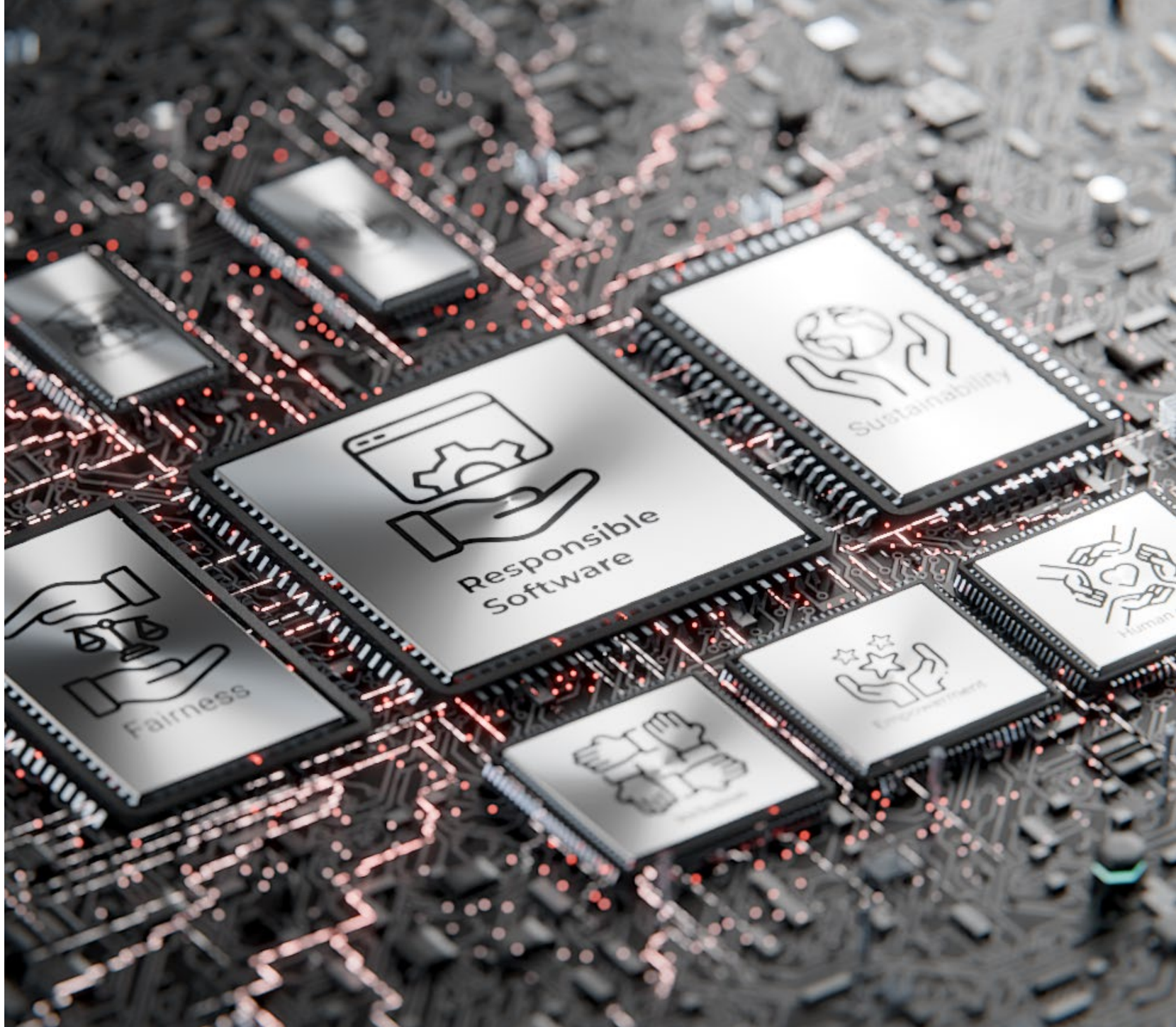
EPFL

Mock Test Debriefing

4 nov.

Cécile Hardebolle

Responsible Software



Agenda for today

1. Debriefing of the mock test

2. Tips for effective revisions

Debriefing of the Mock Test

Format of the exam

- In the Winter exam session
  - Date: **21 January 2026, at 9h15**
    ⚠️ **subject to change until official publication of exam schedule** ⚠️
  - Duration: **2h**
    Note: this is a change compared to last year, duration was 1h30

- Supervised in auditorium, individual
- **Pen and paper (QCM + case studies)**
- Closed book, **one A4 paper sheet** allowed (both sides, free format)

👉 The format of the mock test reflects exactly the format of the exam!

Overall feedback

Thanks a lot to those who sent feedback (26)! 🙏 🙏 🙏

- Level of difficulty of the test:

Response	Average
Easy	8%
Neutral	50%
Difficult	33%

- Comments:
  - **Usefulness of the cheatsheet++**
    - Watch out: changes compared to last year!
  - How is subjectiveness managed during evaluation?
  - Language: add a translation in French

Duration of the mock test

- How long did it take for you to complete the whole test?

Response	Average
[1h ; 1h30 [	32%
[1h30 ; 2h [	24%
2h	12%
> 2h	24%

Votes on questions

Questions with more than 20% vote:

- Single choice questions:
  - 2
  - 3
  - 5
  - 7
  - 8
  - 10
  - 12
- True/false
  - 14
- Case studies
  - 17 (least voted)
  - 18
  - 19

Debriefing of the Mock Test

Single-choice questions

Question 2

URL: ttpoll.eu
Session ID: cs290

A company has developed a complex algorithm to predict whether athletes suspected of doping actually do it. A positive result means that the algorithm classifies athlete as at risk of doping, while a negative result means no risk of doping. The system has been used for 5 years and we have access to data about athletes that were indeed caught for doping. We found that the proportion of athletes predicted to dope amongst all predictions is higher for men rather than for women.
The type of fairness metric we have used is:

19% a. Conditional use accuracy equality

5% b. Error rate balance

5% c. Equal accuracy

✅ 71% d. Demographic parity

Question 3

URL: ttpoll.eu
Session ID: cs290

Imagine that you develop software for people from a single country. If you nonetheless envision cultural differences in this context, which strategy are you probably using?

