Methods in AI Research
Closing lecture

**Dragan Doder, Chris Janssen,
Dong Nguyen, Dominik Klein**

Recap: Methods in AI Research

The course

- **that introduces you to a great variety of Utrecht topics in AI**
- **that teaches you various AI skills**
- **to improve skills that need improving**
- **in which you get acquainted with your fellow students**

Recap: aims of MAIR

Know and understand the basics of

- **Agents**

  Dialogue modelling, autonomous dialogue systems
- **Reasoning**

  Knowledge-based systems, logic, argumentation
- **Cognitive Processing**

  Cognitive modelling, experimentation, interaction
- **Natural Language**

  Natural language processing
- **Machine Learning**

  Decision trees, the learning problem

Today's structure

1. About the mock exam – general comments
2. Discussion on the mock exam and answers to your questions
   - Dragan's part
   - Dong's part
   - Chris' part
3. Individual questions via MS Teams (we will stay online)

Exam set-up

- Online exam on **Remindo**:
  - 8 November 15:15–17:45 (15:15–18:10 with extra time)
  - Make sure you have stable internet connection
    - If you are worried about being stuck without internet during the exam, please make sure you have mobile data and can at least communicate with us by MS Teams / email;
    - The important timing when you **must** have internet is when you *log in* and when you *submit the test;*
    - You can work offline during the time in between, if you **do not** reboot the computer or restart the browser.
  - The exam will be an “open book” exam;
  - You need to work *individually;*
  - There will be a random *ID check* procedure via MS Teams (we can call you to check your ID).

Mock exam vs Final exam

Similarities:

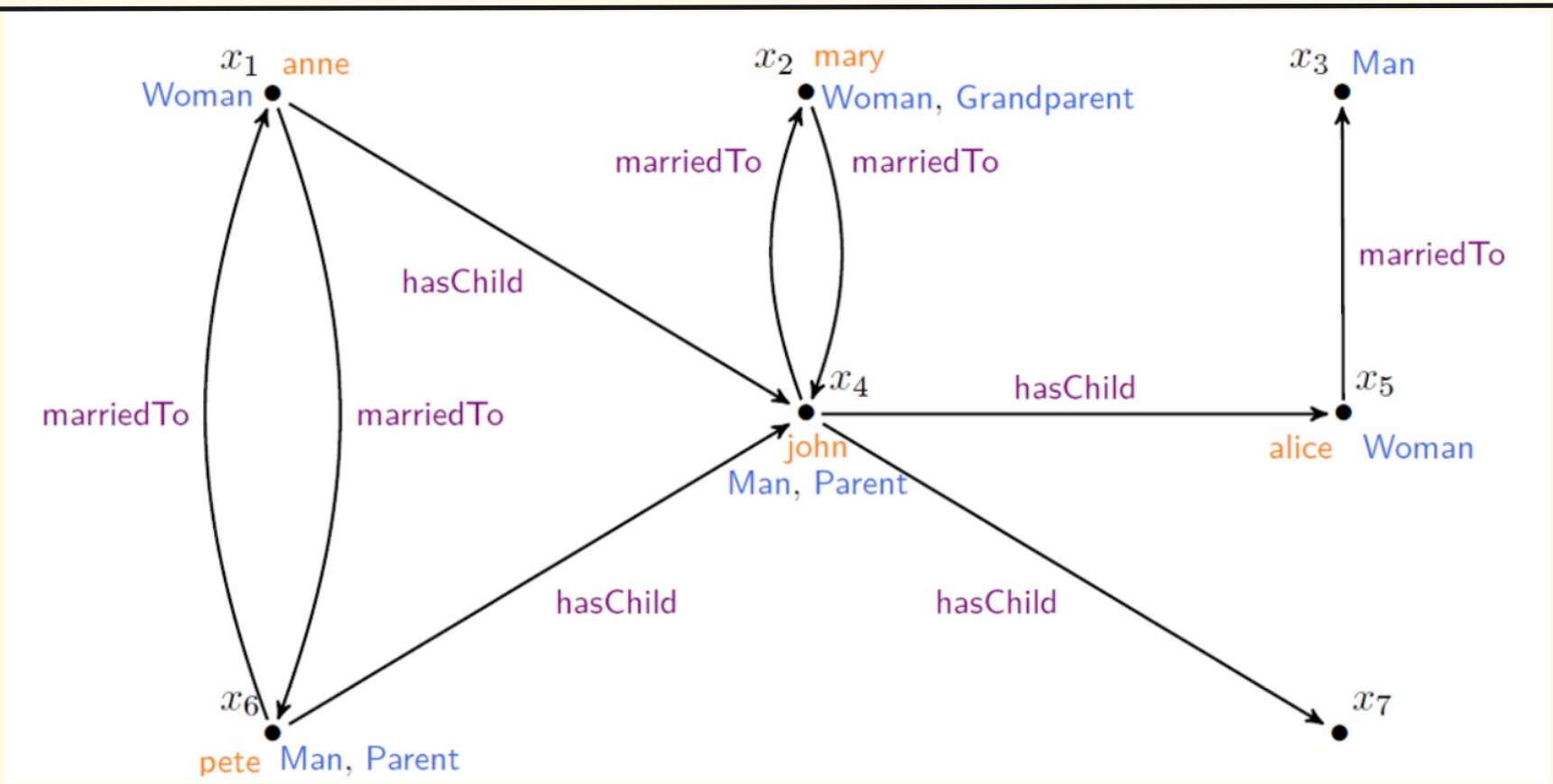
- **Similar degree of complexity**
- **Similar type of questions**

But:

- **There will be more questions on the final exam**
- **Some topics that were not present on the mock exam are possible on the final exam**
- **Points will not be uniformly distributed**

