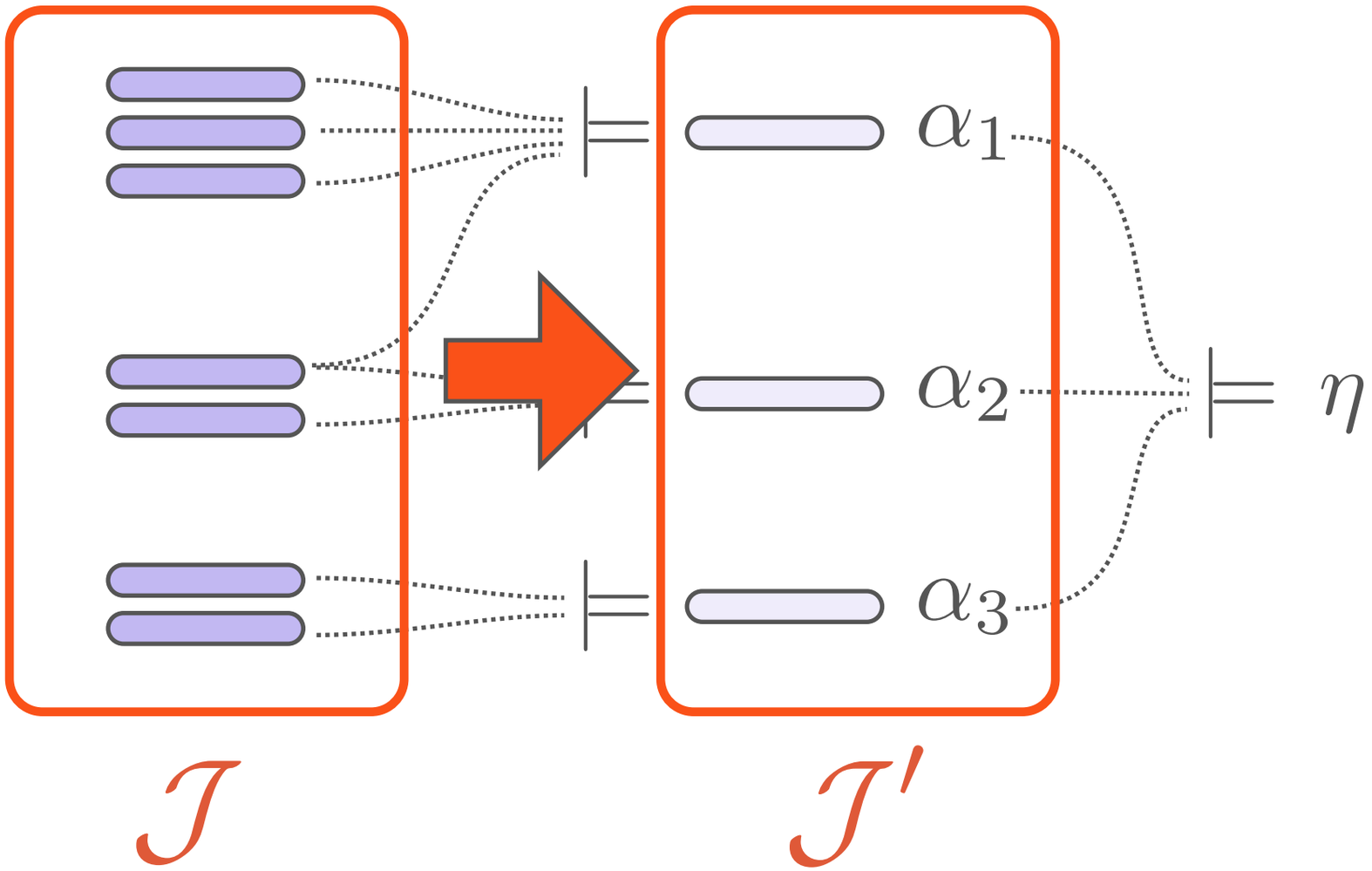
Lemmas for Justifications



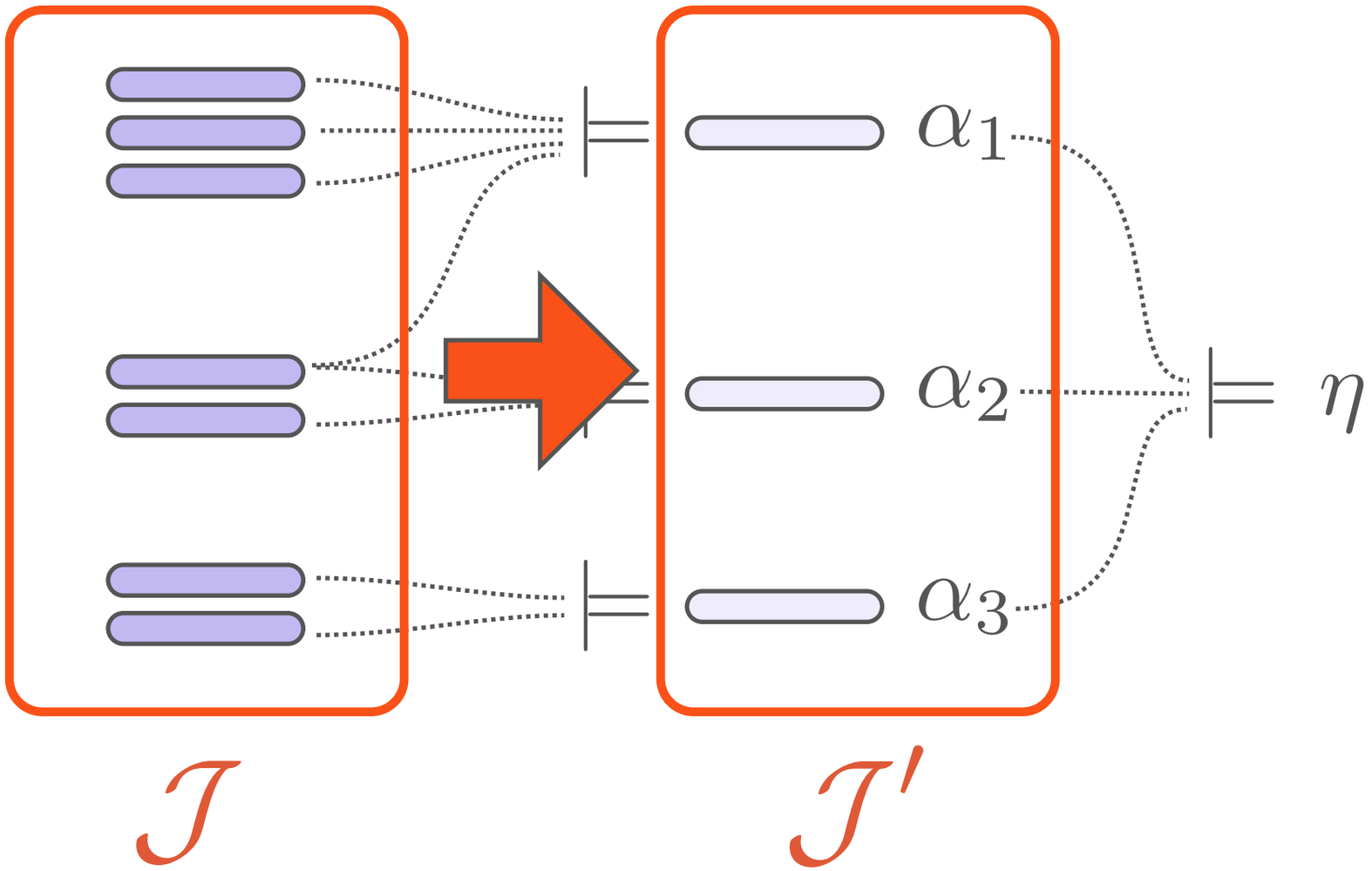
MANCHESTER 1824

The University of Manchester

Lemmas for Justifications

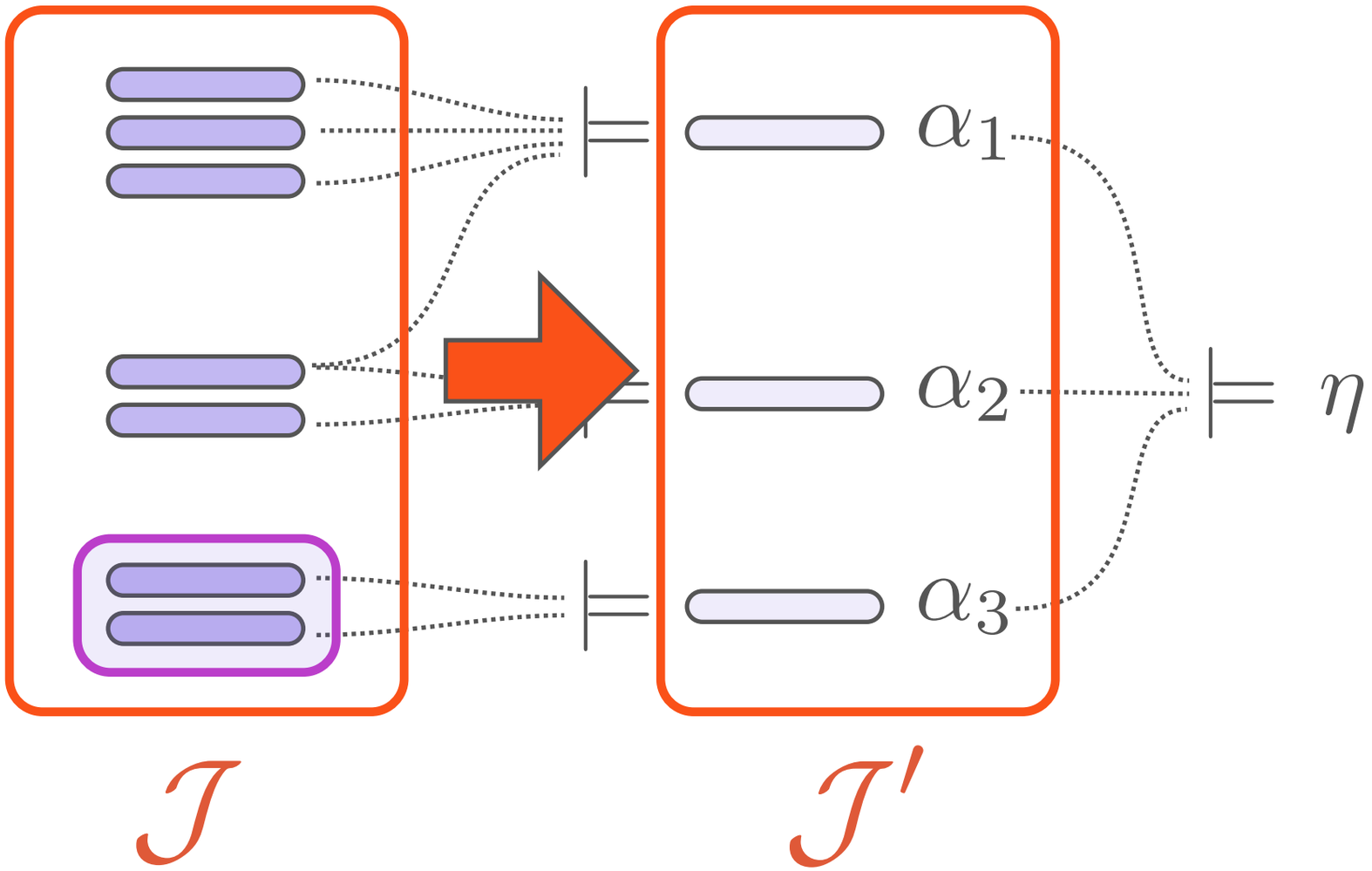


Lemmas for Justifications



Complexity$(\mathcal{J}, \eta) >$ Complexity$(\mathcal{J}', \eta)$

Lemmas for Justifications



Complexity$(\mathcal{J}, \eta) >$ Complexity$(\mathcal{J}', \eta)$

MANCHESTER 1824

The University of Manchester

Lemmas

$$\mathcal{J}$$

$$A \sqsubseteq B$$

$$B \sqsubseteq \exists R.D$$

$$R \sqsubseteq S$$

$$E \equiv \exists S.\top$$

$$\models A \sqsubseteq E$$

The University of Manchester

Lemmas

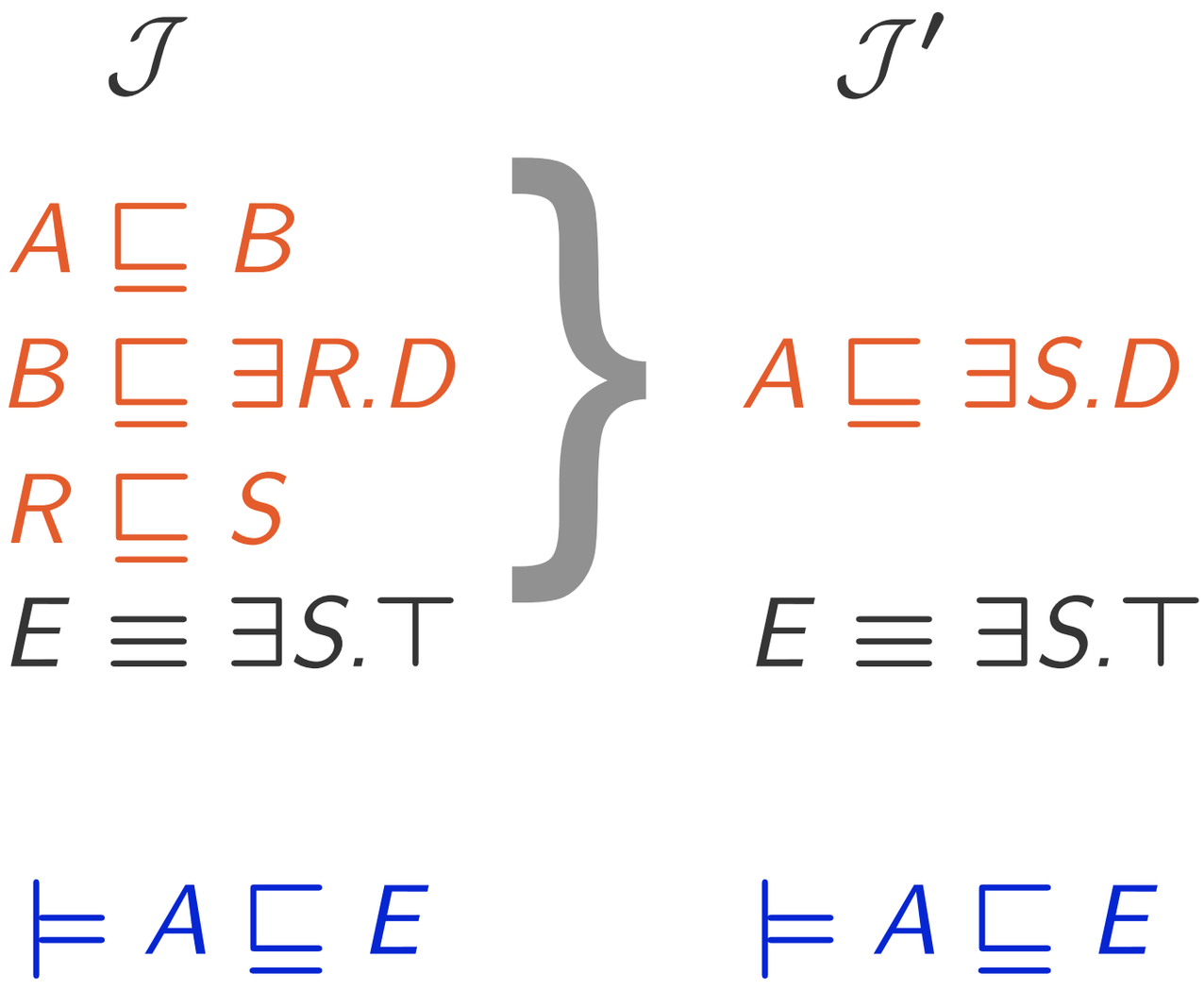
$$\mathcal{J}$$

$$\left.\begin{array}{l} A \sqsubseteq B \\ B \sqsubseteq \exists R.D \\ R \sqsubseteq S \\ E \equiv \exists S.\top \end{array}\right\} \quad A \sqsubseteq \exists S.D$$

$$\models A \sqsubseteq E$$

The University of Manchester

Lemmas



Example ctd.

$$\mathcal{J} = \{$$

$$\begin{aligned}
\text{Person} &\sqsubseteq \neg\text{Movie} \\
\text{RRated} &\sqsubseteq \text{CatMovie} \\
\text{CatMovie} &\sqsubseteq \text{Movie} \\
\text{RRated} &\equiv \exists\, \text{hasScript}.\text{ThrillerScript} \\
&\quad \sqcup\ \forall\, \text{hasViolenceLevel}.\text{High} \\
\exists\, \text{hasViolenceLevel} &\sqsubseteq \text{Movie}
\end{aligned}$$

$\}$

$$\models \text{Person} \sqsubseteq \bot$$

MANCHESTER 1824

The University of Manchester

A lemmatised explanation

Person $\sqsubseteq \bot$

MANCHESTER 1824

The University of Manchester

A lemmatised explanation

$\mathsf{Person} \sqsubseteq \bot$

$\mathsf{Person} \sqsubseteq \neg\mathsf{Movie}$

$\neg\mathsf{Movie} \sqsubseteq \bot$

MANCHESTER 1824

The University of Manchester

A lemmatised explanation

$\mathsf{Person} \sqsubseteq \bot$

$\mathsf{Person} \sqsubseteq \neg\mathsf{Movie}$

$\neg\mathsf{Movie} \sqsubseteq \bot$

$\neg\mathsf{Movie} \sqsubseteq \mathsf{Movie}$

A lemmatised explanation

$\mathsf{Person} \sqsubseteq \bot$

$\mathsf{Person} \sqsubseteq \neg\mathsf{Movie}$

$\neg\mathsf{Movie} \sqsubseteq \bot$

$\neg\mathsf{Movie} \sqsubseteq \mathsf{Movie}$