Question 5

URL: ttpoll.eu
Session ID: cs290

Fill the blanks:
If a piece of software behaves in a ___ way at first glance, but puts people of ____ at ____ , then it is a case of ____ discrimination.

4% a. neutral / several groups / a disadvantage / direct

4% b. positive / specific groups / an advantage / inverse

0% c. negative / several groups / an advantage / direct

0% d. negative / specific groups / a disadvantage / indirect

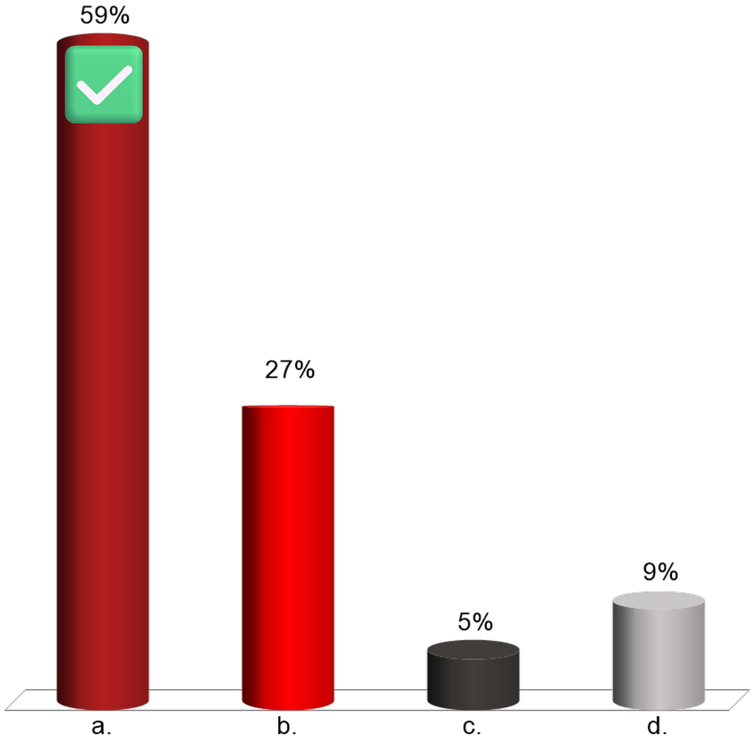
✅ 91% e. neutral / specific groups / a disadvantage / indirect

Question 7

URL: ttpoll.eu
Session ID: cs290

A group of computer scientists with similar background, all experts in software development, are starting a new software project for healthcare. They are aware of cognitive biases and want to minimize the impact of these biases when making design decisions.
Which is the only strategy that could effectively help them in this context?

a. Use a structured approach and slow down the decision-making process
b. Choose one or two of them to play the devil's advocate
c. Systematically include all members of their group to increase heterogeneity
d. Systematically include all members of their group to apply a participatory design method



Question 8

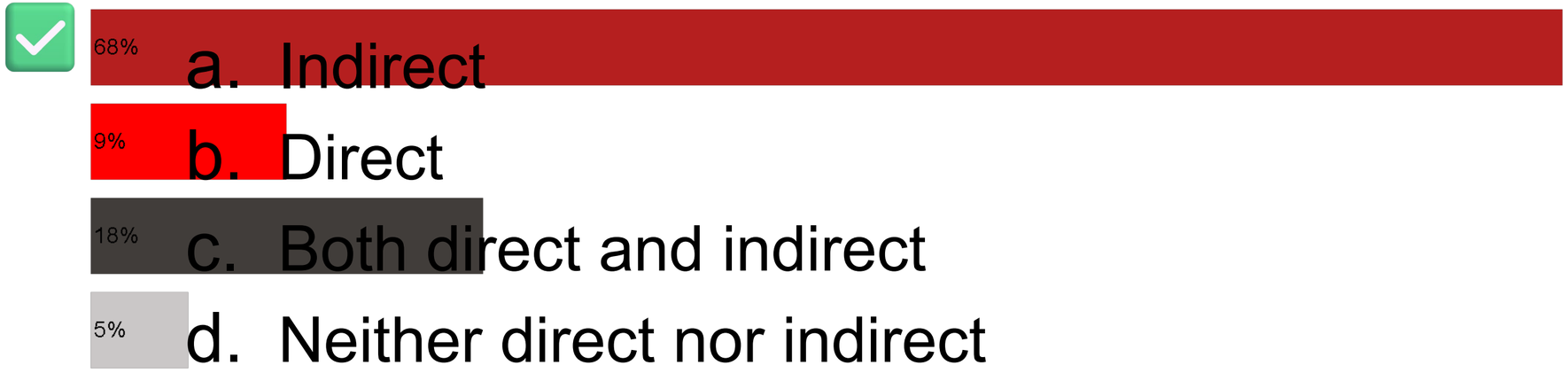
URL: ttpoll.eu
Session ID: cs290

You work on a chatbot to provide students assistance on campus questions. When evaluating the quality of the responses it provides, you identify that the responses contain hallucinations (i.e. content that is incorrect or wrong but looks perfectly plausible) with a 15% rate.
What type of situation are you facing?

Question 10

URL: ttpoll.eu
Session ID: cs290

A bad actor launched a phishing attack on employees of Swiss public institutions to steal their login credentials. An online media outlet reported on it, with the most upvoted comments on the article criticizing the institutions for their inability to counter online threats, harming their reputation.
What type of impact is the harm to reputation as a result of the attack?



Question 12

URL: ttpoll.eu
Session ID: cs290

A start-up developed a machine learning model designed to connect people based on their personal interests. A big company has then bought the start-up and is currently using the algorithm to connect jobseekers with employers. What type of bias is likely to appear?