Dragan’s part

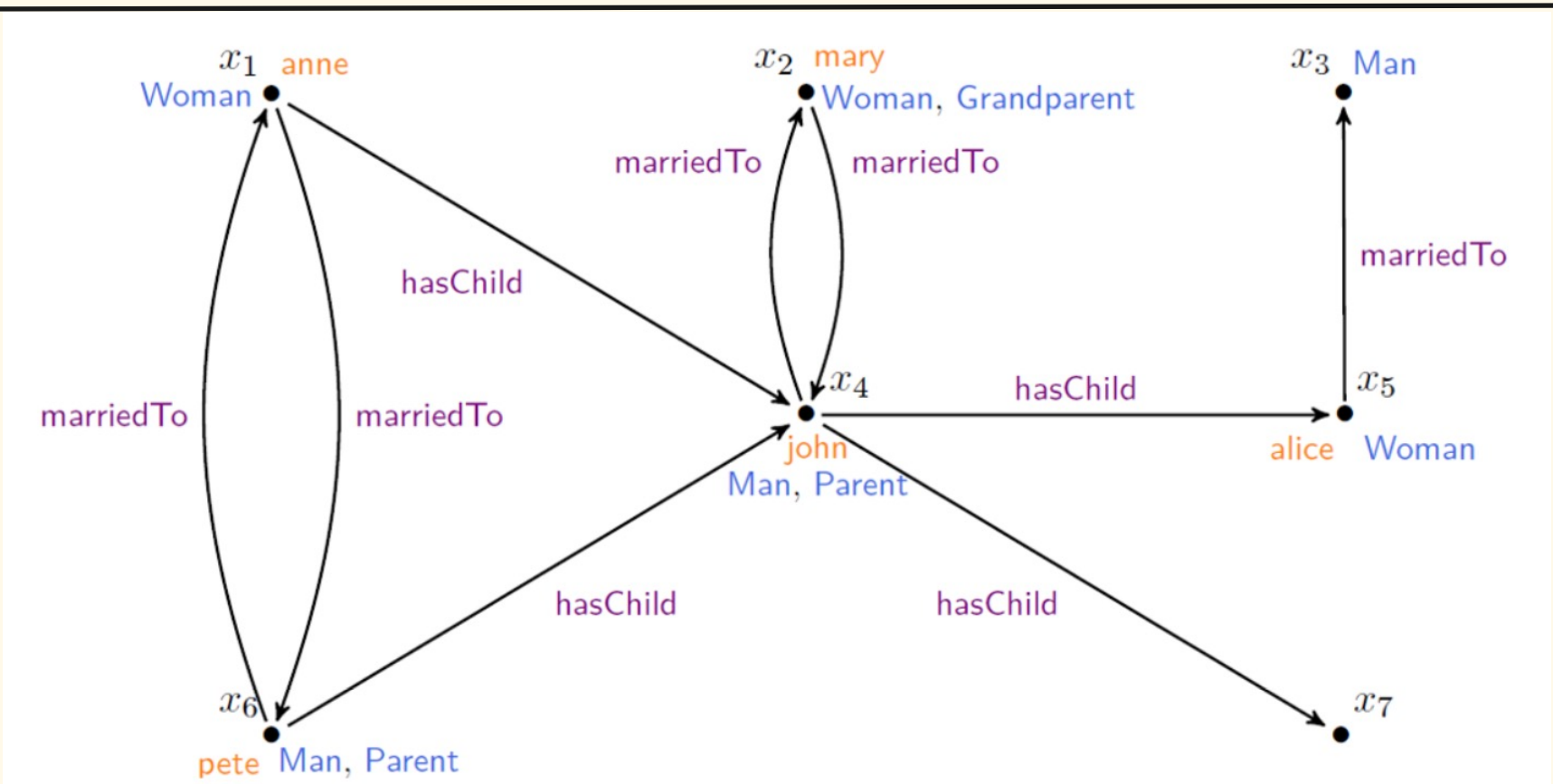
Mock exam – Question 1



$$\exists hasChild.\top \sqcap \neg \exists hasChild.Woman$$

List all the elements of the set obtained by applying the interpretation function from the interpretation depicted above to the given concept.