$\neg\mathsf{Movie} \sqsubseteq \neg\mathsf{RRated}$

$\neg\mathsf{RRated} \sqsubseteq \mathsf{Movie}$

A lemmatised explanation

$\mathsf{Person} \sqsubseteq \bot$

$\mathsf{Person} \sqsubseteq \neg\mathsf{Movie}$

$\neg\mathsf{Movie} \sqsubseteq \bot$

$\neg\mathsf{Movie} \sqsubseteq \mathsf{Movie}$

$\neg\mathsf{Movie} \sqsubseteq \neg\mathsf{RRated}$

$\mathsf{RRated} \sqsubseteq \mathsf{CatMovie}$

$\mathsf{CatMovie} \sqsubseteq \mathsf{Movie}$

$\neg\mathsf{RRated} \sqsubseteq \mathsf{Movie}$

A lemmatised explanation

$\mathsf{Person} \sqsubseteq \bot$

$\mathsf{Person} \sqsubseteq \neg\mathsf{Movie}$

$\neg\mathsf{Movie} \sqsubseteq \bot$

$\neg\mathsf{Movie} \sqsubseteq \mathsf{Movie}$

$\neg\mathsf{Movie} \sqsubseteq \neg\mathsf{RRated}$

$\mathsf{RRated} \sqsubseteq \mathsf{CatMovie}$

$\mathsf{CatMovie} \sqsubseteq \mathsf{Movie}$

$\neg\mathsf{RRated} \sqsubseteq \mathsf{Movie}$

$\neg\mathsf{RRated} \sqsubseteq \exists\mathsf{hasViolenceLevel}.\top$

$\exists\mathsf{hasViolenceLevel}.\top \sqsubseteq \mathsf{Movie}$

A lemmatised explanation

$\mathsf{Person} \sqsubseteq \bot$

$\mathsf{Person} \sqsubseteq \neg\mathsf{Movie}$

$\neg\mathsf{Movie} \sqsubseteq \bot$

$\neg\mathsf{Movie} \sqsubseteq \mathsf{Movie}$

$\neg\mathsf{Movie} \sqsubseteq \neg\mathsf{RRated}$

$\mathsf{RRated} \sqsubseteq \mathsf{CatMovie}$

$\mathsf{CatMovie} \sqsubseteq \mathsf{Movie}$

$\neg\mathsf{RRated} \sqsubseteq \mathsf{Movie}$

$\neg\mathsf{RRated} \sqsubseteq \exists \mathsf{hasViolenceLevel}.\top$

$\mathsf{RRated} \equiv \ldots \sqcup \forall \mathsf{hasViolenceLevel}.\mathsf{High}$

$\exists \mathsf{hasViolenceLevel}.\top \sqsubseteq \mathsf{Movie}$

Complexity Model

Two parts:

- Structure based
  considers structure of axioms and their interactions;
  abstracts from the logic
- Phenomena based
  considers certain patterns;
  takes the logic into account

MANCHESTER 1824

The University of Manchester

Complexity Model

Structure based criteria

- Number of different types of axioms and class expressions
- Signature flow
  How much is the signature spread within a justification?