9% a. Aggregation bias

14% b. Measurement bias

14% c. Intersectional bias

✅ 64% d. Deployment bias

Debriefing of the Mock Test

True/False questions

Question 14

URL: ttpoll.eu
Session ID: cs290

You found a dataset with 5 variables, all self-reported by participants: eye-color, extraversion and 3 health-related variables. When analyzing the data you identify that:
- there are positive and substantial correlations among the 3 health variables
- there is a positive and substantial correlation between eye-color and extraversion
- there is no correlation between eye-color and the health variables

Eye-color is a sensitive attribute:

Debriefing of the Mock Test

Case studies

Question 17 / Case 1: Harm modeling

Category	Type of harm	Social assistant chatbot
Humans	Physical injury	**A)**
Allocation of Resources	Opportunity loss	**B)**
Human Rights	Liberty loss	**C)**
	D)	Most intimate feelings are now "public"
Social System Harms	Social detriment	**E)**

Let's NOT use the vote system

Post your ideas:
https://speakup.epfl.ch
Room key: **80844**

Question 17 / Case 1: Harm modeling

Category	Type of harm	Social assistant chatbot
Humans	Physical injury	**A)**
Allocation of Resources	Opportunity loss	**B)**
Human Rights	Liberty loss	**C)**
	D)	Most intimate feelings are now "public"
Social System Harms	Social detriment	**E)**

Post your ideas:
https://speakup.epfl.ch
Room key: **38348**

Question 18 / Case 2: Value analysis

- We provide 2 stakeholders:
  - Lydia, the webstore manager.
  - Hari, a customer who browses and purchases products on the webshop.
- Goal = identify **2 value-based benefits** and **2 value-based harms for these stakeholders**

  👉 Report 4 times the text for A+B+C+D from the table below

Stakeholder	Key Value	Benefits	Harms	Justification
Stakeholder: **(A)**	Value name and description: **(B)**	Benefit or Harm: **(C)**		It's a value-based benefit/harm for this stakeholder because: **(D)**

Question 18 / Case 2: Value analysis

Which **value-based benefits** did you identify?

- 1 post / value
  - name the value
  - briefly describe for whom and why it’s a benefit

Post your ideas:
https://speakup.epfl.ch
Room key: **31719**

Question 18 / Case 2: Value analysis

Which **value-based harms** did you identify?

- 1 post / value
  - name the value
  - briefly describe for whom and why it’s a harm

Post your ideas:
https://speakup.epfl.ch
Room key: **83726**

Question 18 / Case 2: Value analysis

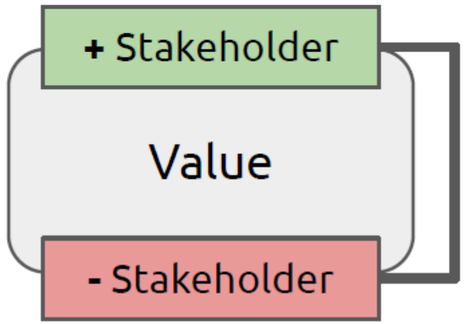
Value based tension map:

Value tensions

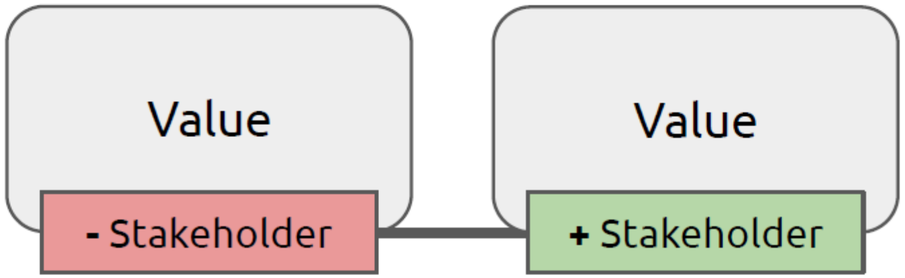
Situations that **oppose** a value-based **benefit** with a value-based **harm**, either **between stakeholders** or **for a single stakeholder**.

C. Value-based tension map

Tension: benefit (+) and harm (-) on the same value



Tension: benefit (+) and harm (-) on different values (requires interpretation of the values in context)



- Map **with the values that you presented in the table/text**
- **Draw + explain** 1 tension

Question 19 / Case 3: Strategies

As the ethics referee of the team, you are asked to anticipate potential consequences of the deployment of the platform in terms of safety and fairness.
Name one strategy seen in the course that you can apply for this task.

👉 **You need to have in mind (or on your A4 paper) the different strategies we have seen in the course and identify situations where they are adapted**

Question 19 / Case 3: Strategies

List the strategies we have seen in class
(**try without looking at your notes**!)

- 1 post / strategy (name of the strategy)

Post your ideas:
https://speakup.epfl.ch
Room key: **51619**

Question 19 / Case 3: Strategies

Here are **4 categories of strategies**
👉 for each category list all the strategies that correspond (duplicates possible):

A. Investigating stakeholders
B. Eliciting values
C. Anticipating impacts
D. Thinking in systems

Strategies by category

Category	Strategies
Investigating stakeholders	• Stakeholder analysis • Bad actors • The people behind the data • Ethics Canvas
Eliciting values	• Analyzing values
Anticipating impacts	• Ethical speculation • Bad actors • STRIDE • Harm modeling • Edge cases • Causal loop diagrams • Inclusive design • Analyzing values • The people behind the data • Datasheets for datasets • Ethics Canvas
Thinking in systems	• Causal loop diagrams
Making decisions	*NOT SEEN YET!*

Question 19 / Case 3: Strategies

As the ethics referee of the team, you are asked to **anticipate potential consequences** of the deployment of the platform in terms of safety and fairness.
Name one strategy seen in the course that you can apply for this task.

Category	Strategies
Anticipating impacts	• Ethical speculation • Bad actors modeling • STRIDE • ~~Harm modeling~~ • Edge cases • Causal loop diagrams • Inclusive design • Analyzing values • People behind the data • Datasheets for datasets • Ethics Canvas

Question 19 / Case 3: Strategies

As the ethics referee of the team, you are asked to anticipate potential consequences of the deployment of the platform in terms of safety and fairness.