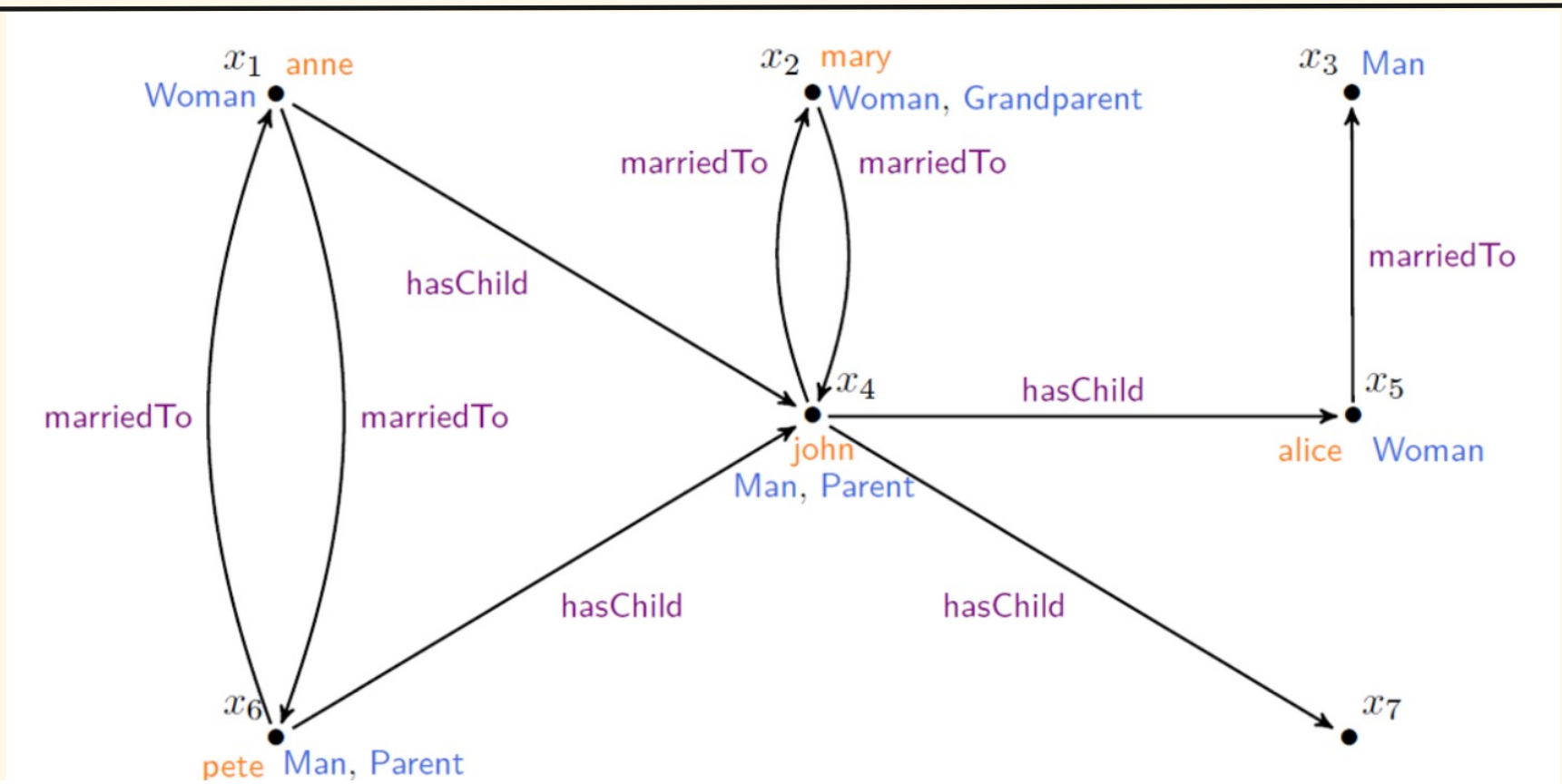
Mock exam – Question 1



$\exists hasChild.\top$ - **The individuals that have *some* "has child" relation with an element of $\top$ (= have at least one child)**

{x1, x6, x4}

Mock exam – Question 1

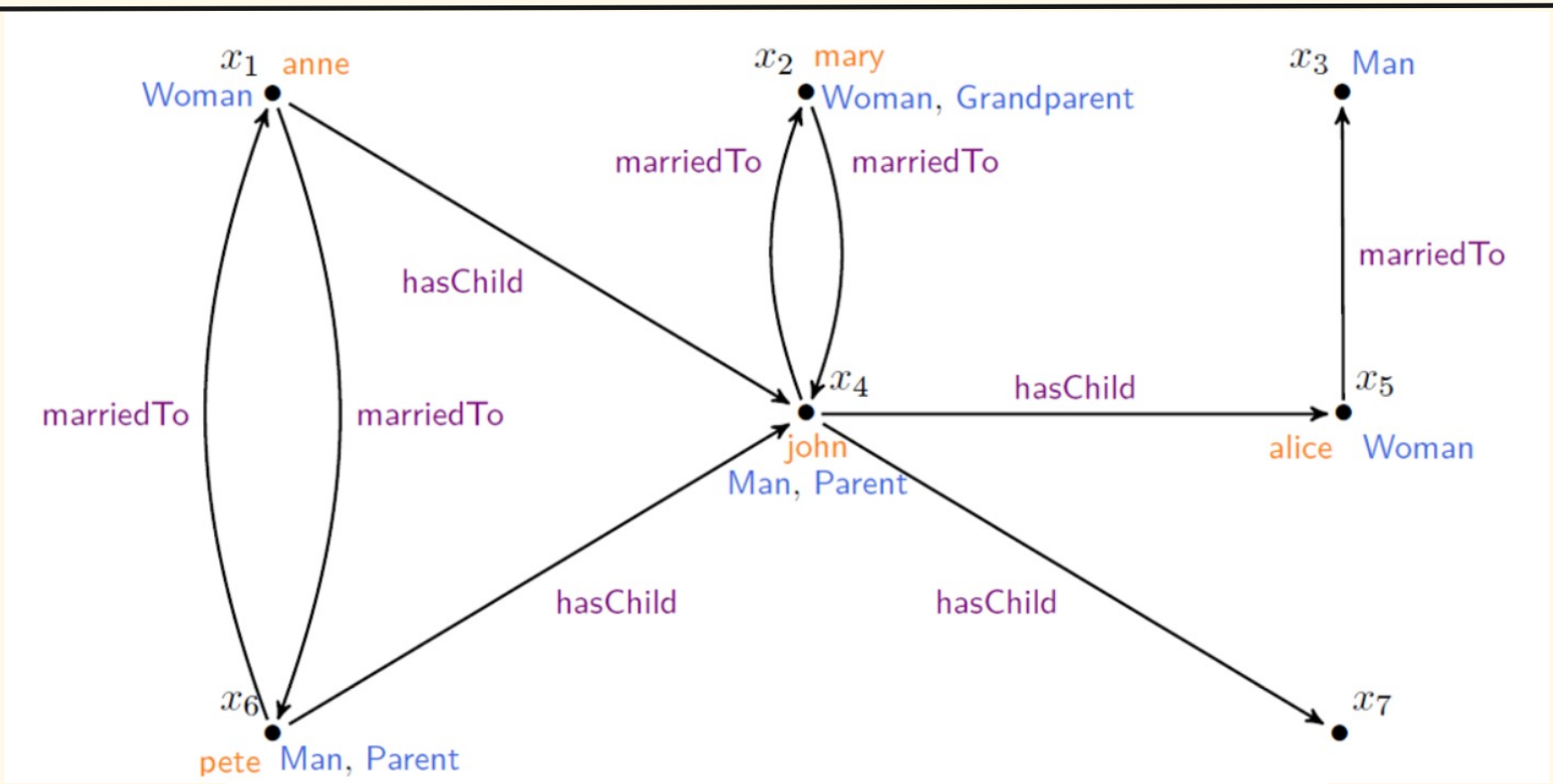


$\exists hasChild.Woman$ - **The individuals that have ***some*** "has child" relation with an element of *Woman***

(= have at least one female child)

