Methods in AI Research
Midway check-up

Dragan Doder, Chris Janssen, Rosalie Iemhoff, Dong Nguyen

Today's structure

1. Exam set-up and structure
2. Your questions
   A. Course organization and procedures
   B. Integration of topics, relationships of themes
3. Group assignment: integration of themes
4. Reflection on integration
5. Your questions
   C. Dong's classes
   D. Chris' classes
6. Remainder: available for 1-on-1 questions (we will stay on this call longer)

1A Exam set-up

- Online exam on **Remindo**:
  - 5 November 09:00-11:30 (09:00–11:55 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).

1B Exam structure

- Types of **questions** in the exam:
  - *Multiple choice questions.* For example:
    - choose the correct answer;
    - which of the following sentences are true? (maybe more than one choice to select per question);
  - *Essay style questions;*
  - *Fill in the blanks type of question*. For example:
    - Word: Insert a missing word/phrase;
    - Calculate: Insert the result (first calculate it “offline” – e.g. use pen&paper).
- Distribution of **points:**
  - Equally distributed per lecturer (Dong, Dragan, Chris)
  - Not equally distributed per lecture;
  - We will cover all the lectures, apart from the bonus lectures; Not each lecture equally extensive
  - Different questions can have different weights.

2A Your questions about organization

Examination

- Is the online going to be online? **yes**
- Are there going to be any old exams we can make to practice for the real exam?
- I think the lessons week by week were very different, which I like, […] but I guess my question is how significant is each part? Because I'm curious if the test will also have a similar structure and how it integrates? Would it be: A few questions from 1.st weeks topic, then 2nd weeks topic, etc equally spread, or are certain things that we should focus less or more on (besides the scientific writing, that will not be examined but it's still good to know)?
- I would like to know how in depth we need to learn the literature. Do we need to know every aspect of the article or just know what the article is about in global?
- What is the best way to study (prevalence of lectures or research papers?)

2A Your questions about organization

Labs

- Well, more of a suggestion: It would have been good to know right from the beginning, that we need to log our time spent on the project.
- I would like to mention that working in a group of 4 is not very practical in a lecture hall at 1,5m distance. It would be much better to work together in a seminar room (the 1,5m distance is less bad then because you can then hear each other etc). I suppose that this has not been your personal choice either, but rather an effect of the current situation, however, a side effect is that more and more groups tend to meet somewhere else.

2B Your questions about integration

- No! It is clear they all have a relationship to AI
- As the lecture provided a good intro to project work with ML pipelines, and I wonder if there are any courses providing deeper insights into ML engineering and project management.

Today's structure

1. Exam set-up and structure
2. Your questions
   A. Course organization and procedures
   B. Integration of topics, relationships of themes
3. **Group assignment: integration of themes**
4. **Reflection on integration**
5. **Your questions**
   C. **Dong's classes**
   D. **Chris' classes**
6. **Remainder: available for 1-on-1 questions (we will stay on this call longer)**

3. Group assignment: Integration of themes

- **Domain:**
  Imagine an "**intelligent calendar**". So, a calendar (e.g. iCal, google calendar, outlook calendar). What you consider intelligent can vary, but perhaps you can think of aspects such as:
  - making intelligent inferences
  - integrating information from a (perhaps different) source(s)
  - integrating information in a smart way

  You can take inspiration from a digital calendar that you interacted with, or perhaps you can envision a future, improved calendar.

- Discuss: How do methods apply to this domain?
  Think of methods from past and future classes; labs.
  Try to think of methods that are rooted in different disciplines
- Be prepared to discuss with us
- Report back in XXX minutes.

Methods

Lectures so far	Lectures to come	Labs
Dialogue systems	Knowledge-based systems	Domain modeling
Data and vector processing	Description logic	Machine learning
Natural language	Non-monotonic reasoning	Natural language processing
Machine learning	Formal argumentation	Implement interactive system
Cognitive modeling	Logical reasoning	Scientific writing
Experimentation (design and analysis)		Experimental design
Scientific writing		Statistics (various forms)
Designing responsible AI systems (Chris: interaction)		
Designing responsible AI systems (Dong: societal impact)		

Today's structure

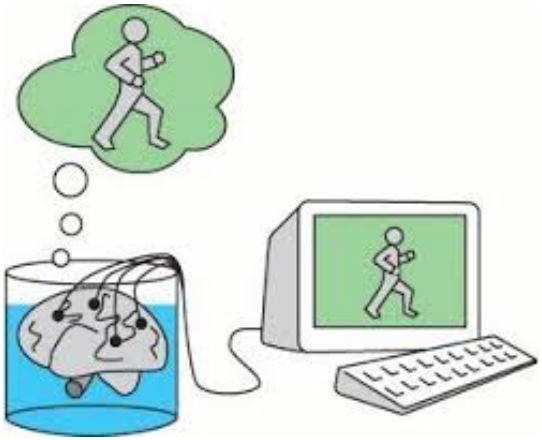
1. Exam set-up and structure
2. Your questions
   A. Course organization and procedures
   B. Integration of topics, relationships of themes
3. Group assignment: integration of themes
4. **Reflection on integration**
5. **Your questions**
   C. **Dong's classes**
   D. **Chris' classes**
6. **Remainder: available for 1-on-1 questions (we will stay on this call longer)**