MANCHESTER 1824

The University of Manchester

Complexity Model

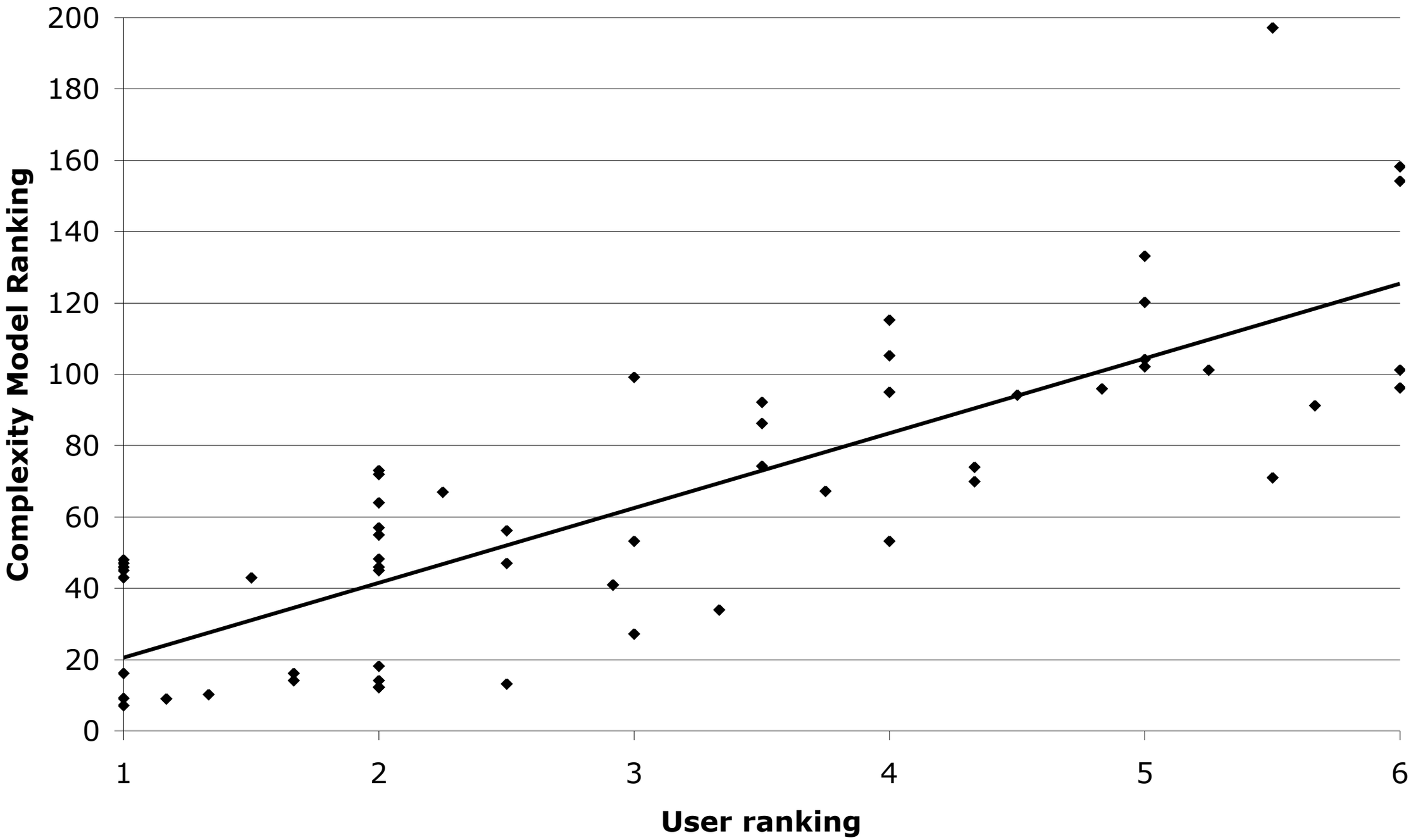
Phenomena based criteria

- Non-explicit synonyms of Top
  $\mathcal{J} \models \top \sqsubseteq A$ and this is not explicitly asserted in $\mathcal{O}$
- Universal implication
  $\mathcal{J}$ contains $\forall R.C \sqsubseteq \ldots$:
  $\forall R.C$ includes all individuals that have no R-successor
- Presence of GCIs (general concept inclusions)
  $\mathcal{J}$ contains $C \sqsubseteq D$, where C is not a class name

MANCHESTER 1824

The University of Manchester

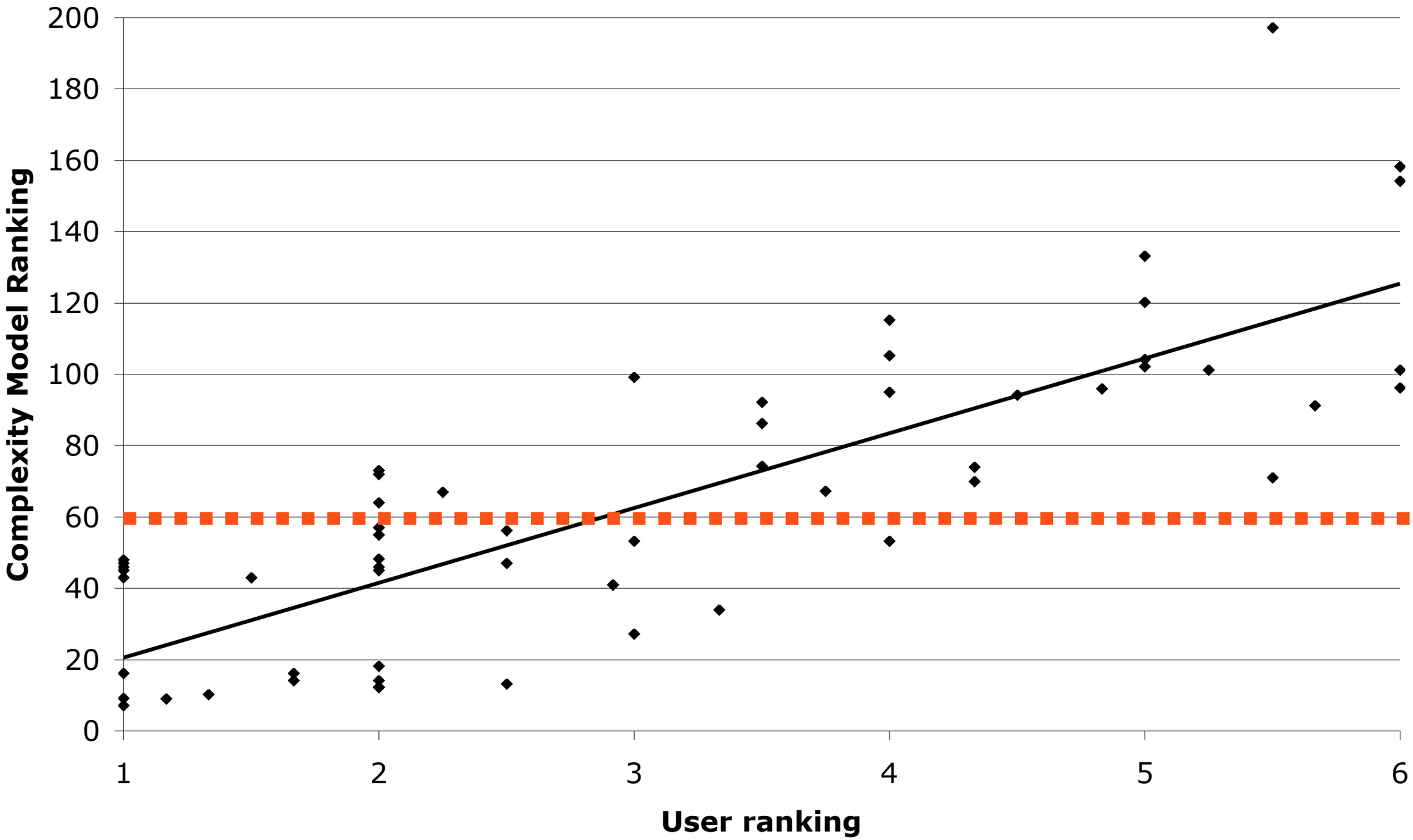
Participant Ranking versus Model Prediction