{x4}

Mock exam – Question 1



$$\exists hasChild.\top \sqcap \neg \exists hasChild.Woman$$

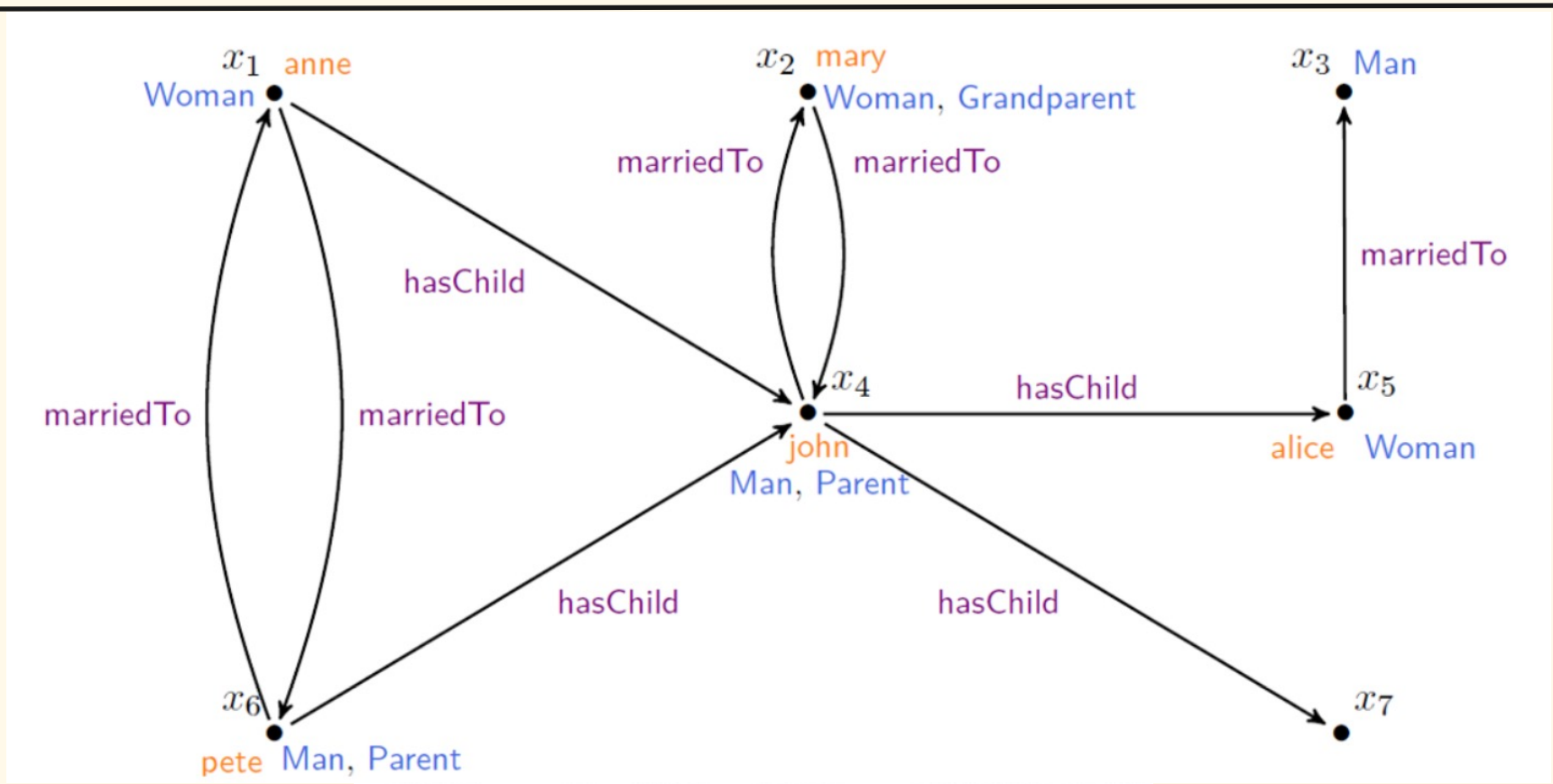
(those who have a child and do not have a female child)

The individuals from {x1,x4,x6} that are not from {x4}

Thus, {x1,x6}

Mock exam

$\exists hasChild.\top \sqcap \neg\exists hasChild.Woman$

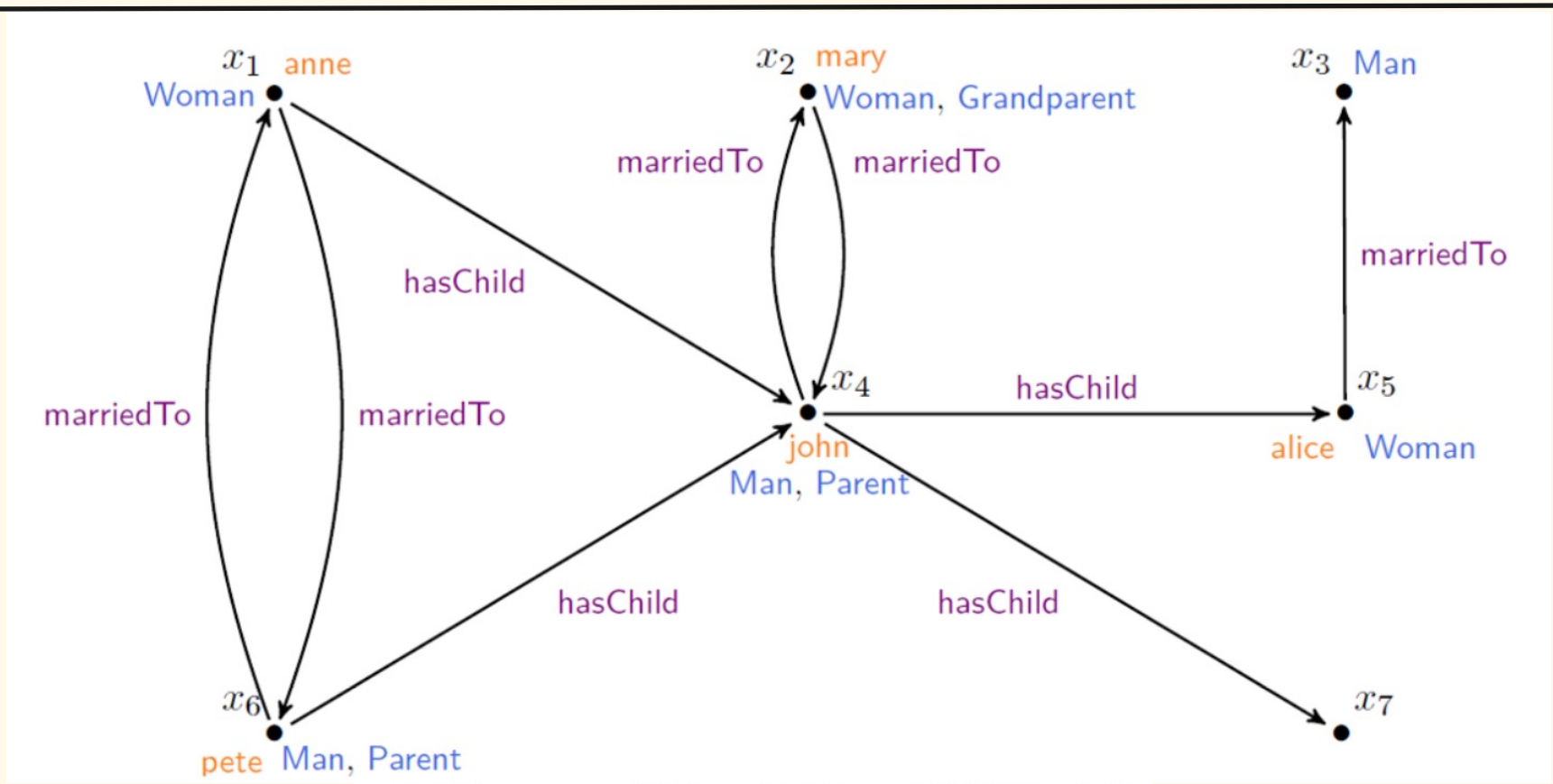


$\exists MarriedTo.\exists HasChild.T$

List all the elements of the set obtained by applying the interpretation function from the interpretation depicted above to the given concept.