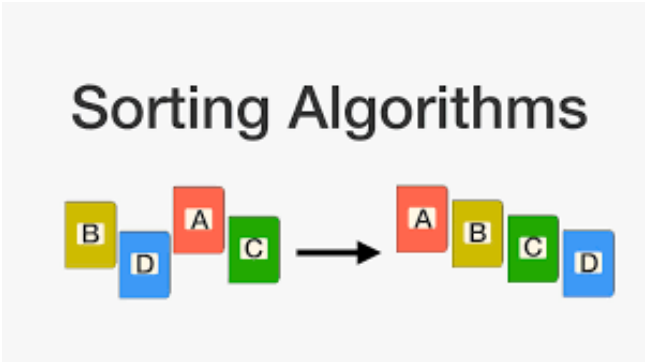
4 Reflection on Integration (so far)

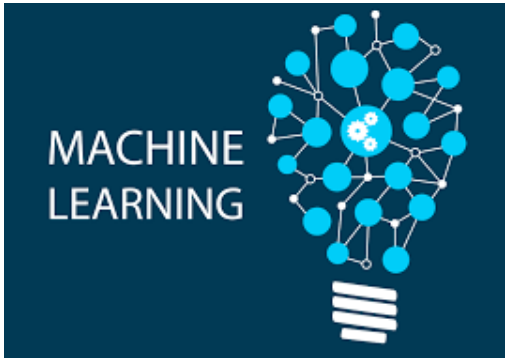
Tied to

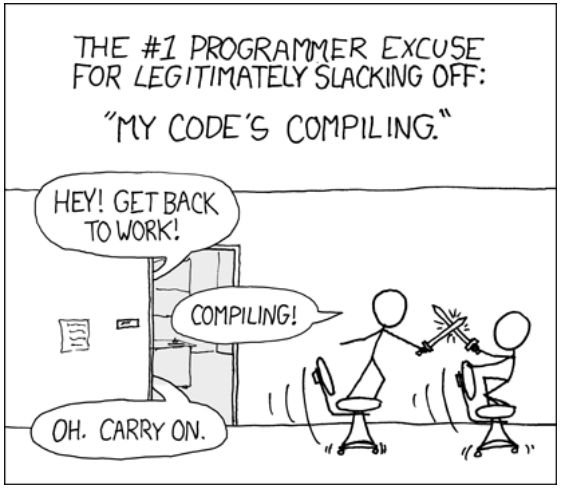
A. Previous lectures:
   - Rosalie's Introduction
   - Dong's class on responsible AI
B. Learning objectives
C. Your future career

4. Reflection: tied to introduction









4. Reflection: tied to introduction

Research in Artificial Intelligence takes place in various disciplines:

- computer science
- psychology
- linguistics
- mathematics
- logic
- neurology
- philosophy
- humanities

Flavors:

- machine driven versus data driven
- theory versus engineering

⋮

These fields have different research methods:

- algorithms
- statistics
- natural language processing
- mathematical proofs
- logical methods
- engineering

⋮

Ok… what now?

Maybe we shouldn't use AI?

Ok… what now?

Maybe we shouldn't use AI?

But: what is the alternative/current situation?

People:

- make mistakes
- are biased
- can't always explain their decisions
- are not consistent

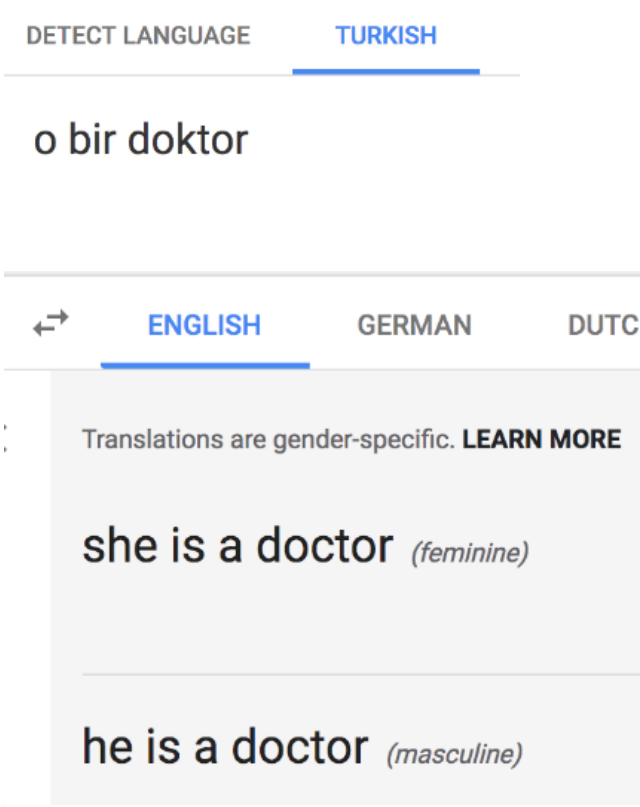
Dong's class (designing responsible AI)