MANCHESTER 1824

The University of Manchester

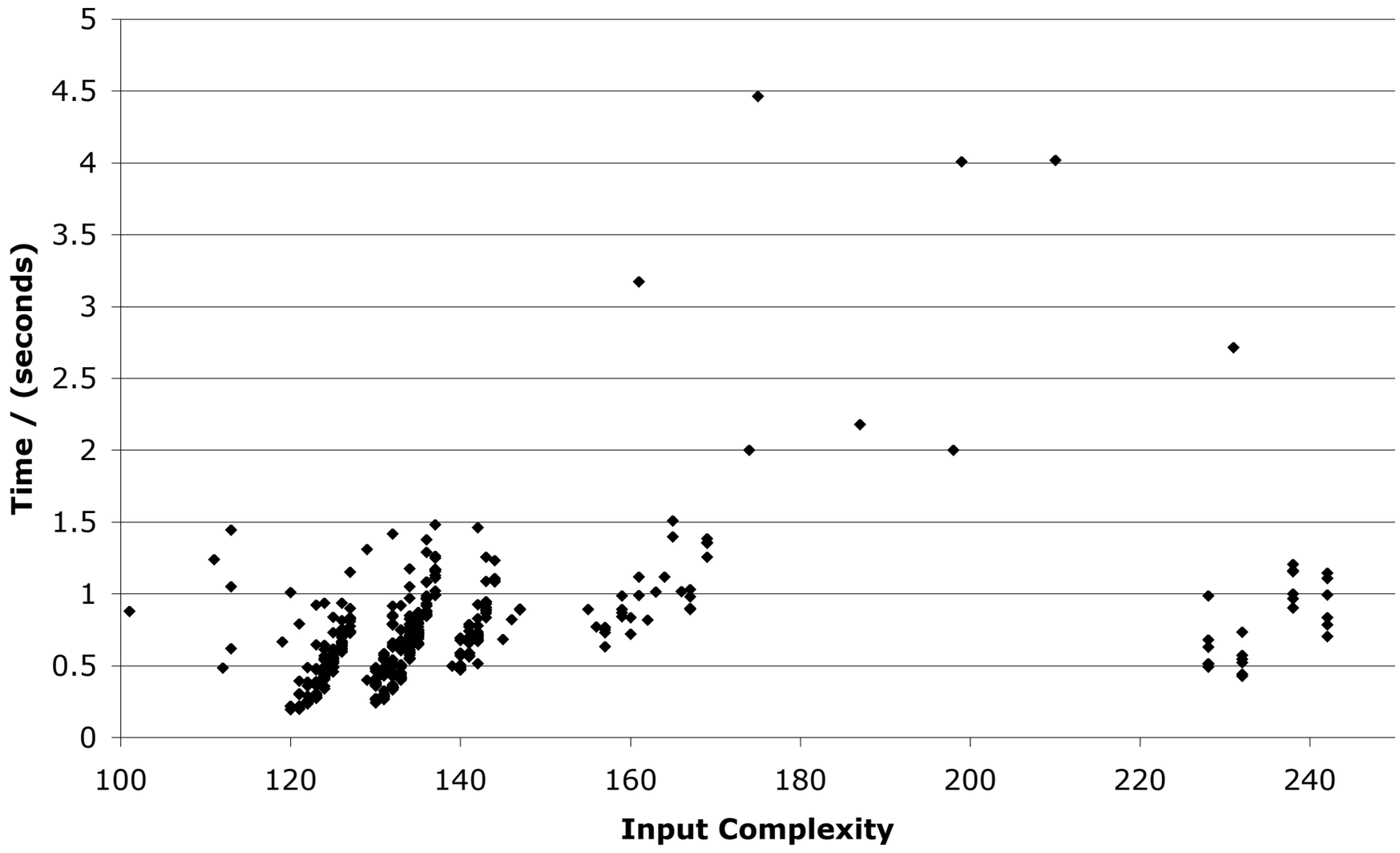
Participant Ranking versus Model Prediction



MANCHESTER 1824

The University of Manchester

Computing Lemmatised Justifications



MANCHESTER 1824

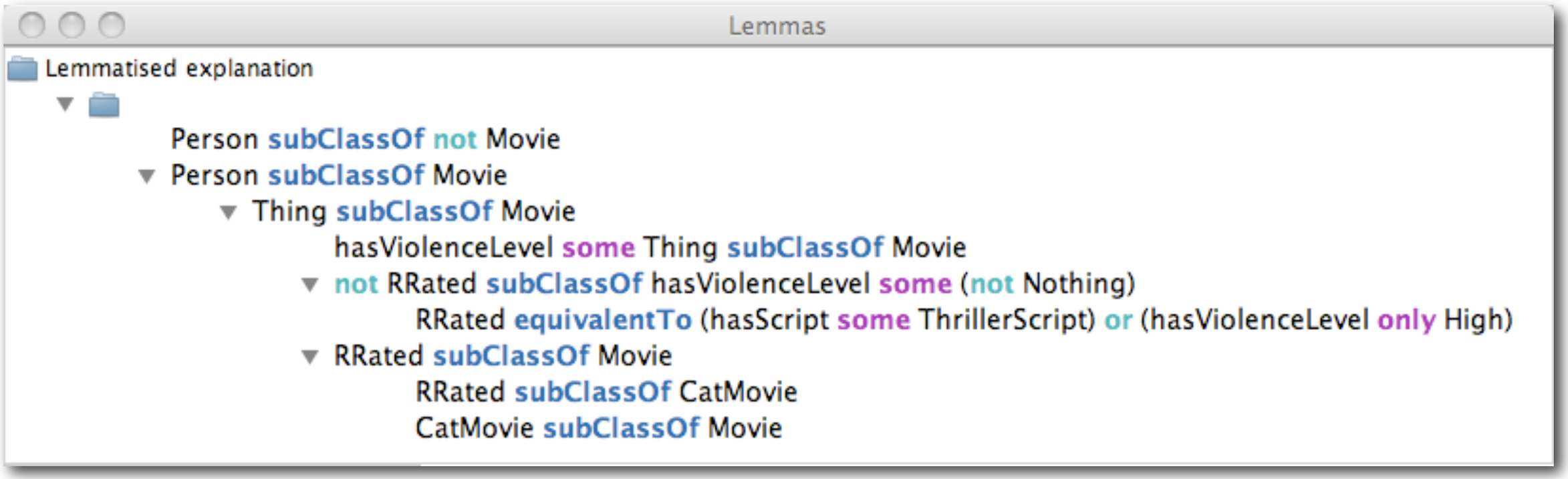
The University of Manchester

Time for a demo ...

- See the Explanation workbench at http://owl.cs.manchester.ac.uk/explanation

- This tool is work in progress.

Time for a demo ...



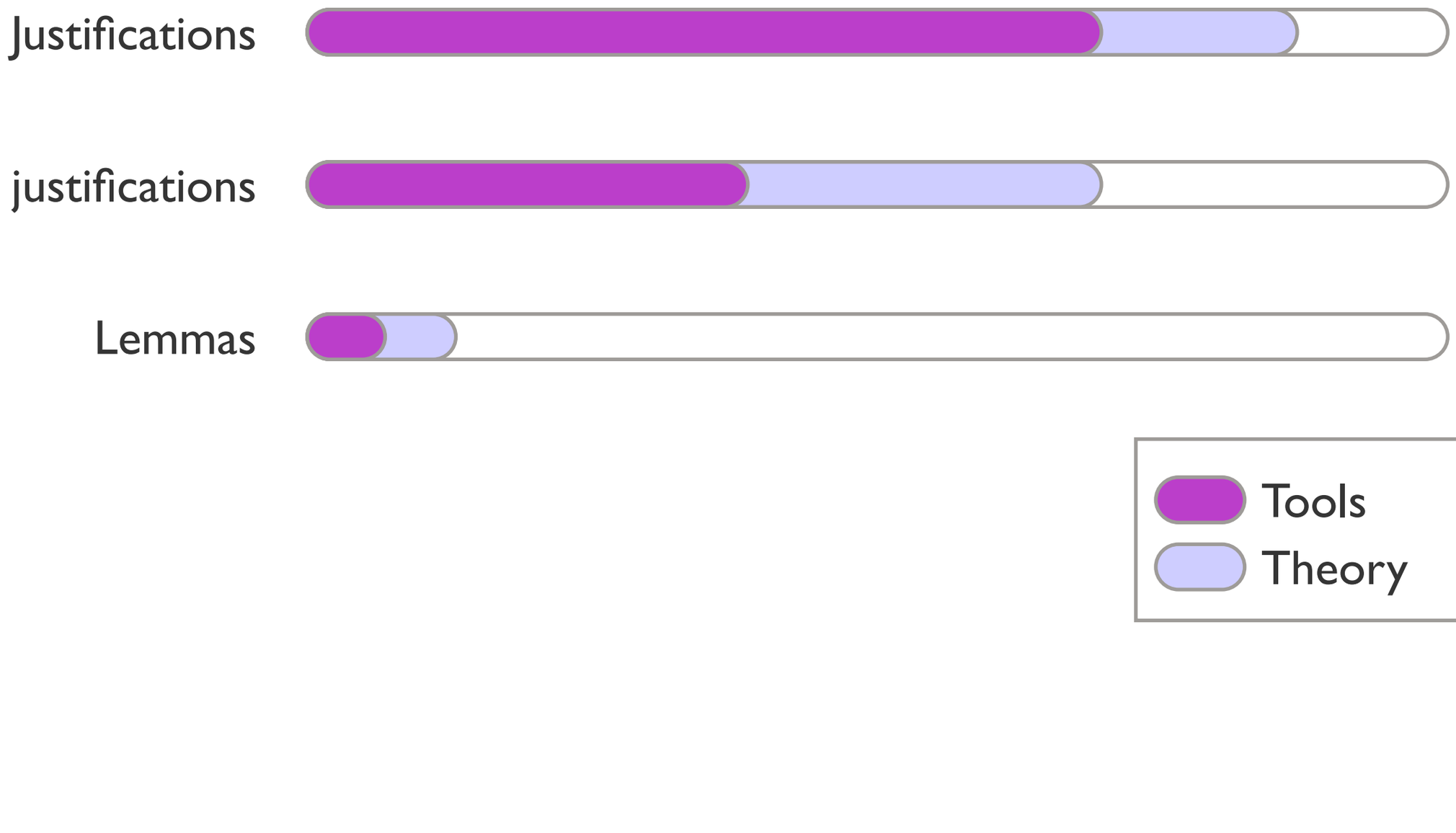
MANCHESTER 1824

The University of Manchester

Conclusions

- Justifications for entailments in real ontologies can be difficult or impossible to understand
- Lemmatised justifications presented as a solution
- Steps are driven by a complexity model
- Practical to compute lemmatised justifications
- For use in other services,
  e.g. construction of justification-oriented proofs

State of the art



Model interaction and computation

Current Explanation

- Revolves around explaining entailments
  - Esp. undesirable ones
- Focus 1: Isolating parts of the ontology
  - Justifications and Laconic justifications
- Focus 2: explicating how parts entail
  - Lemmas and Proofs

Current Explanation

- Revolves around explaining entailments
  - Esp. undesirable ones
- Focus 1: Isolating parts of the ontology
  - Justifications and Laconic justifications
- Focus 2: explicating how parts entail
  - Lemmas and Proofs

Proof-based Explanations!

MANCHESTER 1824

The University of Manchester

What about models?