Mock exam

$\exists hasChild.\top \sqcap \neg\exists hasChild.Woman$



$\exists MarriedTo.\exists HasChild.T$

Those that are married to someone from "has a child" class

= Those that are married to someone from {x1,x6,x4}

={x1,x6,x2}

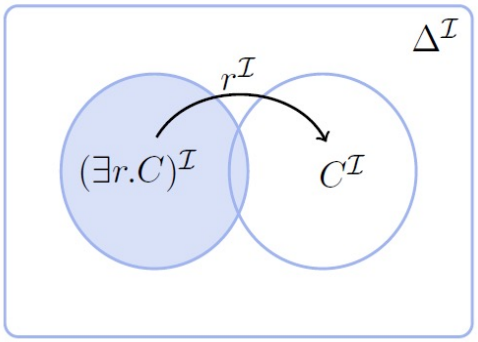
Mock exam – Question 1

Tips:

- **Check the slides L12 =>**
- **Check the slides for Live session 1 of Dragan**
- **Do the exercises 10&11 from the exercises document on reasoning (on BB)**

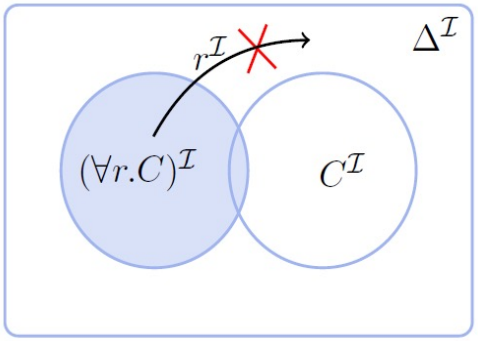
Existential restriction

- At least one value of a class
- $(\exists r.C)^{\mathcal{I}} = \{x \mid r^{\mathcal{I}}(x) \cap C^{\mathcal{I}} \neq \emptyset\}$

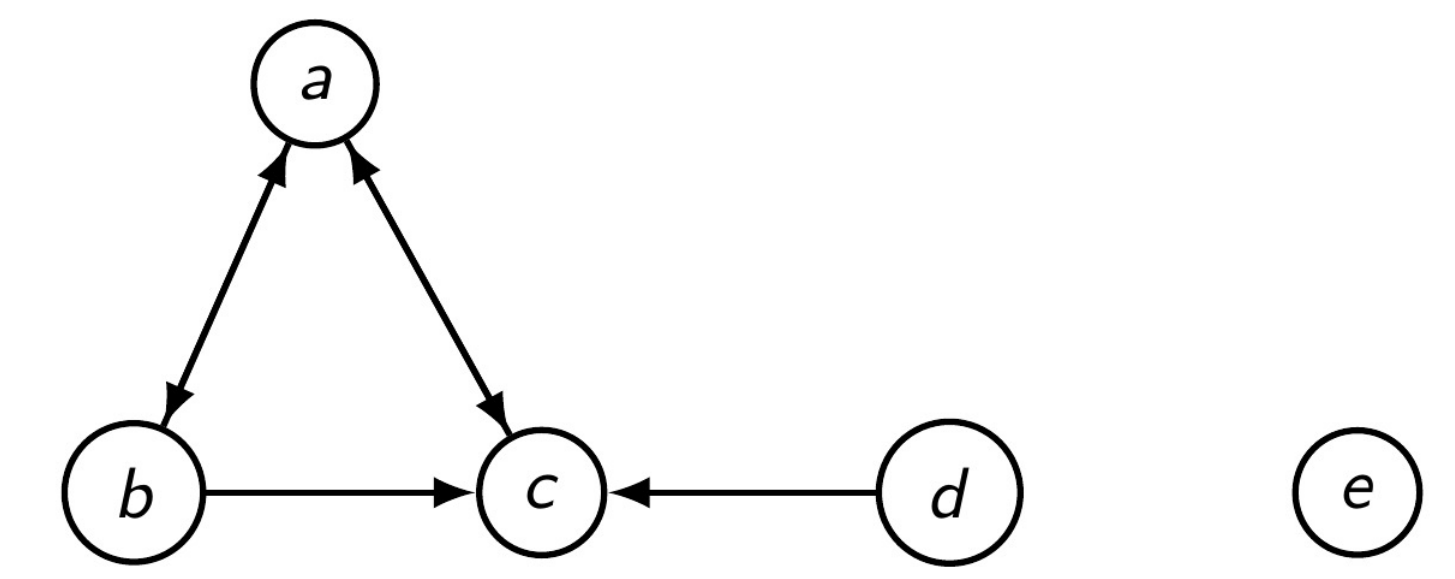


Value restriction

- All values of a class
- $(\forall r.C)^{\mathcal{I}} = \{x \mid r^{\mathcal{I}}(x) \subseteq C^{\mathcal{I}}\}$