Multidisciplinary solutions needed

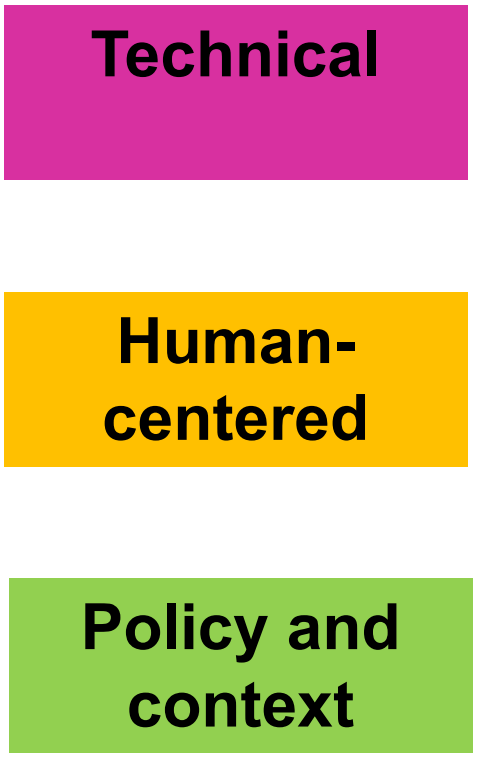
Technical: Enhancing training data, new metrics, interpretability methods, etc.

Human-centered solutions: Google translate changed it user interface. User experiments.

Policy solutions: E.g., the "right to explanation" (but regulations are heavily debated)



Multidisciplinarity



Multidisciplinarity

Lectures
Dialogue systems
Data and vector processing
Natural language
Machine learning
Cognitive modeling
Experimentation
Scientific writing
Designing responsible AI
Knowledge-based systems
Description logic
Non-monotonic reasoning
Formal argumentation
Logical reasoning

Technical

Human-centered

Policy and context

Labs
Domain modeling
Machine learning
Natural language processing
Interactive system
Scientific writing
Experimental design
Statistics

Multidisciplinarity

Lectures
Dialogue systems
Data and vector processing
Natural language
Machine learning
Cognitive modeling
Experimentation
Scientific writing
Designing responsible AI
Knowledge-based systems
Description logic
Non-monotonic reasoning
Formal argumentation
Logical reasoning

Technical

Human-centered

Policy and context

Labs
Domain modeling
Machine learning
Natural language processing
Interactive system
Scientific writing
Experimental design
Statistics

Multidisciplinarity

Lectures
Dialogue systems
Data and vector processing
Natural language
Machine learning
Cognitive modeling
Experimentation
Scientific writing
Designing responsible AI
Knowledge-based systems
Description logic
Non-monotonic reasoning
Formal argumentation
Logical reasoning

Technical

Human-centered

Policy and context

Labs
Domain modeling
Machine learning
Natural language processing
Interactive system
Scientific writing
Experimental design
Statistics

Multidisciplinarity

Lectures
Dialogue systems
Data and vector processing
Natural language
Machine learning
Cognitive modeling
Experimentation
Scientific writing
Designing responsible AI
Knowledge-based systems
Description logic
Non-monotonic reasoning
Formal argumentation
Logical reasoning

Technical

Human-centered

Policy and context

Labs
Domain modeling
Machine learning
Natural language processing
Interactive system
Scientific writing
Experimental design
Statistics

Dong's class (designing responsible AI)

Multidisciplinary solutions needed

- You have been taught multiple methods

- You have been *integrating* ideas from fields (e.g., in labs, discussions)
  - ***Interdisciplinairy*** learning

4B. Integration: tied to learning objectives

At the end of the course the students will:

- Know and understand important techniques in the different fields of AI, such as [..]
- Be able to choose from and use different research methods in AI. More specifically, they will be able to:

  (a) Implement different AI techniques in a [..] programme;

  (b) Test and evaluate an AI system (technical capabilities, performance, usability);

  (c) Write a technical report and a research paper on an AI system, its evaluation and its place in the broader context of AI.

- Note that the students are expected to achieve all these goals at a basic level, to give them a first “taste”. They will know where to find more information, and be in a good position to take further classes in which the above skills are deepened.

4B. Integration: tied to learning objectives

At the end of the course the students will:

- **Know and understand important techniques** in the different fields of AI, such as [..]
- Be able to **choose from and use different research methods in AI**. More specifically, they will be able to:

  (a) **Implement different AI techniques** in a [..] programme;

  (b) **Test and evaluate an AI system** (**technical** capabilities, performance, **usability**);

  (c) **Write a technical report and a research paper** on an AI system, its evaluation and its place in the broader context of AI.

- Note that the students are expected to achieve all these goals at a basic level, to give them a first “taste”. They will know where to find more information, and be in a good position to take further classes in which the above skills are deepened.

4B. Integration: tied to learning objectives

At the end of the course the students will:

- **Know and understand important techniques** in the different fields of AI, such as [..]
- Be able to **choose from and use different research methods in AI**. More specifically, they will be able to:

  (a) **Implement different AI techniques** in a [..] programme;

  (b) **Test and evaluate an AI system** (**technical** capabilities, performance, **usability**);

  (c) **Write a technical report and a research paper** on an AI system, its evaluation and its place in the broader context of AI.