Explain the strategy:

a) Justify why this strategy is appropriate for this task (1 sentence).
b) Describe briefly how to apply this strategy (2-3 sentences).

- Steps / phases / questions
- When to use it

Question 19 / Case 3: Strategies

Which **safety** **issue** did you identify?

- 1 post / issue: brief description

Post your ideas:
https://speakup.epfl.ch
Room key: **20158**

Question 19 / Case 3: Strategies

Which **fairness** **issue** did you identify?

- 1 post / issue: brief description

Post your ideas:
https://speakup.epfl.ch
Room key: **00774**

Tips for effective revisions

Revision techniques

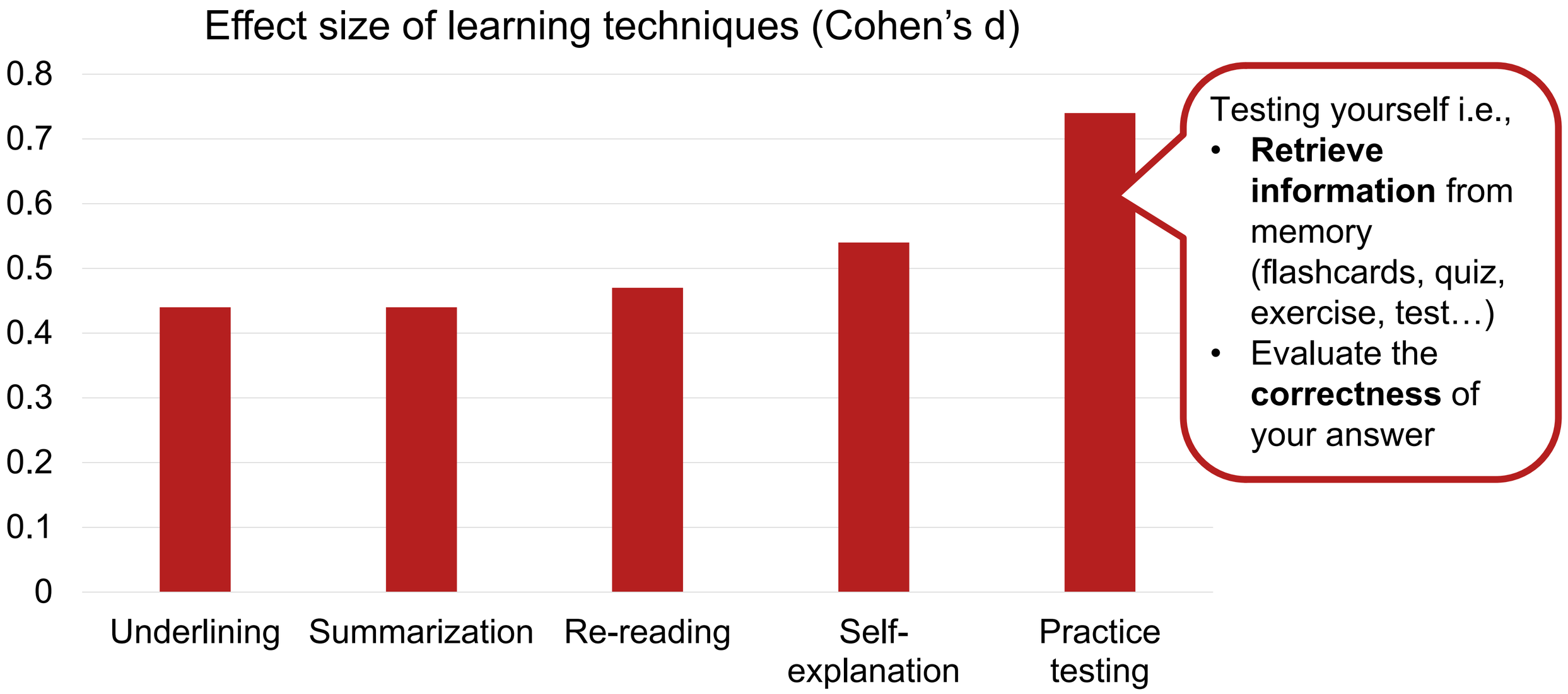
URL: ttpoll.eu
Session ID: cs290

Select the technique(s) that you generally use for your revisions:

a. Re-read notes/course documents
b. Highlight important points in notes/course documents
c. Create summaries from notes and/or course documents
✅ d. Create summaries from memory (without looking at notes)
✅ e. Use flashcards / self-made questions
f. Re-read the solutions of exercises / case studies
✅ g. Re-do exercises / case studies then check the solutions
h. Other