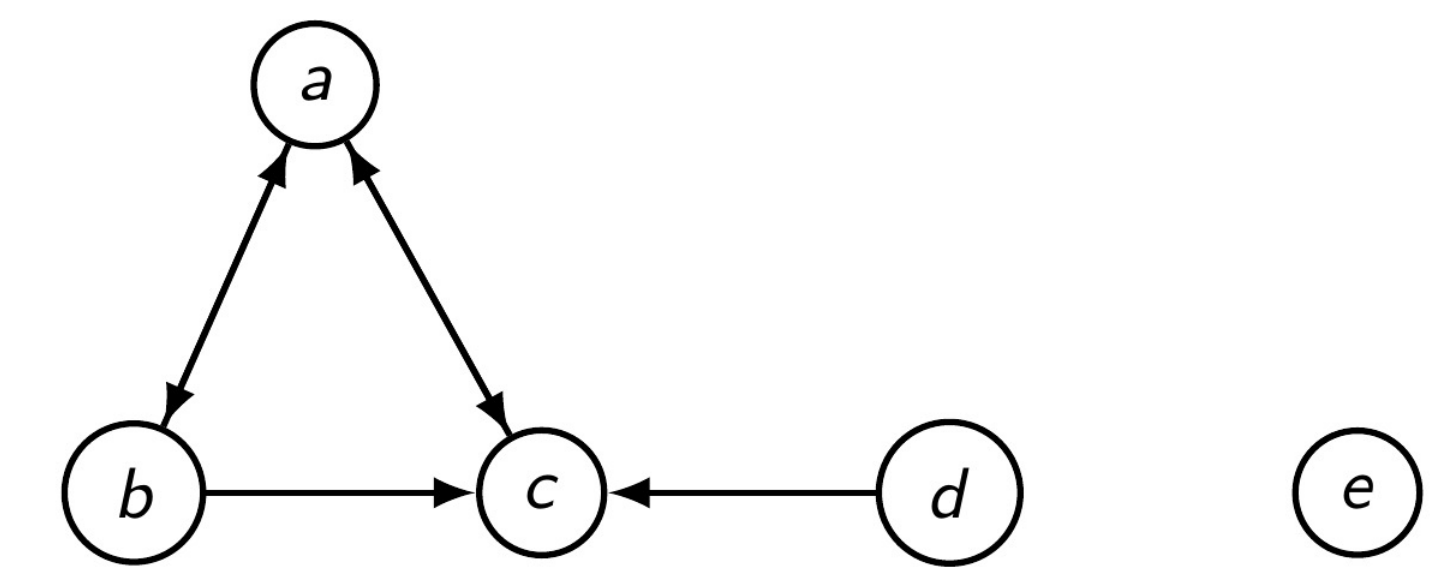
Mock exam – Question 3



Consider the argumentation graph depicted above. Which of the following statements about this graph hold? (Select all the statements that hold.)

a. **The grounded extension contains at most one argument**
b. **There exists a preferred extension which is not stable**
c. **Every argument from the graph belongs to at least one complete extension**
d. **Each preferred extension contains at least three arguments**
e. **No complete extension contains c**

Mock exam – Question 3



Let's calculate the extensions

(complete = conflict-free + admissible + contains every argument it defends)

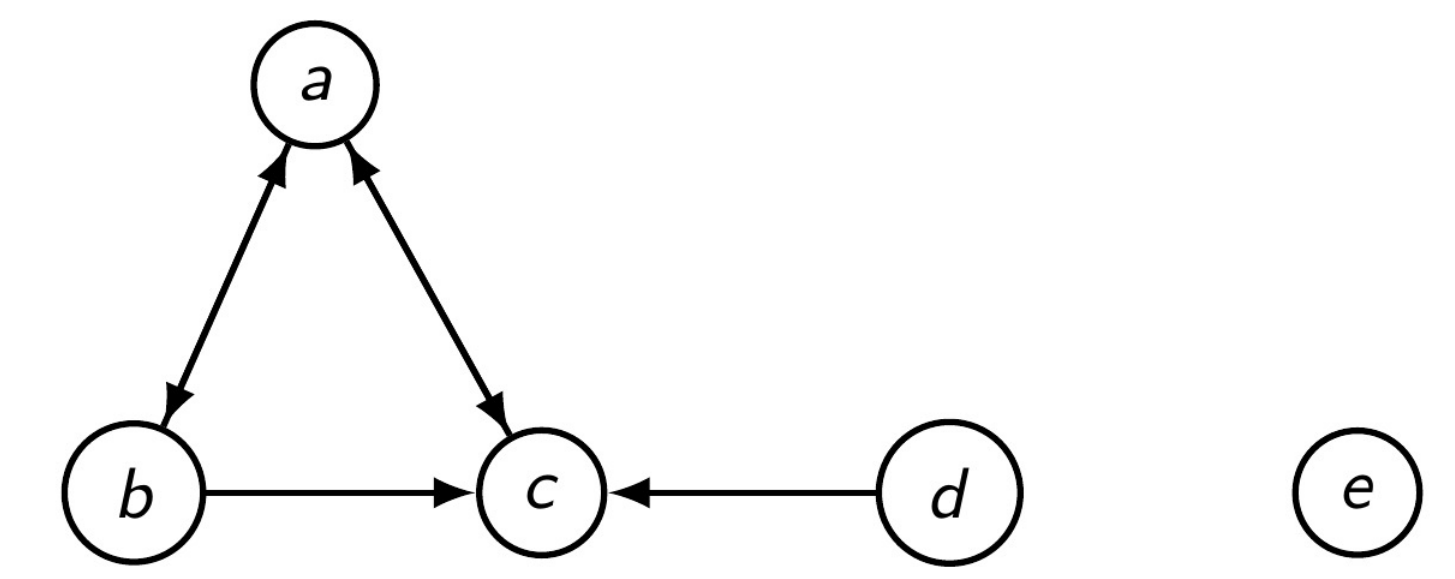
- **Starting point: non-attacked arguments: d, e**
- **Non-attacked arguments belong to <u>every</u> extension**

  Reason: they are (formally) defended by every set S

  (no attackers – no attackers that are not attacked by S)

  (one consequence: the grounded extension is the empty set iff every argument is attacked)

Mock exam – Question 3



- **Grounded**: {d,e}

  (tip: check the iterative procedure for grounded extension – Lecture 14)

- **Preferred**: d and e are inside, so c is not; we can choose between a and b

  {a,d,e}, {b,d,e}

- Both preferred ext. are stable (they attack all other arguments)

Mock exam – Question 4

Assume M is a model of the formula

$$\exists x \exists y (Friend(George, x) \wedge Friend(George, y) \wedge \neg(x = y))$$

Which of the following statements necessarily hold true?

The elements of M are humans

M has at least two elements ✓

M has at most one element

There are two different constants of M that correspond to the constant symbol George

George is not a friend of himself

$$\exists x Friend(x, George)$$

There exist constant symbols that denote the friends of George

$$\exists x \exists y \, (Friend(George, x) \wedge Friend(George, y) \wedge (x = y)$$ ✓

Your questions

The first question had a statement we have not seen before. Would like to know more about that

- **Q1 does not deal with statement. If Q1 is still unclear, call on 1to1**

When can complete, preferred, stable and grounded extensions be the empty set?

- **Empty set is the grounded (and therefore a complete) extension iff every argument in a graph is attacked (see slides Live session 2)**

Are there more excercises to practice with?