- Note that the students are expected to achieve all these goals at a basic level, to give them a first “taste”. They will know where to find more information, and be in a good position to take further classes in which the above skills are deepened.

4C Integration: tied to profession

4C Integration: tied to profession

So…

Keep up the good work!
Keep learning new things.

Today's structure

1. Exam set-up and structure
2. Your questions
   A. Course organization and procedures
   B. Integration of topics, relationships of themes
3. Group assignment: integration of themes
4. Reflection on integration
5. **Your questions**
   C. **Dong's classes**
   D. **Chris' classes**
6. **Remainder: available for 1-on-1 questions (we will stay on this call longer)**

5B Your questions about Dong's lectures

What do we need to know?

- How specific should our knowledge on the machine learning be, when it comes to the exam?
- Explain a little more about backprop!
  - Take a ML course! ☺ See also YouTube (e.g. CS231n Winter 2016: Lecture 4: Backpropagation, Neural Networks 1)
- In 'Limits of Learning' they talk about Bayes optimal classifier, which is not mentioned in the lecture. Should we know the details of the literature that is not mentioned in the lectures?
- To be clear, all techniques and concepts we learned, while constructing the dialog system are subject to be reproduced in the exam?

5B Your questions about Dong's lectures

Responsible ML: There seem to many people that say things like: AI can fix this or will improve this. How did this idea originate that AI can fix a lot? Moreover, of course AI should be designed with ethical considerations, but what about the people using it? Even though the AI system is super fair, people can still use it for malicious things, right?

5B Your questions about Dong's lectures

Can you maybe explain one more time the cosine similarity, the Euclidian distance and the relationship between these two?

$$\textbf{cosine similarity}$$

$$\frac{\boldsymbol{a} \bullet \boldsymbol{b}}{\|\boldsymbol{a}\|\|\boldsymbol{b}\|} = \frac{\sum \boldsymbol{a_i}\, \boldsymbol{b_i}}{\sqrt{\sum \boldsymbol{a_i^2}}\sqrt{\sum \boldsymbol{b_i^2}}}$$

if a and b have unit length:
cosinesim(a,b) = $\sum \boldsymbol{a_i}\, \boldsymbol{b_i}$

$$\sum a_i{}^2 \;=\; \|\boldsymbol{a}\|_2^2 = 1$$

Euclidian distance

$$\|\boldsymbol{a} - \boldsymbol{b}\|_2 = \sqrt{\sum (a_i - b_i)^2}$$

$$\|\boldsymbol{a} - \boldsymbol{b}\|_2^2 = \sum (a_i - b_i)^2$$

$$= \ldots\ldots$$

$$= 2 - 2\ \text{cosinesim(a,b)}$$

5B Your questions about Chris' lectures

Content clarification

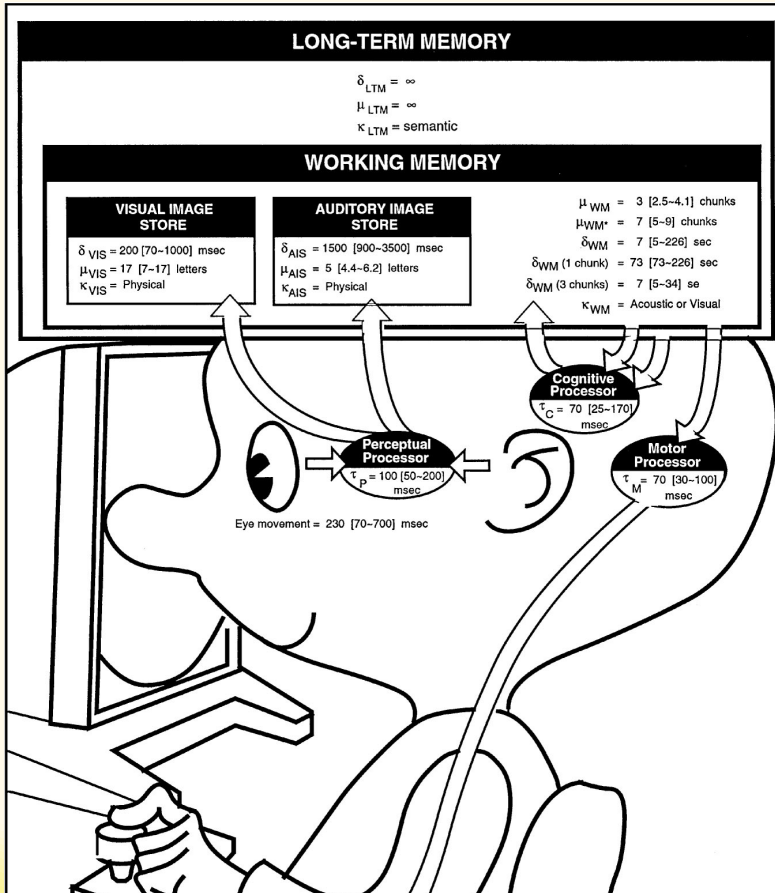
- Should we learn by heart all the guidelines for design ?
- Can you explain one time more clearly: What is the difference between cognitive model and cognitive architecture? I get a cognitive model, but a cognitive architecture is hard to get my head around.