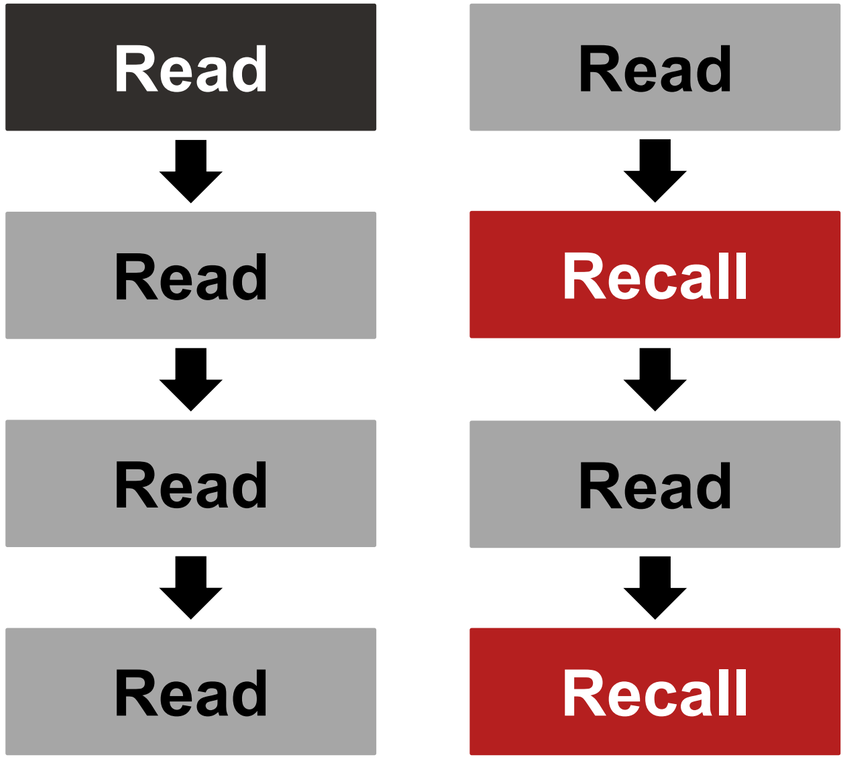
Learning techniques

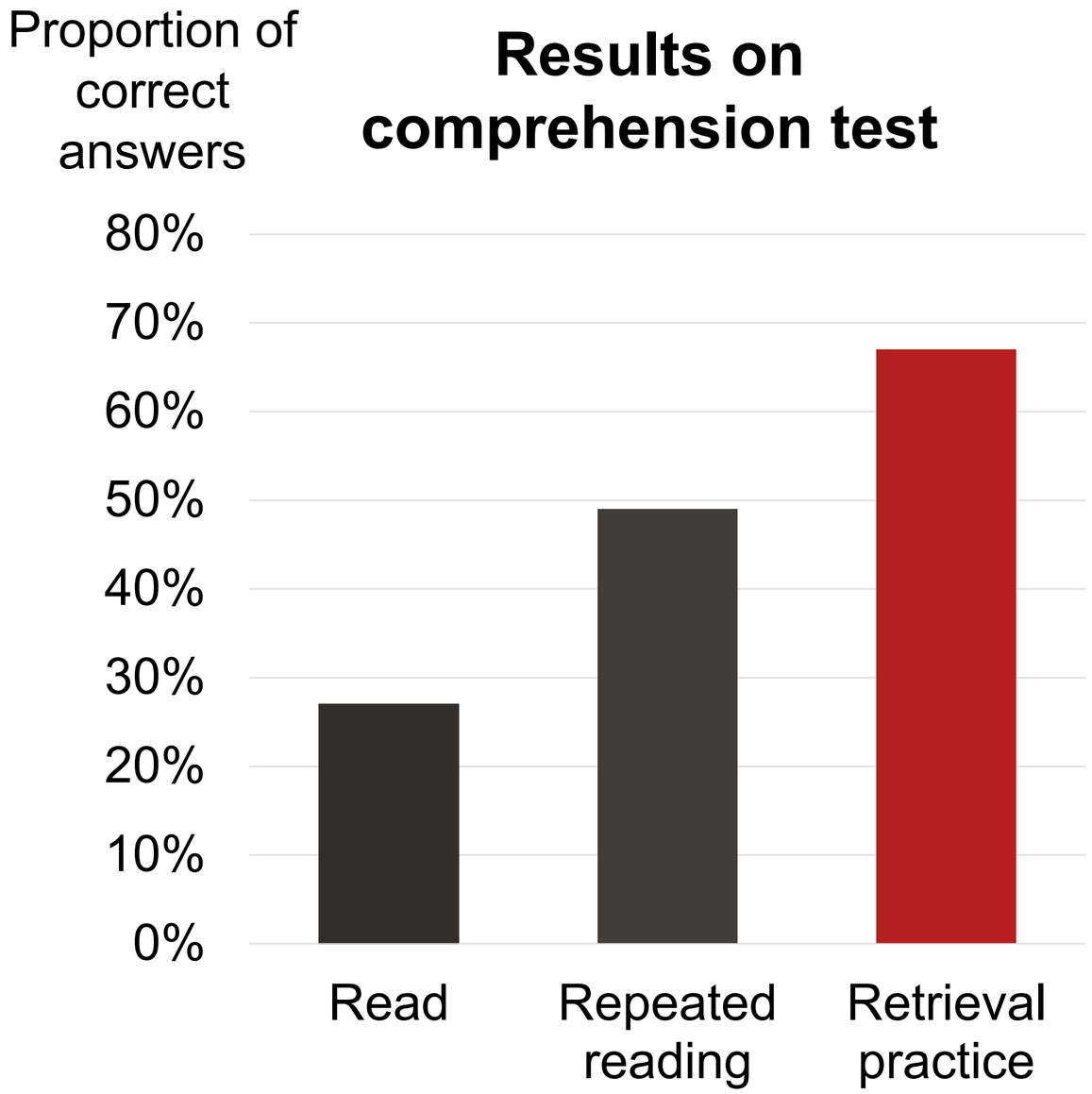
(Donoghue & Hattie, 2021)