- **Exercises L11&L12 on BB (also the book Knowledge representation and reasoning. L14 – exercises on the slides (see also the link from the slides for Live session 2)**

I would like you to explain the concept of model a bit more in detail.

- **1 to 1 call, https://www.youtube.com/watch?v=a4HdbgEgnYE**

The literature that is on the Blackboard pages for L12 and L14 are different from what is stated in the Course Manual.

- **BB literature is subset of CM, sufficient for the lectures. All links are provided.**

Your questions

When an empty set is part of the complete extension shouldn't it also be part of the preferred and therefore also the stable? That is because the empty set is not a subset of anything so it should be maximal. Moreover, in situations where we have an empty set it is implied that the empty set attacks every argument. Therefore, it should also be in the stable extension as it is guaranteed to attack all non complete extension member arguments. I don't understand why this is not the case in practice so any explanation would be great.

- 1 to 1 for more details

I struggle to understand slide 37 of lecture 11 on satisfiability relation with entailment. Could you walk us through it one more time?

- Next page

Your questions

Some semantic concepts: satisfiability and validity

Satisfiability

A formula α is satisfiable iff there is a model m such that $m \models \alpha$ holds (e.g. $p \wedge q$)

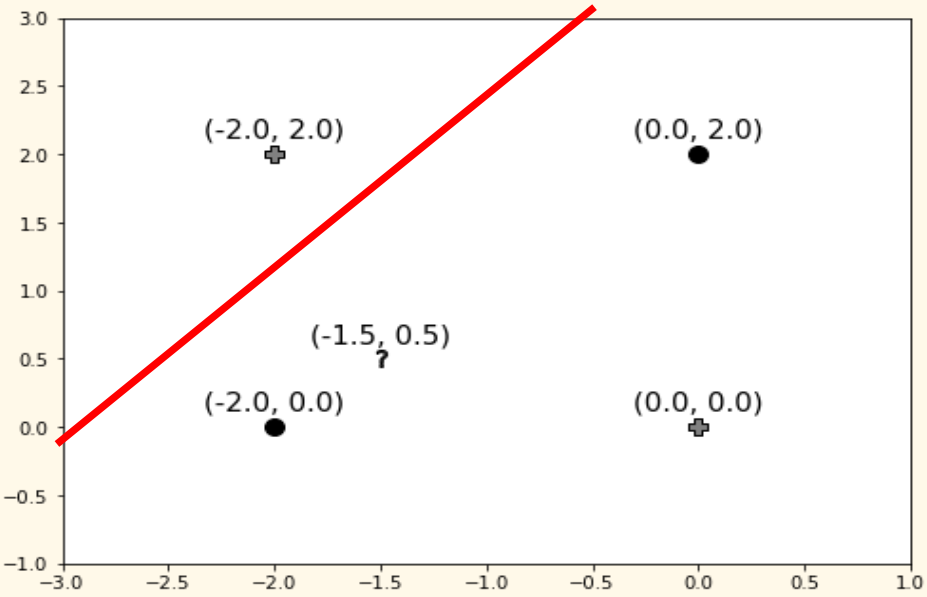
α is satisfiable iff $Mod(\alpha) \neq \emptyset$

Validity

A formula α is valid (notation $\models \alpha$) iff $m \models \alpha$ holds for every model m (e.g. $p \vee \neg p$)

Dong’s part

Exam question



- Decision boundary = straight line
- Training set→ so you can ignore the test point!
- Don't forget to provide the accuracy

The figure below shows you a dataset with four training points (labeled with • and +) and a test point (marked with ?). We have two features: x1 (shown on the x-axis) and x2 (shown on the y-axis). (It's the same figure as in the previous question).

You also fit a logistic regression classifier on this training set with the two features x1 and x2. What is the best **possible accuracy** the classifier can achieve on **this training set?** Describe a **decision boundary** that would achieve this.

Exam question

"Your friend tells you about a decision tree classifier she trained. The splits are chosen based on the misclassification rate. She wanted to limit the maximum depth of the tree and therefore tried the following values for the maximum depth: 5, 10 and 15. Her final tree uses a maximum depth of 15, because that resulted in the highest accuracy on the training set. (1) Please explain why this is not a good approach to set the maximum depth; and 2) What she should have done instead."

Your questions to me

- Logistic regression is still a bit unclear to me. Could you go quickly over this again with some examples?
- Should we be able to do gradient descent by heart? More broadly, doing a logistic regression on a dataset provided by you.
- What is the best moment in developing a machine learning model to tune a hyperparameter?
- in the case of oracle models, could this be a model that has completely memorised the full data (including testing), so basically a lookup table? That should also give us an estimate of the upper bound scores (e.g., due to noise). Is this a correct way to use an oracle model?
- To what detail should we know each corresponding chapter from Jurafsky and Martin? In particular the first chapter on dialogue systems.
- Is bag of words represented as a neural network or a decision tree?
- When do we use L1 and L2 regularization?

Chris’ part

Your answer to questions

- **Modeling question: most answered correct**

- **Remarks:**
  - **Number your answers**
  - **Answer the question, don't add additional information that is not asked for**
  - **Give an explanation. So not "*Biological band is not relevant*" – tell me WHY**

Your answers to questions

- Jasmine runs a reaction time experiment, with a between-subjects manipulation of whether people drank a cup of coffee before the experiment yes or no. However, she had a confound in her experiment: the coffee drinkers all did the experiment at 9 am, the non-coffee drinkers all did the experiment at 11 am. Which of the following types of validity is being violated due to this confound? If you are in doubt between multiple, select the most severe violation.

- Remarks:
  - Explain your answer
  - Reread Cairns paper if you were confused
  - Most confusion between internal and construct

Your answers to questions

- In class we discussed the article by Amershi and colleagues (“Microsoft paper”) with the guidelines for Human-AI interaction. The article provides 18 guidelines across 4 “top-level categories”.
- For this question, please think of an AI-infused technology that you are familiar with and where one design principle is not applied well. This should be your own example, not one that I gave in the lecture or that is extensively discussed in the paper.
- With your example in mind, please answer these questions. Make sure to number your answers, so I see which answer belongs to what subquestion:
  1. Explain briefly what your AI-infused technology example is. The description can be very brief; it should be at such a level that I (Chris) understand what you are talking about. You do not need to go into the full technical details of the technology.
  2. Write down 1 guideline that is NOT adhered to in this technology (give the guideline only).
  3. Write down what top-level category this guideline belongs to (again, only the name).
  4. Explain briefly why you assess that the technology does not adhere to this design principle. You do not need to go deep into the technical realization, it can be a more conceptual description similar to how we discussed examples in class.