Cognitive Architecture & Cognitive Model

- **Cognitive *Architecture* (John Anderson, 2007):**
  "a specification of the structure of the brain at the level of abstraction that explains how it achieves the function of the mind"

- **Explains:**
  - **What brain can & can't do (function)**
  - **How it does that (structure)**
  - **What *general* parameters and equations govern behavior**

So, in other words:
General framework about general performance



Cognitive Architecture & Cognitive Model

- **Cognitive *Model***
  - **A model of a *specific* task or process**
  - **Developed *within* or *inspired by* an architecture**

So, in other words: It is more specific, about a specific context

Cognitive Architecture & Cognitive Model

- **Architecture questions: *(i.e., is about general functioning of cognition)***
  - **What is our *general* memory capacity?**
  - **How do we forget information over time in *general*?**
  - **How much visual information can we process per time unit? *(i.e.,: general information processing capability)***
  - **How do we control our hands in *general*?**
- **Model questions: *(i.e., typically about a specific domain/issue / question)***
  - **How do we calculate 101 x 7 – 3 (given architecture)**
  - **How do we control a car (given architecture)**
  - **How do we divide time between driving and making a phone call (given architecture)**

What is perhaps confusing

- One model can not test all aspects of a cognitive architecture
- So.. Architecture is developed and refined over time, but testing multiple models (multiple domains)
- One model can test one specific idea / hypothesis

5B Your questions about Chris' lectures

Content clarification

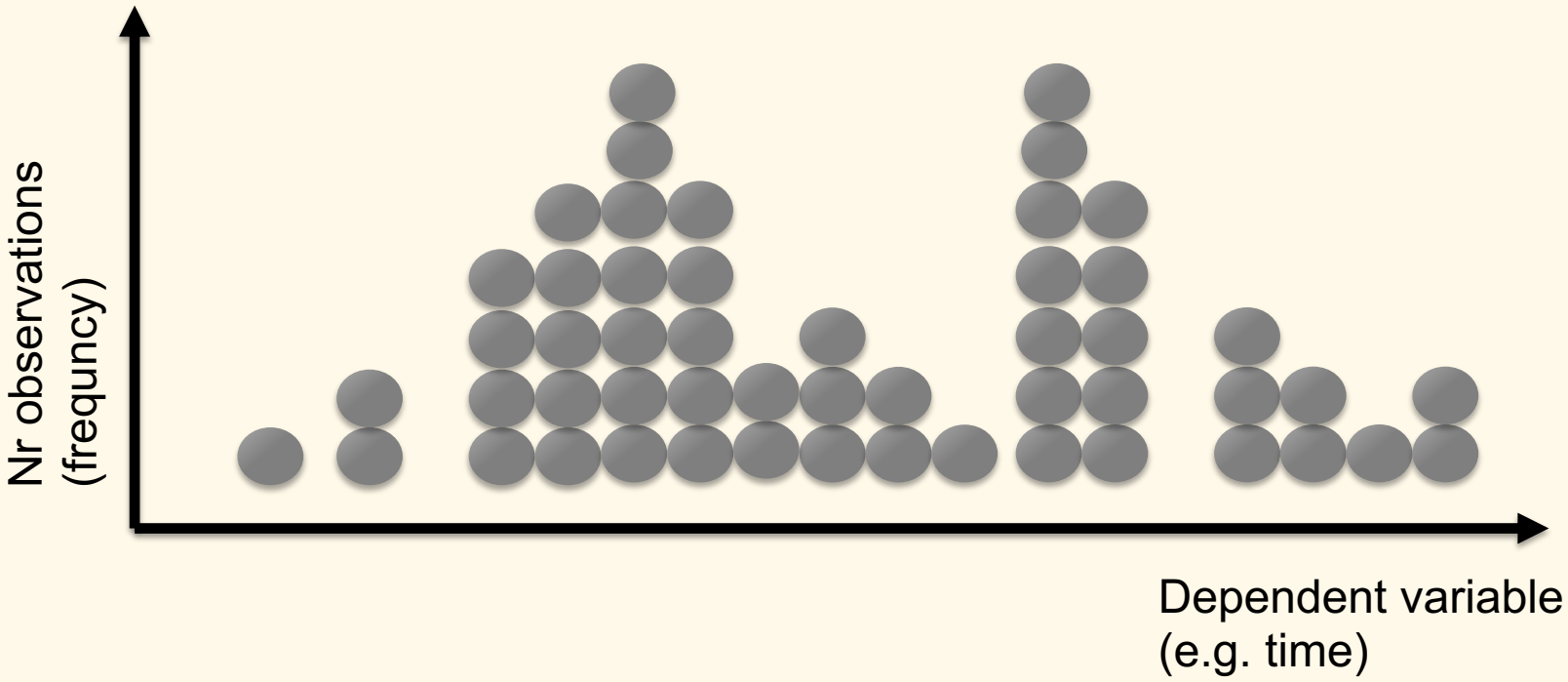
- What is difference between qualitative and quantitative studies?
- To what extend should we know all examples? Because some examples, like the horse and MABAMABA were explicitly mentioned in the quizes