Retrieval practice

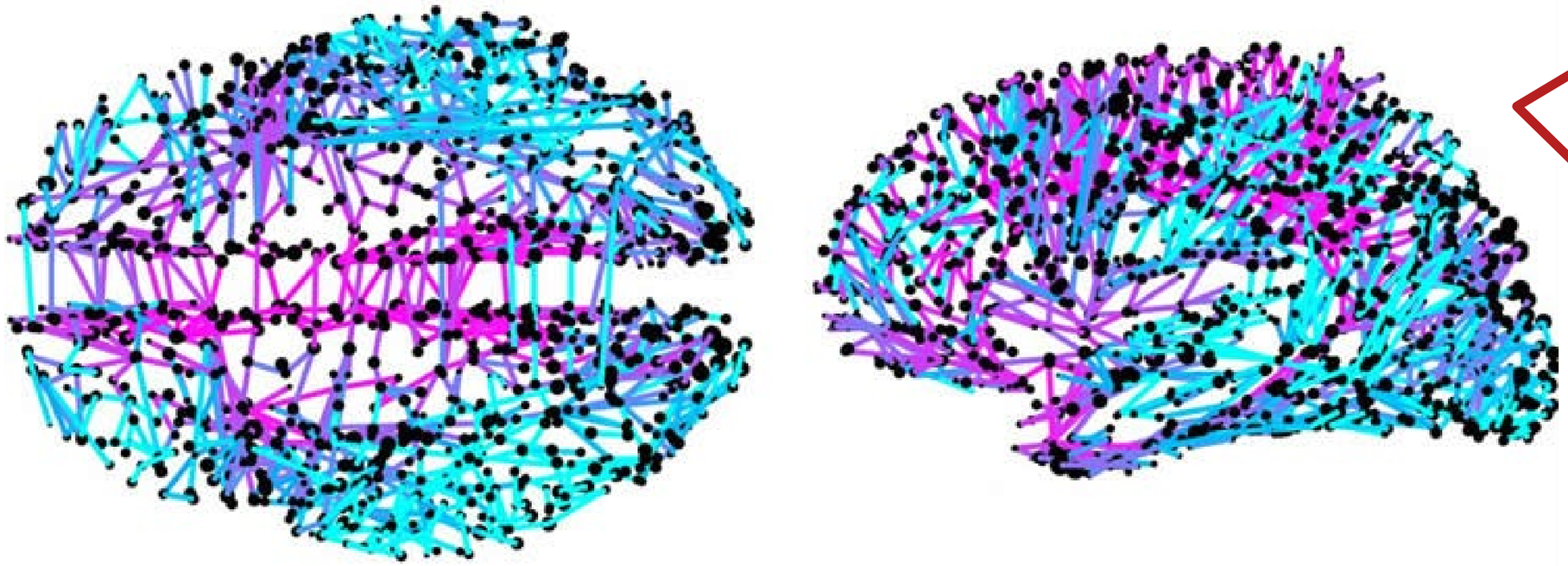
(Karpicke and Blunt, 2011)





Recall = *reconstruct*

(Mišić et al., 2015)



Times at which different connections in the brain are used to spread information

Information is NOT “stored” into memory like in a “drawer”, instead it is **reconstructed** from a **network of connections**

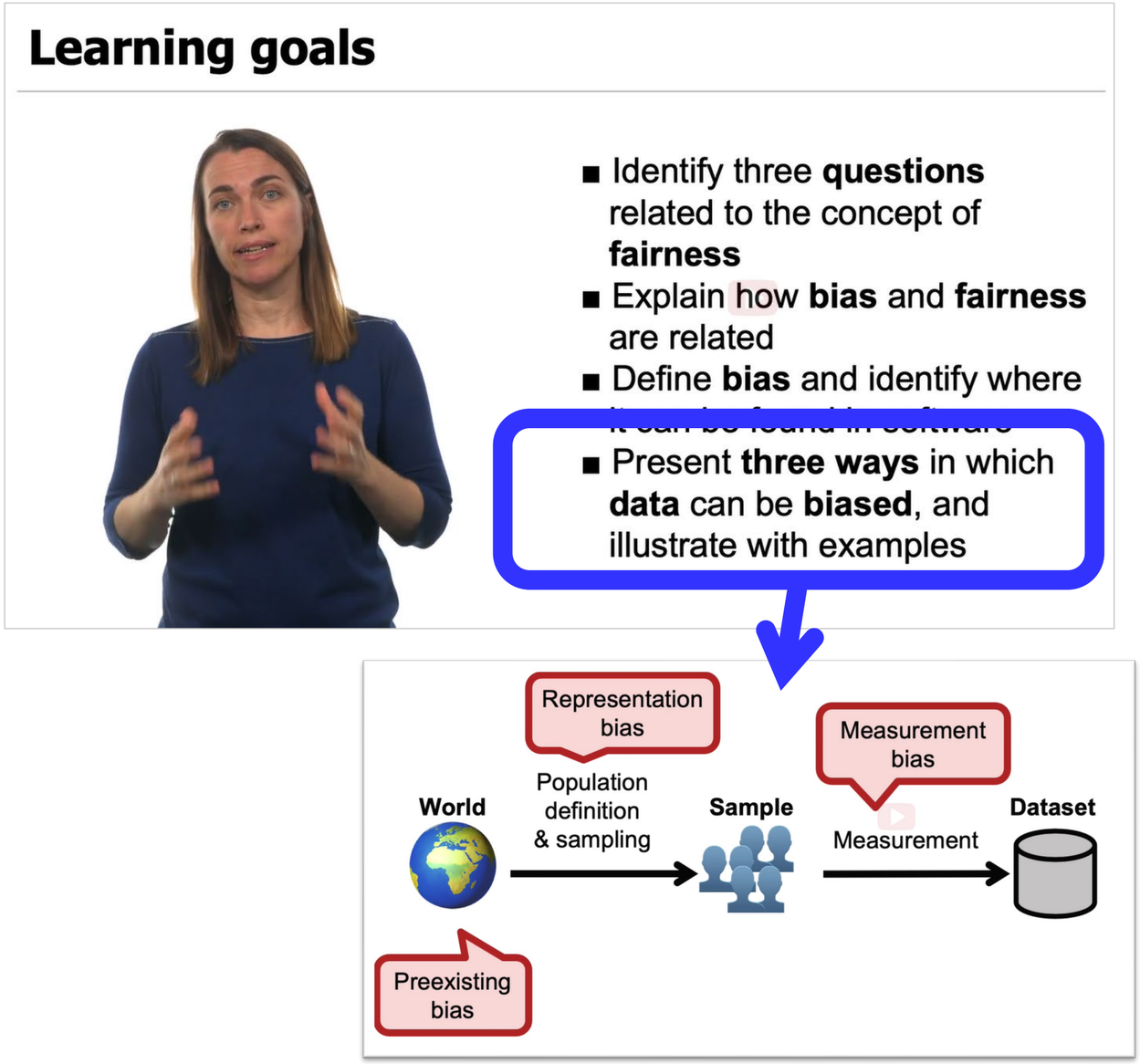
👉 **You should practice reconstruction**

Flashcards

Use the **learning goals** at the beginning of the videos to create **flashcards**

1 goal = 1 card (or more):