- Logics are all about the models!
  - Ontologies describe a set of models
    - Users rarely, if ever, think about models
- Model-theoretic notions of quality
  - Categorial theories
  - Verified ontologies
- If you don't know the models...do you understand?

MANCHESTER 1824

The University of Manchester

Obviously Yes

- Model-oblivious users obviously function
  - And often well
    - At various tasks
- But clearly, there is information in models
  - That might help
  - Er...at something!!!

Problems

- Models come in all shapes and sizes
  - Often infinite
  - Often hard to compute
- Infinite numbers of models
  - Even without isomorphism
- So, which ones are good for what?

Find some task!!

- A Classic Story: Non-entailment
  1. A countermodel is an (the best) explanation
  2. We get them for free from reasoners!
  3. Dump and display!!
- Sound familiar?

MANCHESTER 1824

The University of Manchester

Eek

MANCHESTER 1824

The University of Manchester

Eek

MANCHESTER 1824

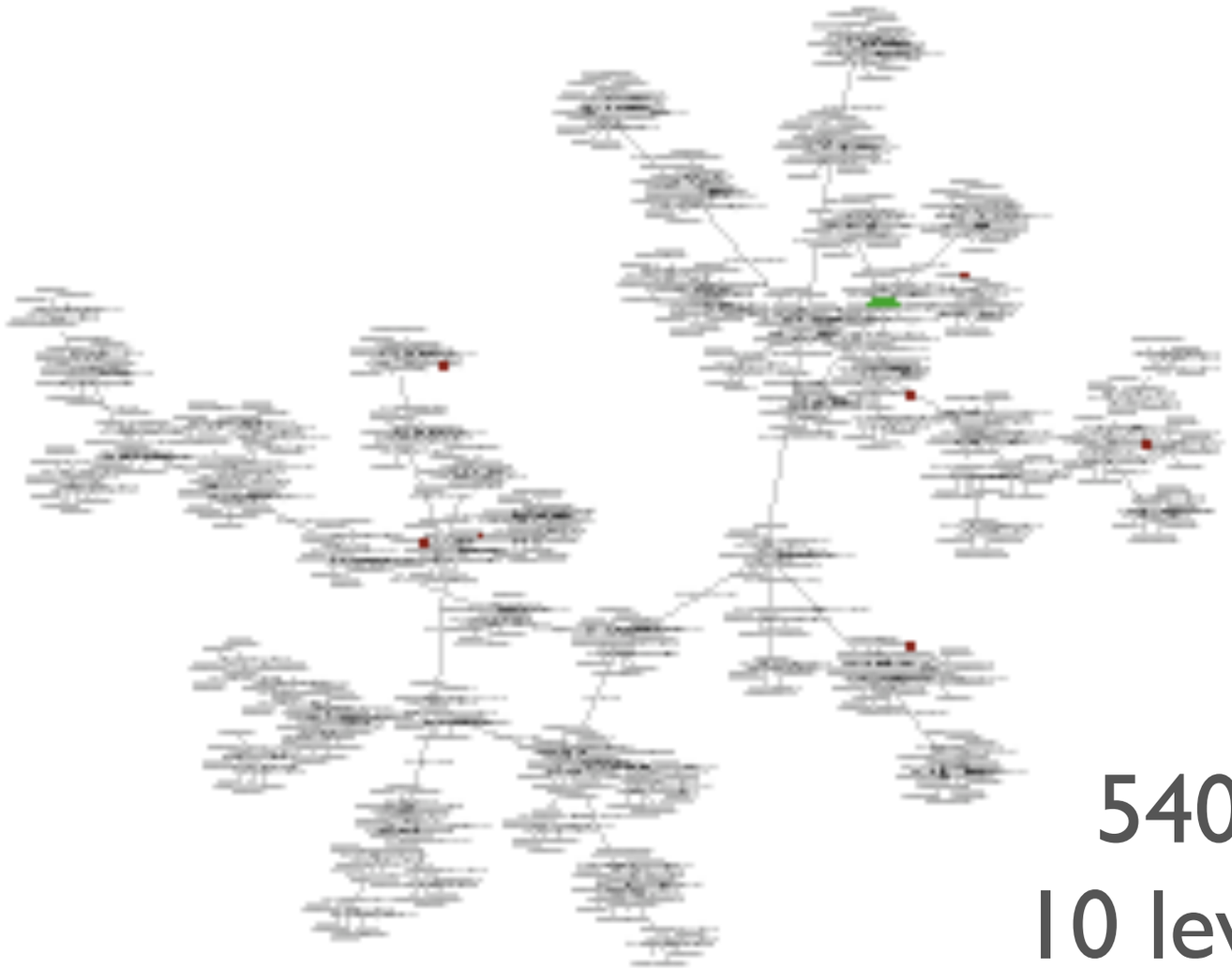
The University of Manchester

Eek

MANCHESTER 1824

The University of Manchester

Eek



540 nodes

10 levels deep

Red nodes indicate “blocking”

MANCHESTER 1824

The University of Manchester

Additional Problem

- Which non-entailments are wrong?
  - Detecting non-entailment is "easy"
    - Detecting wrong non-entailments….?!?!?
      - Domain dependence?
      - User state of mind?
- No canonical test cases

Other tasks?

- Ontology profiling
  - Problem: Model size kills reasoner
    - Goal: Additional constraints to guide the reasoner
- Concept understanding
  - What does some instance "look like"

Looking at Tweezers

Models

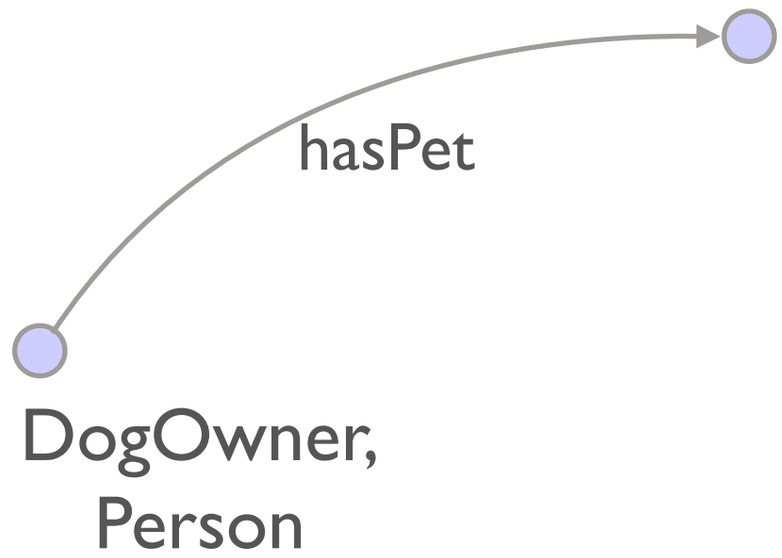
```
DogOwner subClassOf Person and hasPet some Dog
```

Models

```
DogOwner subClassOf Person and hasPet some Dog
```

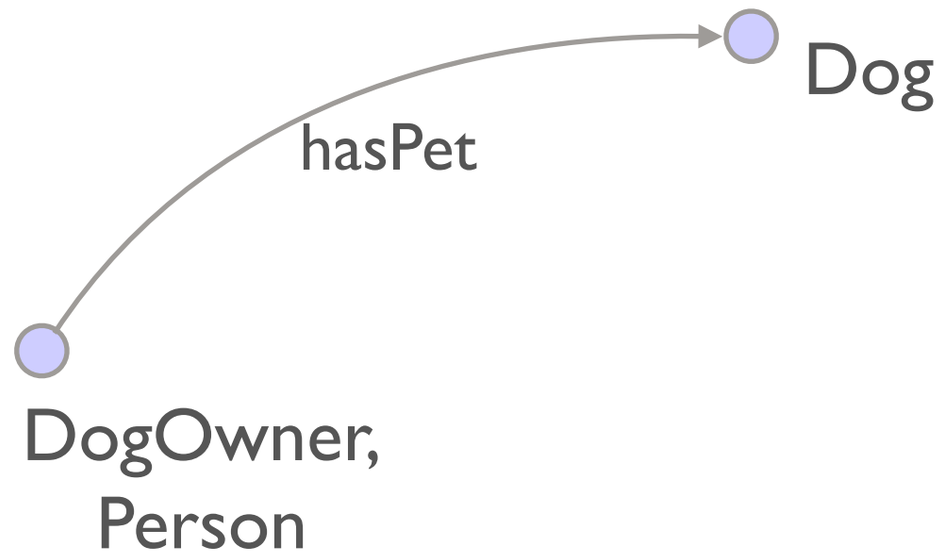
Models

```
DogOwner subClassOf Person and hasPet some Dog
```