5B Your questions about Chris' lectures

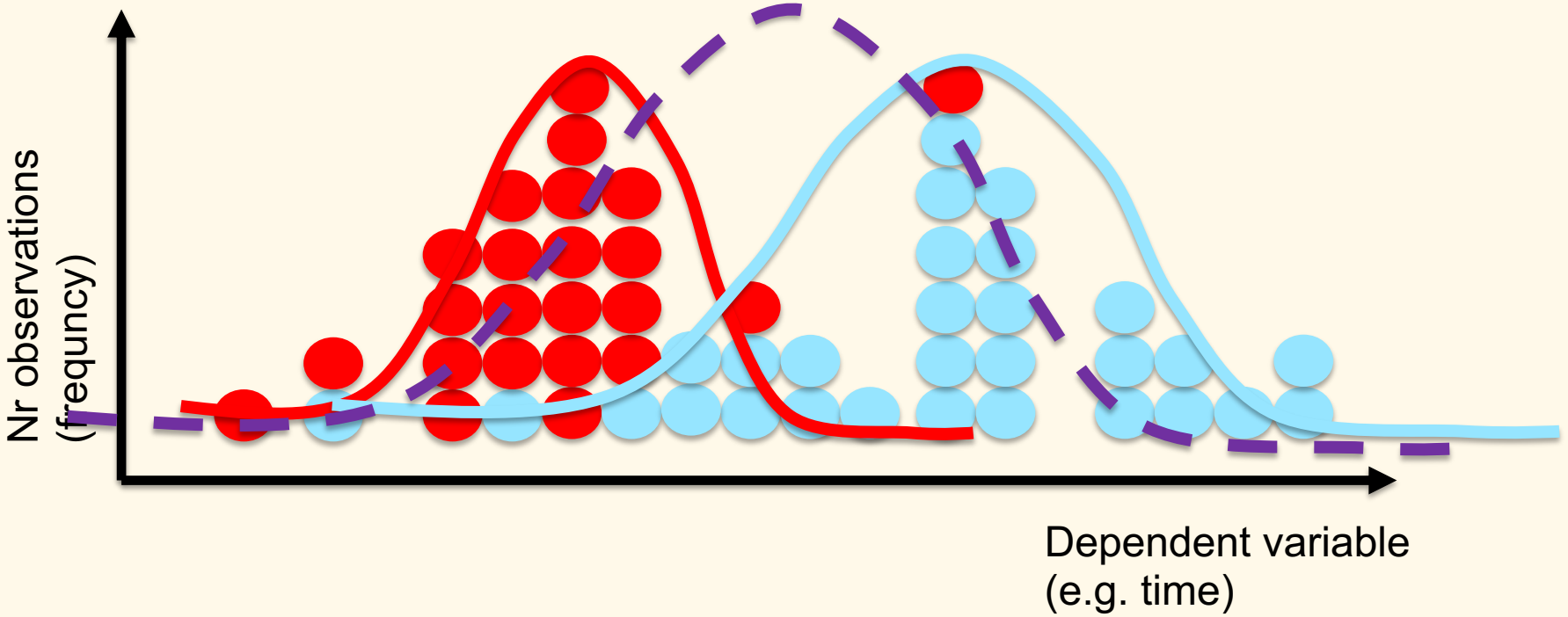
Bigger picture:

- How important is it to write in neutral (without the use of "we" => rather "the team")? Does it depend on the kind of publication?
- In my bachelor's, we have discussed the negative side of P-value statistics and how it could be better to for example use bayesian statistics. What is you opinion about this? Should P-value statistics be replaced by other types of statistics? Would that improve psychology research?

Classic frequentist approach



Classic frequentist approach



Some of the theoretical assumptions

- ***theoretical a priori*** **motivation to guide the intervention. Science (broader: understanding) motivates the *test a priori.***