- Recto = goal or question
- Verso = answer



Free recall

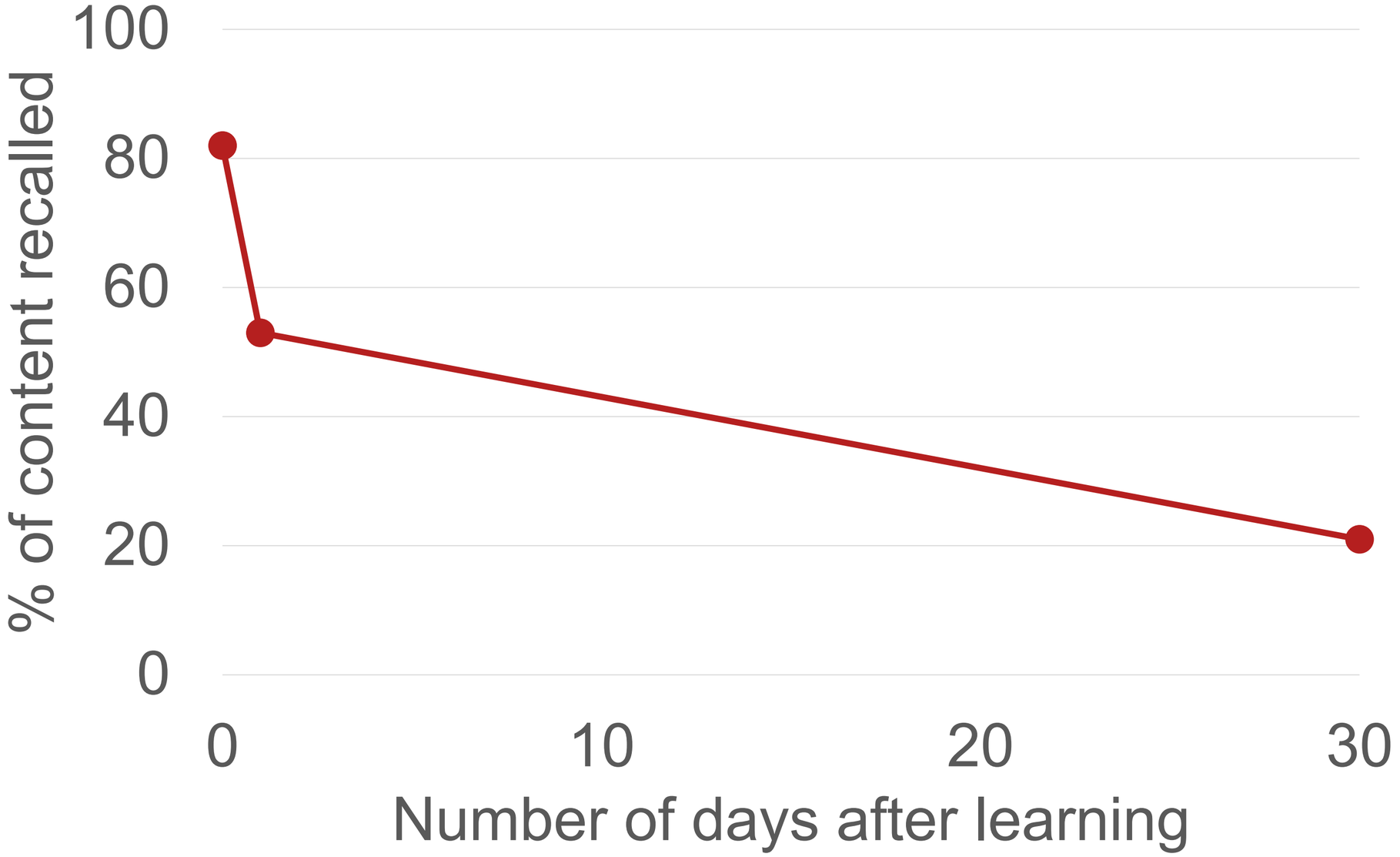
- Take a **blank sheet** of paper
- **Note down** everything you remember on:
  - A video
  - A module
  - A chapter
- Then **check** against your notes

Other suggestions

- **Re-do the case studies without the “proposed answers”**
  Then check with the “proposed answers”
- **Re-do the blank test without the “solution”**
  + **time yourself**: practice strategy and speed!
- Reuse the quizzes done in class
  ⚠️ Watchout: except questions tagged “exam type”, they are not designed like exam questions (multiple correct answers)
  -> you can transform them into single choice questions
- Prepare your allowed A4 sheet of notes for the exam
  (pay attention to structure + synthetization)