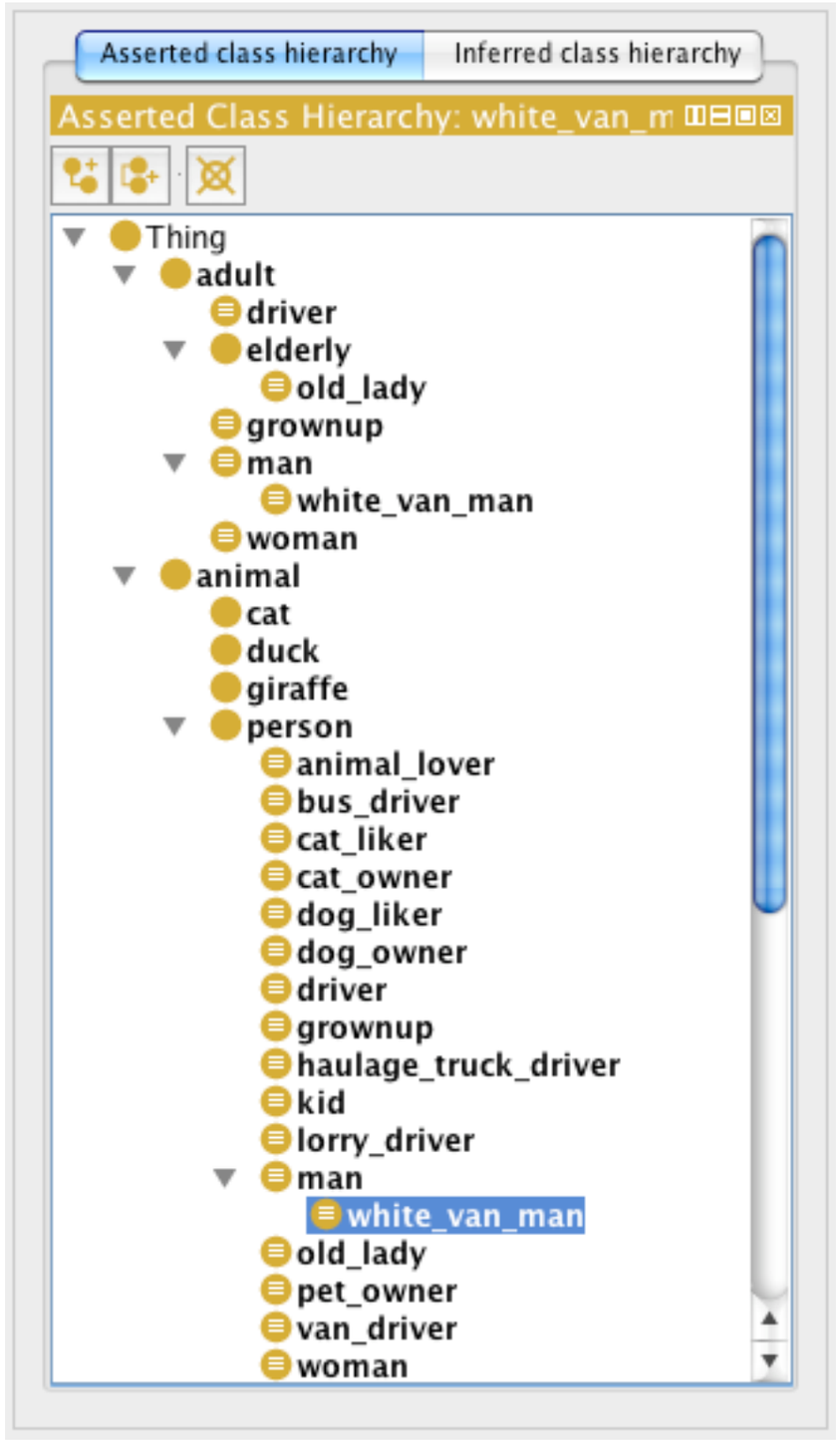


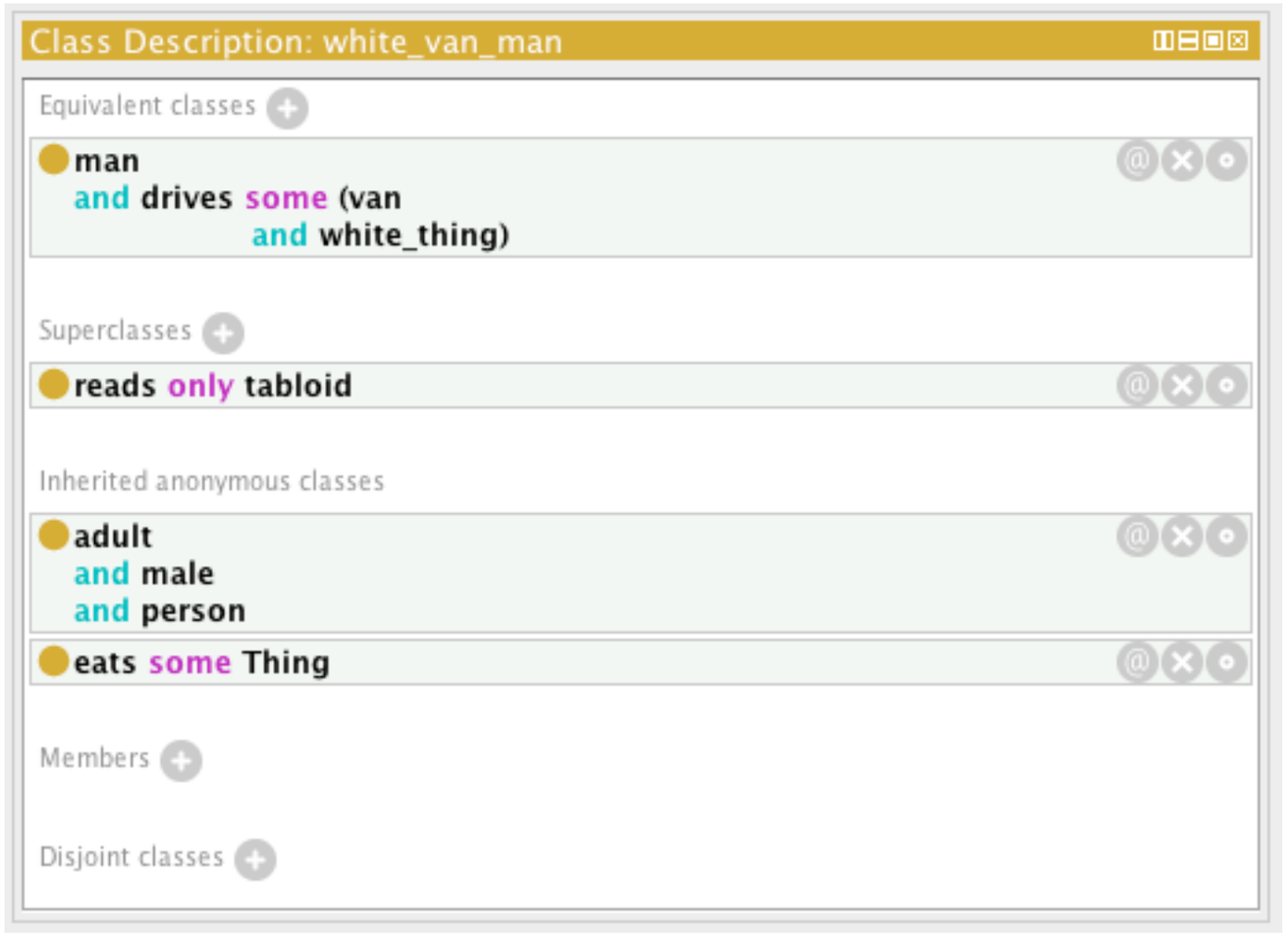
Models

```
DogOwner subClassOf Person and hasPet some Dog
```