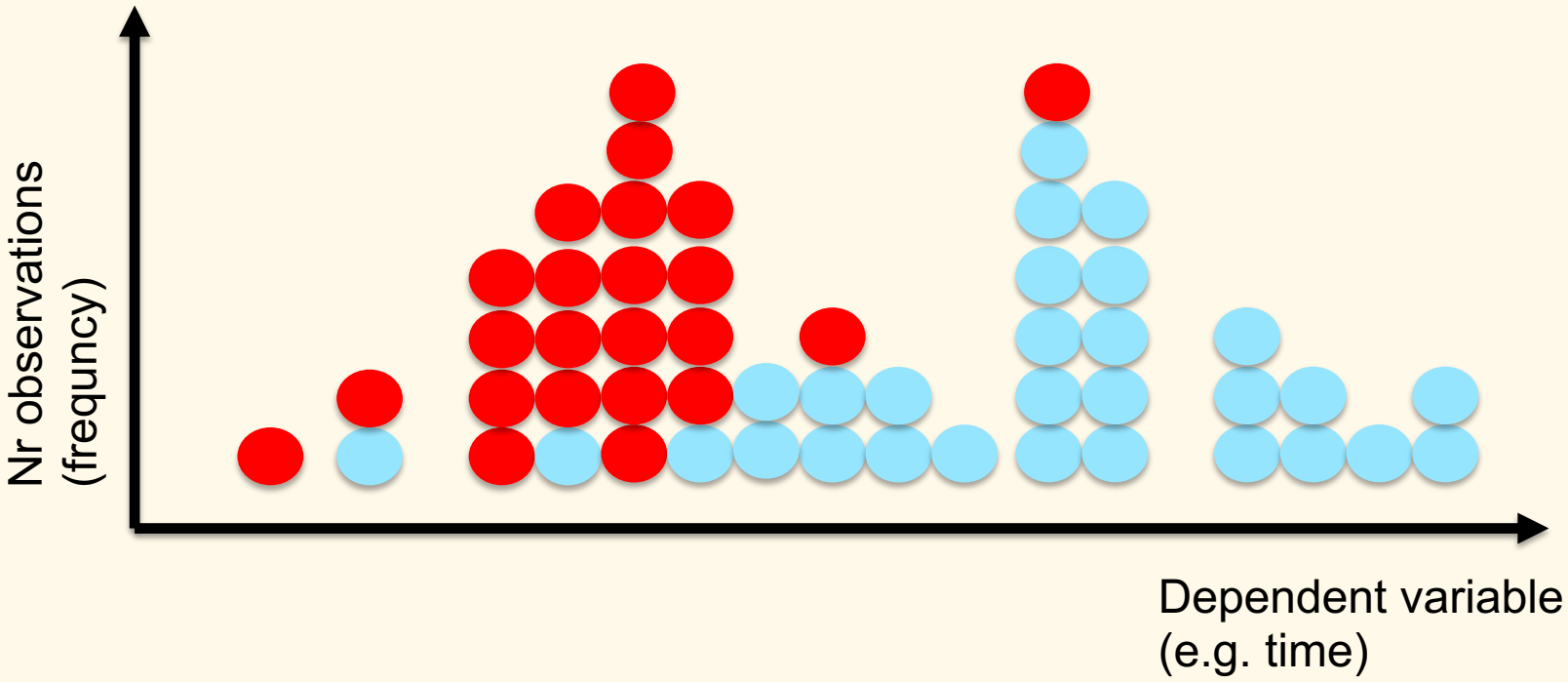
- **A priori understanding of *how many observations* are needed (power analysis). Only do statistical test ("counting frequencies") after all data is in**

Problem: “(p-)hacking”

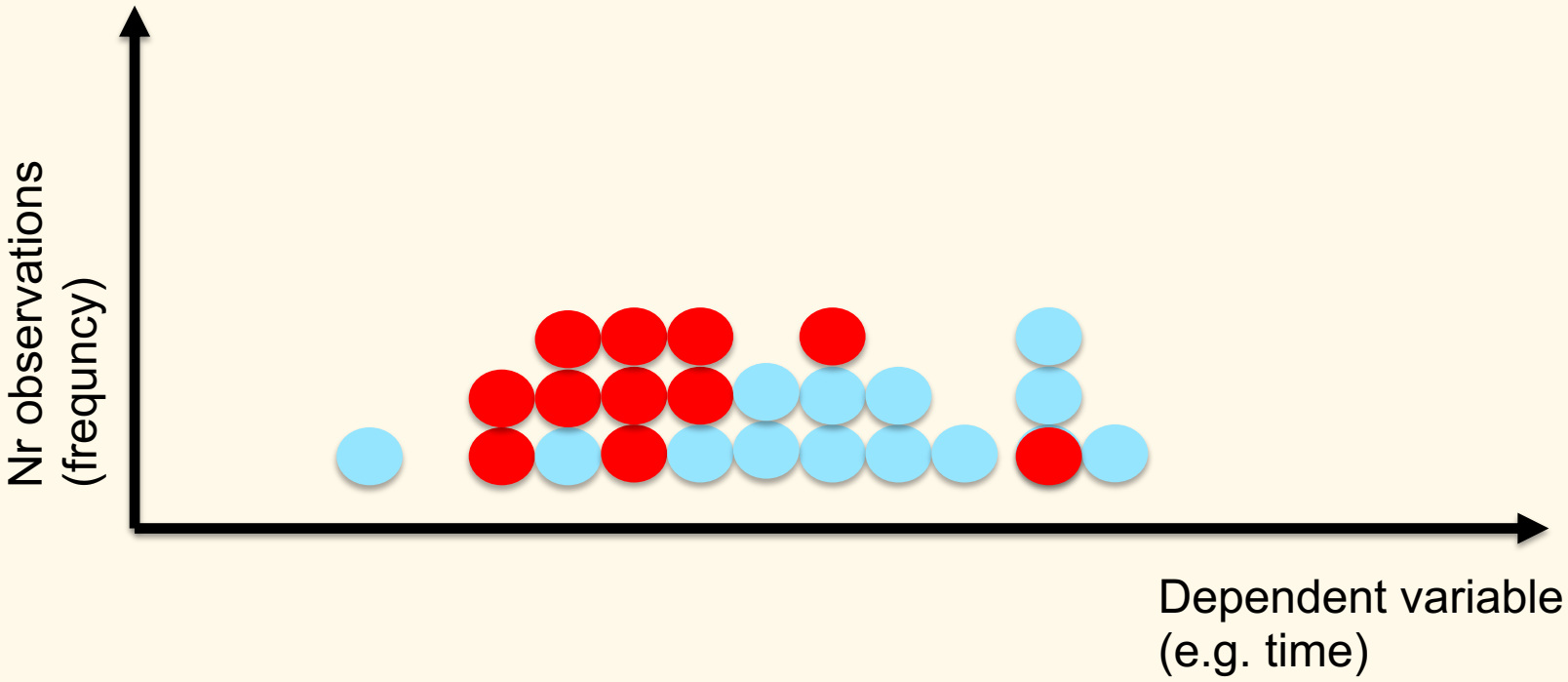
- ***hack of a priori motivation:*** **testing multiple ideas on the data (“fishing”). Only reporting things that have a difference, then making up the theory afterwards (“harking”)**