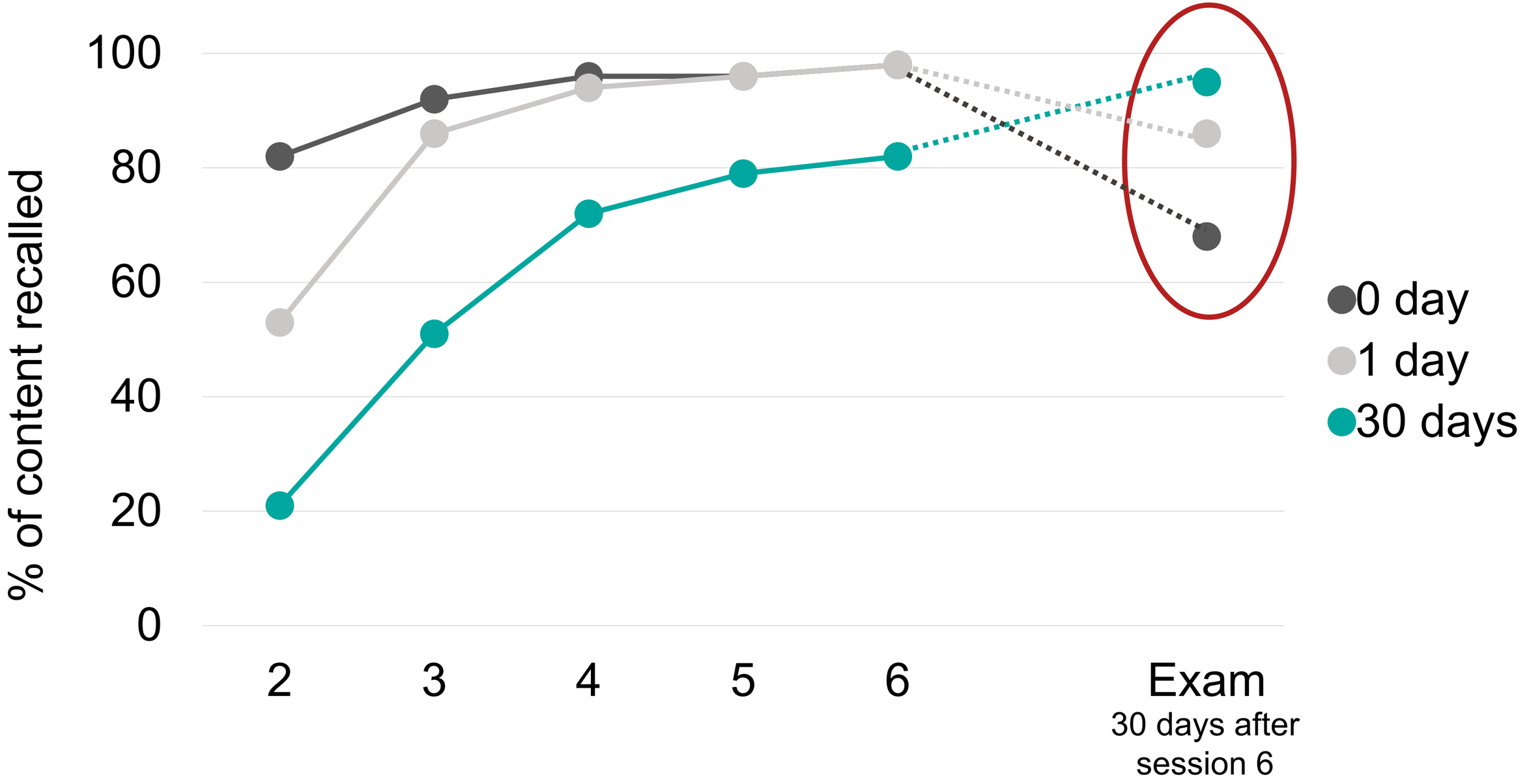
We forget… all the time!

(Bahrick, 1979 ; Dunlosky, 2013)



Spaced learning

(Bahrick, 1979 ; Dunlosky, 2013)



Spaced learning in practice

- **In the course:**
  in part 2, we will practice again with strategies seen in part 1

Date	Week	Lecture (Monday 15h15-17h) in STCC Cloud C	Exercise session (Tuesday 10h15-12h)	Independent study (due before the following Monday)
08/09	1	Getting started	Introduction notebook	Introduction videos and quizzes
15/09	2	Introduction cases (**in CO3**)	Safety 1 notebook	Safety 1 videos and quizzes
22/09	3	public holiday	Safety 2 notebook	Safety 2 videos and quizzes
29/09	4	Safety 2 cases	Fairness 1 notebook	Fairness 1 videos and quizzes
06/10	5	Fairness 1 cases (**in CO3**)	Fairness 2 notebook	Fairness 2 videos and quizzes
13/10	6	Fairness 2 cases	Graded notebook 1	-
20/10			Autumn break	
27/10	7	Graded 1 debriefing	Mock test	-
03/11	8	Mock test debriefing (**in CO3**)	Sustainability 1 notebook	Sustainability 1 videos and quizzes
10/11	9	Sustainability 1 cases	Sustainability 2 notebook	Sustainability 2 videos and quizzes
17/11	10	Sustainability 2 cases	Empowerment 1 notebook	Empowerment 1 videos and quizzes
24/11	11	Empowerment 1 cases	Graded notebook 2	-
01/12	12	Graded 2 debriefing	Empowerment 2 notebook	Empowerment 2 videos and quizzes
08/12	13	Empowerment 2 cases	Graded case	Conclusion videos and quizzes
15/12	14	Conclusion cases	Conclusion review	-
Revisions				
TBD	Exams	Written exam		

- **For your revisions:**
  - “Ideal” interval for an 18-week semester is 2-3 weeks (10-20%)
  - For best effect, intervals should be increased over time

What’s next?

We start Sustainability 1!

Tomorrow, Tuesday 4: notebook on the **carbon footprint** of algorithms

By Monday 10:

- Watch **videos 5.1 to 5.4** + do the **quizzes**
- Finish the notebook
  (and any other leftover from previous weeks)

On Monday 10:

- Interactive questions on the theory
- Work on the **case studies together in class**

References

- Donoghue, G. M., & Hattie, J. A. C. (2021). A Meta-Analysis of Ten Learning Techniques. Frontiers in Education, 6. https://www.frontiersin.org/article/10.3389/feduc.2021.581216
- Karpicke, J. D., & Blunt, J. R. (2011). Retrieval Practice Produces More Learning than Elaborative Studying with Concept Mapping. Science, 331(6018), 772–775. https://doi.org/10.1126/science.1199327
- Mišić, B., Betzel, R. F., Nematzadeh, A., Goñi, J., Griffa, A., Hagmann, P., Flammini, A., Ahn, Y.-Y., & Sporns, O. (2015). Cooperative and Competitive Spreading Dynamics on the Human Connectome. Neuron, 86(6), 1518–1529. https://doi.org/10.1016/j.neuron.2015.05.035
- Bahrick, H. P. (1979). Maintenance of knowledge: Questions about memory we forgot to ask. Journal of Experimental Psychology: General, 108(3), 296. https://doi.org/10.1037/0096-3445.108.3.296
- Dunlosky, J., Rawson, K. A., Marsh, E. J., Nathan, M. J., & Willingham, D. T. (2013). Improving Students' Learning With Effective Learning Techniques: Promising Directions From Cognitive and Educational Psychology. Psychological Science in the Public Interest, 14(1), 4–58. https://doi.org/10.1177/1529100612453266