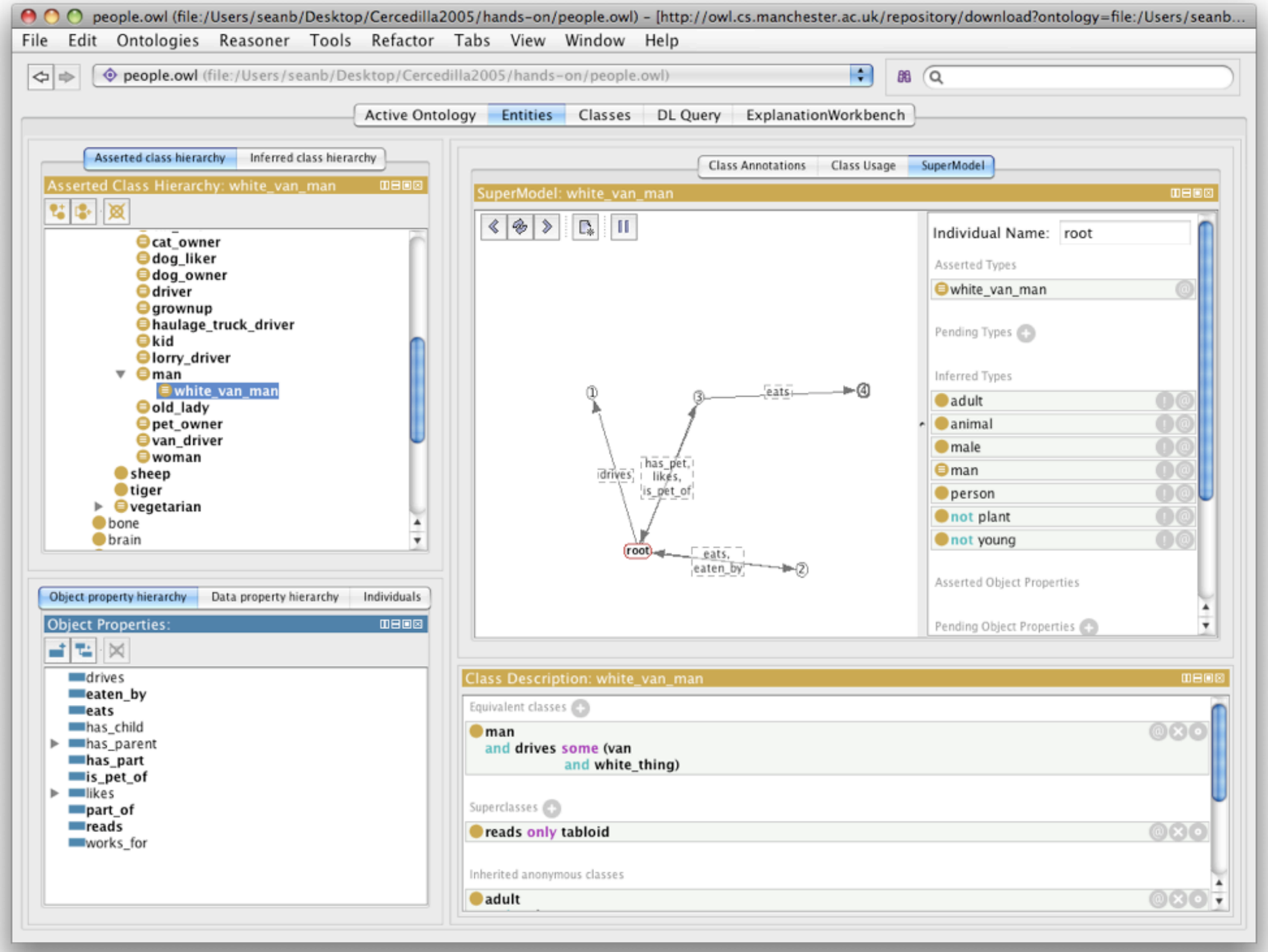


SuperModel

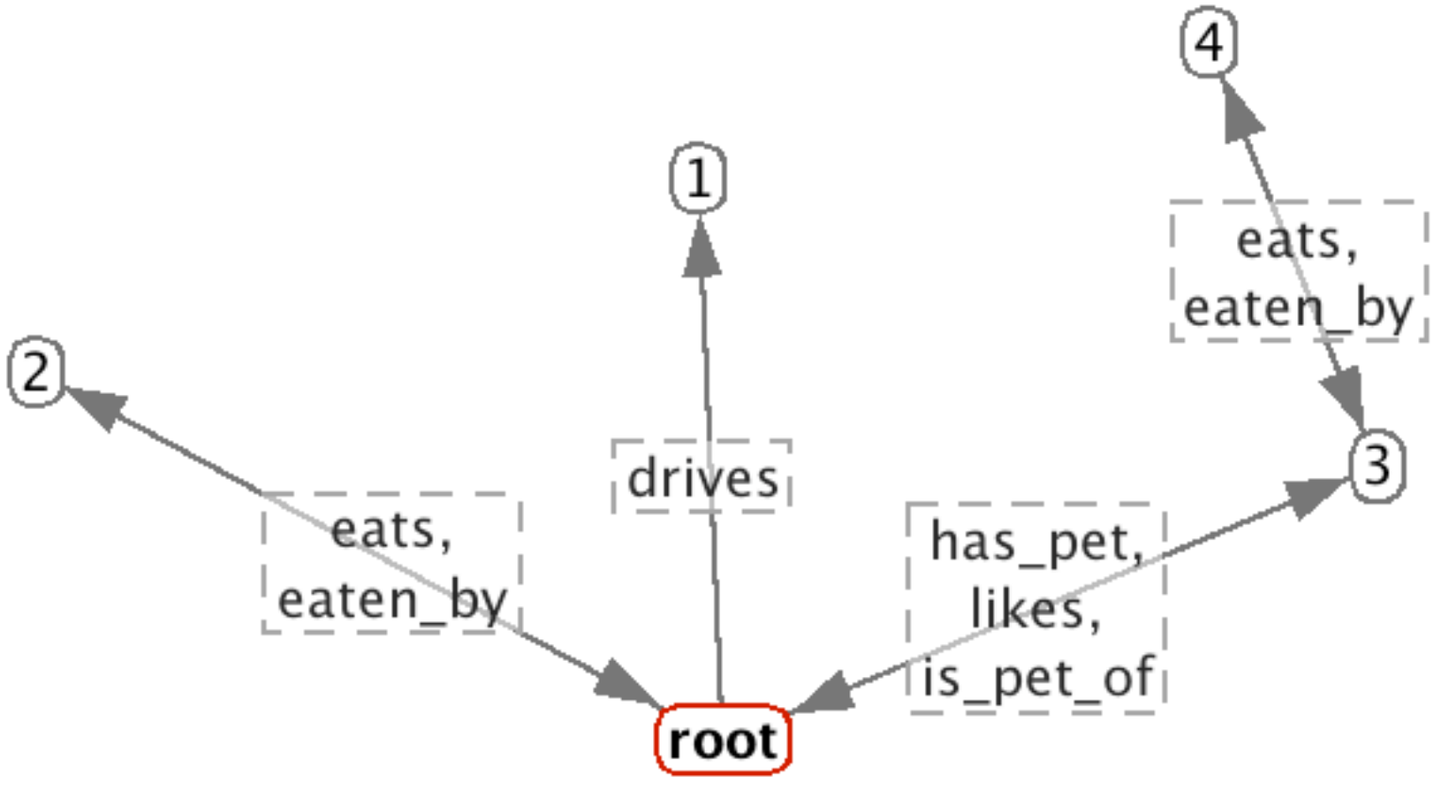

Asserted class hierarchy
Inferred class hierarchy
Asserted Class Hierarchy: white_van_m
Thing
adult
driver
elderly
old_lady
grownup
man
white_van_man
woman
animal
cat
duck
giraffe
person
animal_lover
bus_driver
cat_liker
cat_owner
dog_liker
dog_owner
driver
grownup
haulage_truck_driver
kid
lorry_driver
man
white_van_man
old_lady
pet_owner
van_driver
woman


Class Description: white_van_man
Equivalent classes
man
and drives some (van
and white_thing)
Superclasses
reads only tabloid
Inherited anonymous classes
adult
and male
and person
eats some Thing
Members
Disjoint classes

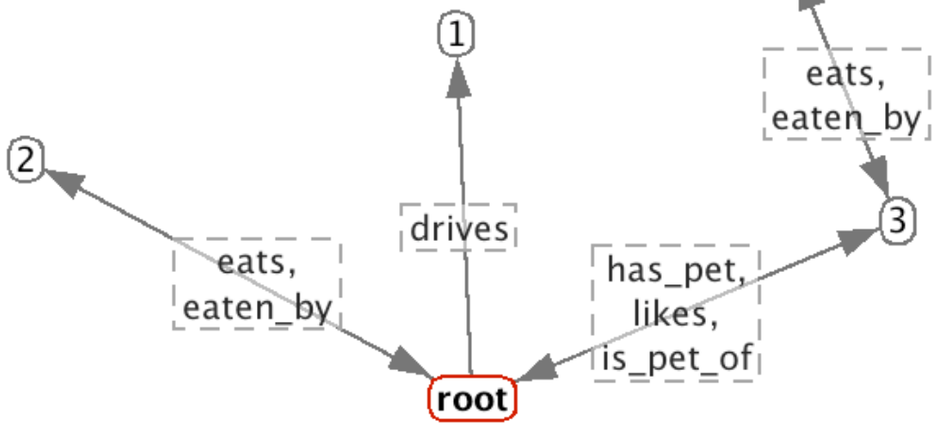
SuperModel


people.owl (file:/Users/seanb/Desktop/Cercedilla2005/hands-on/people.owl) - [http://owl.cs.manchester.ac.uk/repository/download?ontology=file:/Users/seanb...
File Edit Ontologies Reasoner Tools Refactor Tabs View Window Help
people.owl (file:/Users/seanb/Desktop/Cercedilla2005/hands-on/people.owl)
Active Ontology Entities Classes DL Query ExplanationWorkbench
Asserted class hierarchy Inferred class hierarchy
Asserted Class Hierarchy: white_van_man
cat_owner
dog_liker
dog_owner
driver
grownup
haulage_truck_driver
kid
lorry_driver
man
white_van_man
old_lady
pet_owner
van_driver
woman
sheep
tiger
vegetarian
bone
brain
Class Annotations Class Usage SuperModel
SuperModel: white_van_man
drives
has_pet,
likes,
is_pet_of
eats
eats,
eaten_by
root
Individual Name: root
Asserted Types
white_van_man
Pending Types
Inferred Types
adult
animal
male
man
person
not plant
not young
Asserted Object Properties
Pending Object Properties
Object property hierarchy Data property hierarchy Individuals
Object Properties:
drives
eaten_by
eats
has_child
has_parent
has_part
is_pet_of
likes
part_of
reads
works_for
Class Description: white_van_man
Equivalent classes
man
and drives some (van
and white_thing)
Superclasses
reads only tabloid
Inherited anonymous classes
adult