Your answers to questions

Things to check in your answer:

- Is answer numbered.
- When asked to give a name (subparts 2,3): give only name
- Did you not go into too much technical detail about the application (the "nitty gritty"). If that is needed to explain your point, that might be OK. Focus on parts that are relevant to question (and design principle), so I understand it and don't get lost. ➔ *balancing act of clear writing*
- Avoid circular answers ("X is the case, because it is doing X")
- Avoid answers that I can't verify ("X is the case because Donald Trump says so")
- Is this really your own example, or something you just copied of the slides or the paper? I want to see your *own example*.
- In an essay question like this, most points would go to your explanation :-)

Your answers to questions

- **Some things I noticed:**
  - ~1/3 of students did not answer
  - Most descriptions were clear. Most students took about 1-3 sentence to explain the tech in general and 2-3 sentences to explain what was wrong and how this ties to technology.

Your questions to me

- The questions you ask are often quite open questions, could you maybe go through one such question so it is clear what constitutes a good explanation?
- There is a lot of material, so it's kind of hard to estimate what the most important bits are.
- Will it be time-consuming in this part since we have to write a lot about each question?

Your questions to me

- Can we have a few more examples of Newell's bands?
- I would like to have the 4 validity types explained again with examples if possible. Same goes for Newell's band and Marr's level question.

Newell and Marr

- Check slides 73-90 of lecture
- Check slides 9 & 10 of live lecture

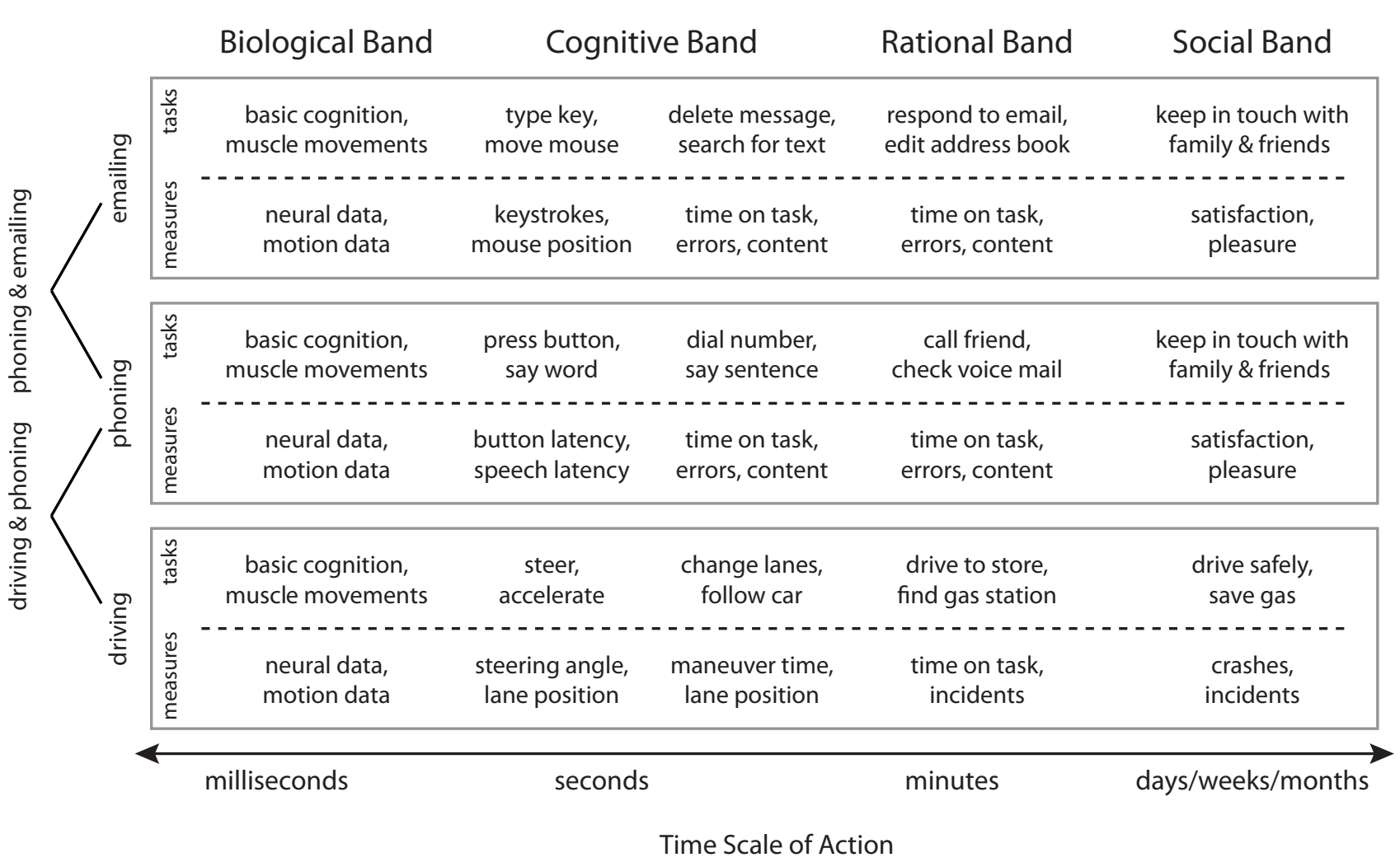


Figure 1.3: The Abstraction Continuum (derived from Newell, 1990).

(From Salvucci & Taatgen (2011) The multitasking mind

What to build? For humans

Levels of Abstraction - David Marr:

- Computational: Why?
- Algorithmic: What?
- Implementation: How?

More detailed about "inside" of brain

More abstracted away from the brain

David Marr (1982) Vision: A computational investigation

4 type of validity (Cairns, 2016)

1. **Construct**
2. **Internal**
3. **External**
4. **Ecological**

More about details of experiment;
Relatively more objective assessment

More about relevance to real-world;
Relatively more subjective assessment

Note: during the exam

- **Chris cannot be online due to family circumstances**
  - **If you run into any questions about my material, please contact Dragan Doder via MS Teams. He knows how to reach me**

NOW WHAT?

Good luck with the exam and with your Master studies!

The MAIR team

P.S.: Please fill out the course evaluation on Caracal, so we can continue to improve this course

(will be sent to you end of next week)