- ***P-hack of number of observations:*** **continue data collection untill you find a significant difference**

Classic frequentist approach

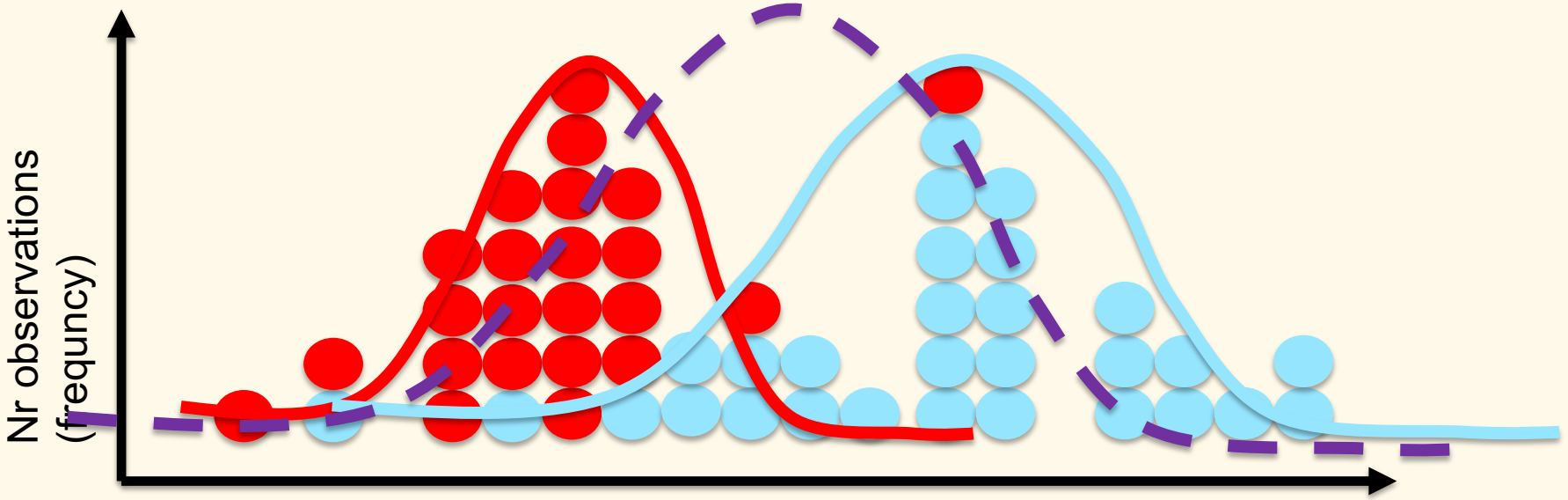


Classic frequentist approach



Solution: Bayes?

Bayes' approach



Which situation is more *likely*?
1 distribution (H0) or 2 distributions (H1) or even others…

Express in *likelyhood ratio,* distinguish 3 situations:

1. L(H1) >>> L(H0). → Accept H1 (conclusion possible)
2. L(H1) <<< L (H0) → Accept H0 (conclusion possible)
3. L(H1) ≈ L(H0) → Not sufficient evidence, no conclusion → *collect more data*

Bayes and p-hacking

- ***(p-) hack of a priori motivation:*** **testing multiple ideas on the data ("fishing"). Only reporting things that have a difference, then making up the theory afterwards ("harking") → not solved!!**
- ***P-hack of number of observations:*** **continue data collection untill you find a significant difference**

  -> solved. Collect data until certain
- **New hack: which hypotheses to consider? What "priors"?**

Conclusion…

- **Hackers can hack anything**

- **Bayes can solve some problems, and can be a useful tool**
  - **Also allowed in labs if you want to (not required)**
  - **Tip: explore jasp: https://jasp-stats.org/**

Today's structure

1. Exam set-up and structure
2. Your questions
   A. Course organization and procedures
   B. Integration of topics, relationships of themes
3. Group assignment: integration of themes
4. Reflection on integration
5. Your questions
   C. Dong's classes
   D. Chris' classes
6. Remainder: available for 1-on-1 questions (we will stay on this call longer)