SuperModel


1
2
3
4
eats,
eaten_by
drives
eats,
eaten_by
has_pet,
likes,
is_pet_of
root

SuperModel



Individual Name: root

Asserted Types

- white_van_man

Pending Types

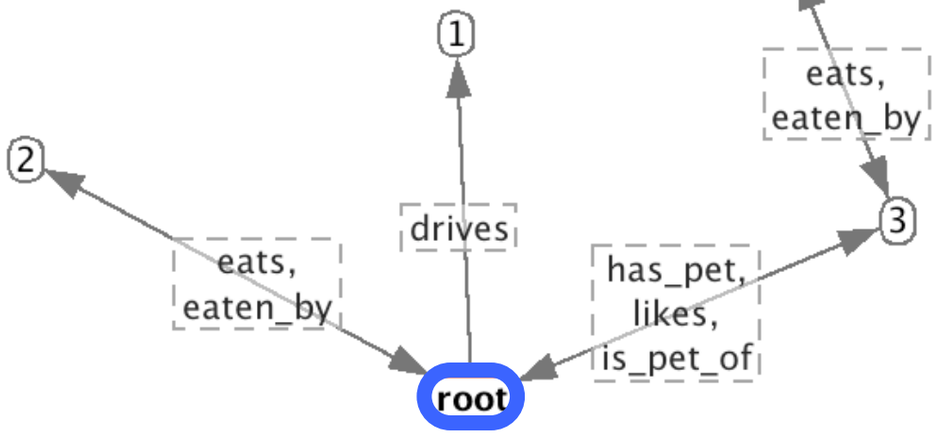
Inferred Types

- adult
- animal
- male
- man
- person
- not plant

Inferred Types

- van
- vehicle
- white_thing

SuperModel



Individual Name: root

Asserted Types

white_van_man

Pending Types

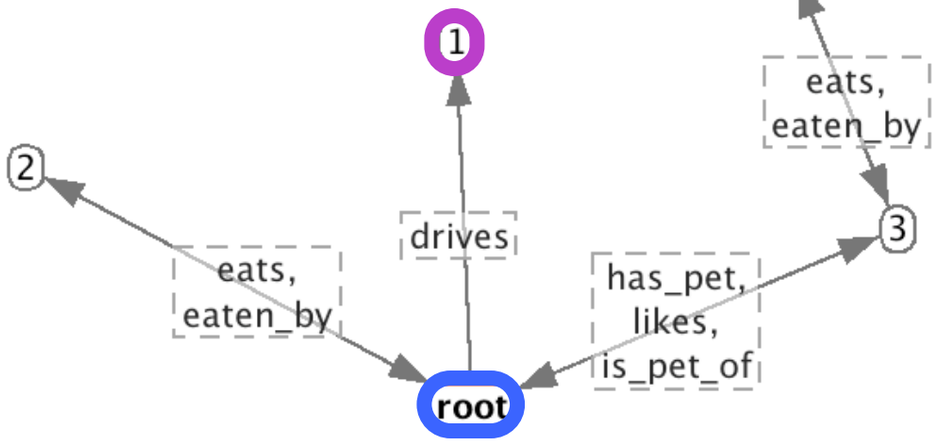
Inferred Types

adult
animal
male
man
person
not plant

Inferred Types

van
vehicle
white_thing

SuperModel



Individual Name: root

Asserted Types

white_van_man

Pending Types

Inferred Types

adult

animal

male

man

person

not plant

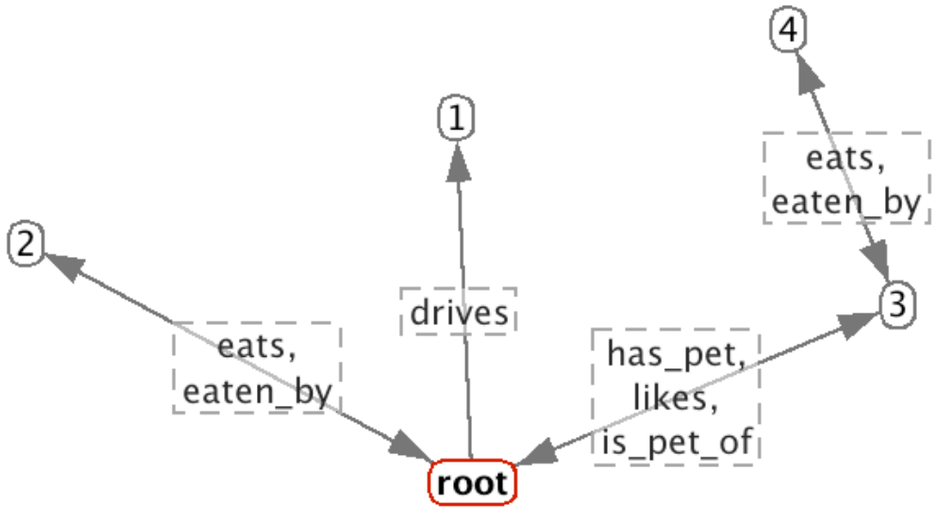
Inferred Types

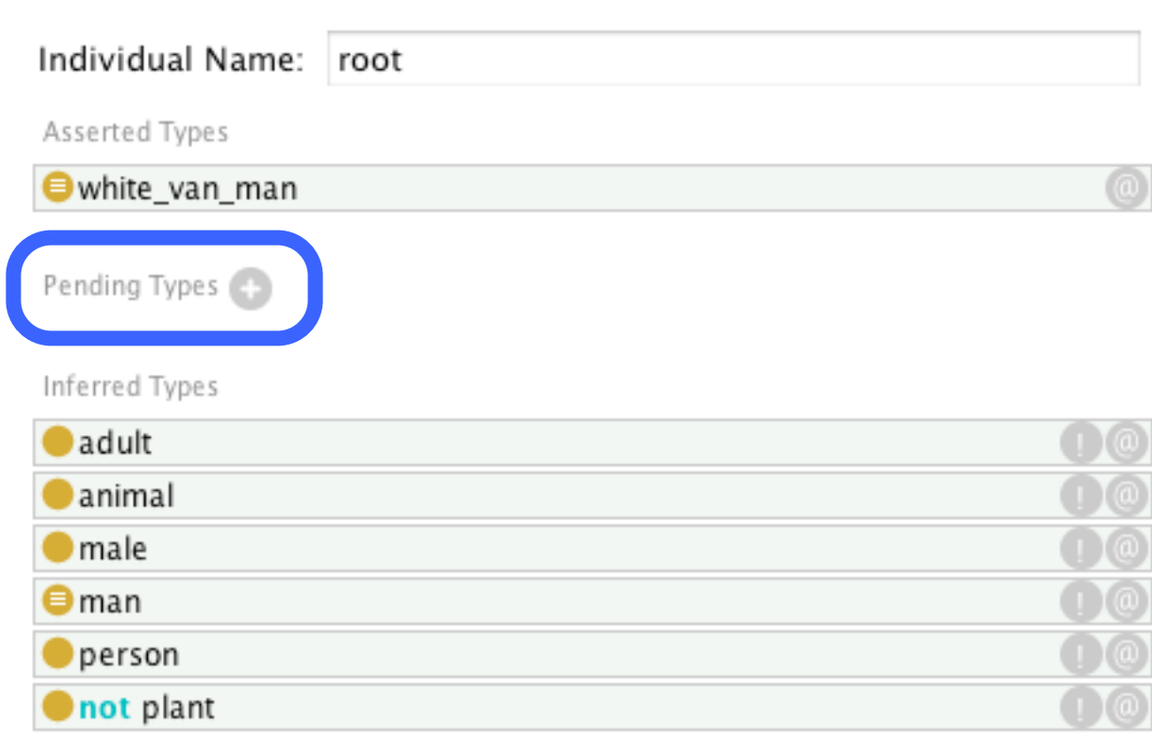
van

vehicle

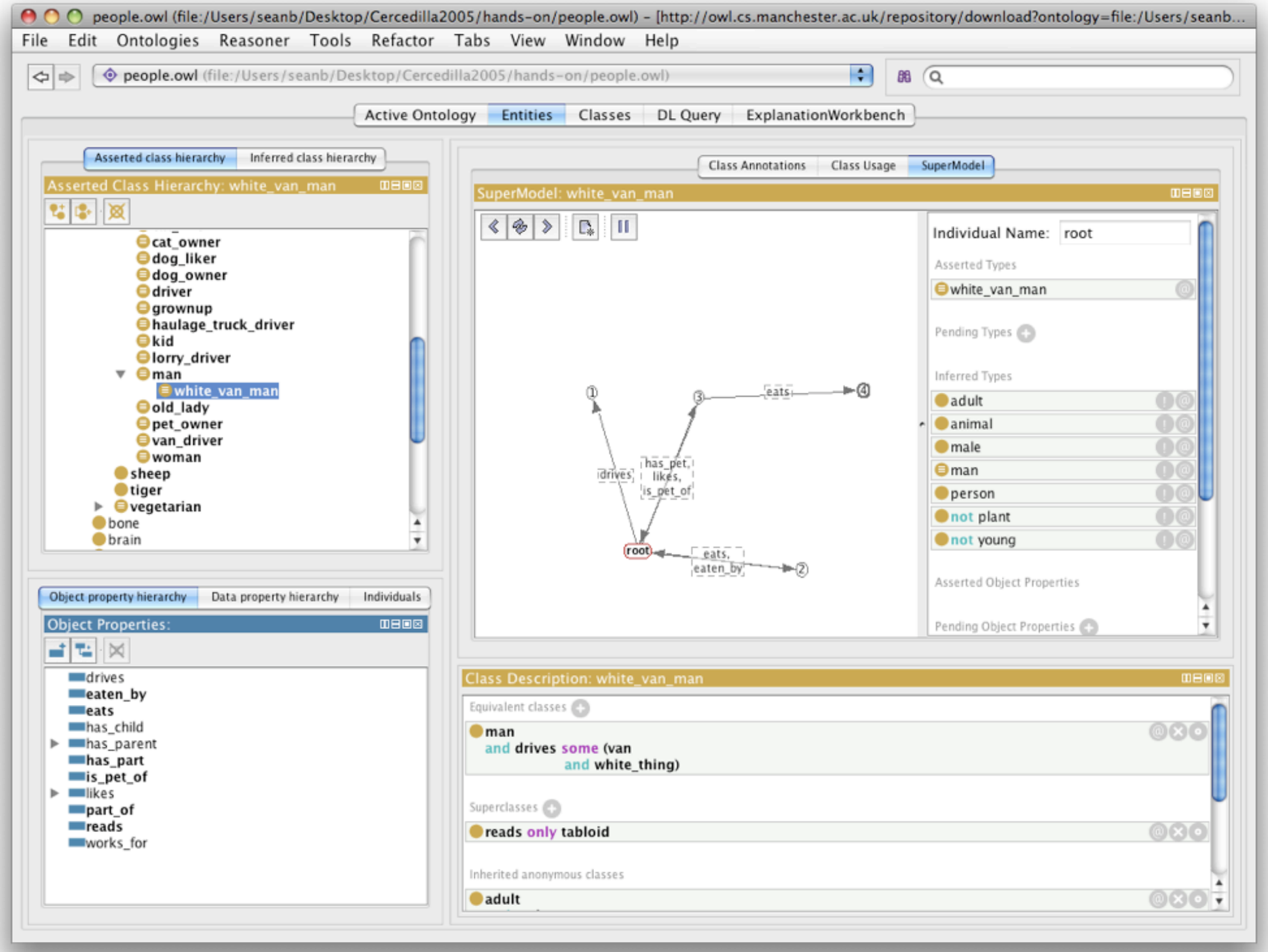
white_thing

SuperModel


1
2
3
4
eats,
eaten_by
drives
eats,
eaten_by
has_pet,
likes,
is_pet_of
root


Individual Name: root
Asserted Types
white_van_man
Pending Types
Inferred Types
adult
animal
male
man
person
not plant

SuperModel


people.owl (file:/Users/seanb/Desktop/Cercedilla2005/hands-on/people.owl) - [http://owl.cs.manchester.ac.uk/repository/download?ontology=file:/Users/seanb...
File Edit Ontologies Reasoner Tools Refactor Tabs View Window Help
people.owl (file:/Users/seanb/Desktop/Cercedilla2005/hands-on/people.owl)
Active Ontology Entities Classes DL Query ExplanationWorkbench
Asserted class hierarchy Inferred class hierarchy
Asserted Class Hierarchy: white_van_man
cat_owner
dog_liker
dog_owner
driver
grownup
haulage_truck_driver
kid
lorry_driver
man
white_van_man
old_lady
pet_owner
van_driver
woman
sheep
tiger
vegetarian
bone
brain
Class Annotations Class Usage SuperModel
SuperModel: white_van_man
drives
has_pet,
likes,
is_pet_of
eats
eats,
eaten_by
root
Individual Name: root
Asserted Types
white_van_man
Pending Types
Inferred Types
adult
animal
male
man
person
not plant
not young
Asserted Object Properties
Pending Object Properties
Object property hierarchy Data property hierarchy Individuals
Object Properties:
drives
eaten_by
eats
has_child
has_parent
has_part
is_pet_of
likes
part_of
reads
works_for
Class Description: white_van_man
Equivalent classes
man
and drives some (van
and white_thing)
Superclasses
reads only tabloid
Inherited anonymous classes
adult

MANCHESTER 1824

The University of Manchester

Time for a demo ...

See SuperModel at
http://www.cs.man.ac.uk/~bauerj/supermodel/

MANCHESTER 1824

The University of Manchester

- ...