

AI in an Open Society

course 2025

Lecture 8: bringing it all together

Dr. Daan van Rooij

Lecturer Experimental Psychology

Summary:

- Exam
- Mock exam
- Assignment
- Lectures so far

Setup of this course: schedule part 1

Schedule

The color coding refers to the lecture type: **Methodological**, **thematic**, or **practical**.



Lecture				Literature	Professor	
	Date	Time	Week			
1	Wed 22-04	13:15 – 15:00	17	Introduction & IOS	1	Van Rooij
2	Fri 24-04	11:00 – 12:45	17	<u>Modeling individual decision making</u>	2	Van Maanen
3	Wed 29-04	13:15 – 15:00	18	Autonomy, psychology of interaction with AI systems	3	Hortensius
4	Fri 01-05	11:00 – 12:45	18	Emergence of collective patterns in relation to norms	4	Klein
5	Wed 06-05	13:15 – 15:00	19	Attitude & Linguistic Models	5	Van der Vegt
6	Fr 08-05	11:00 – 12:45	19	Trust in AI	6	<u>Grimmelikhuijsen</u> + Liefoghe
7	Wed 13-05	13:15 – 15:00	20	Medical AI / Digital Twins	7	Van Rooij + Bontje
8	Wed 20-05	13:15 – 15:00	21	Putting everything together and future outlook		Van Rooij
	Fri 29-05	11:00 – 13:00	22	<u>Mid-term exam</u>		

Setup of this course: schedule part 1

1. Elliott, K., Price, R., Shaw, P. et al. (2021) Towards an equitable digital society: Artificial Intelligence (AI) and Corporate Digital Responsibility (CDR). *Society*, 58, 179–188. doi: 10.1007/s12115-021-00594-8
2. Palada, H., Neal, A., Vuckovic, A., Russell, M., Samuels, K. & Heathcote, A. (2016). Evidence accumulation in a complex task: Making choices about concurrent multiattribute stimuli under time pressure. *Journal of Experimental Psychology: Applied*, 22 (1), 1-23. doi: 10.1037/xap0000074
3. Rahwan, I., Cebrian, M., Obradovich, N., Bongard, J., Bonnefon, J.-F., Breazeal, C., ... Wellman, M. (2019). *Machine behaviour. Nature*, 568, 477–486. doi:10.1038/s41586-019-1138-y
4. Douven, I. & Hegselmann, R. (2021), Mis- and disinformation in a bounded confidence model, *Artificial Intelligence* 291, 103415, doi:10.1016/j.artint.2020.103415
5. van der Vegt, I., Kleinberg, B., & Gill, P. (2023). Proceed with caution: on the use of computational linguistics in threat assessment. *Journal of Policing, Intelligence and Counter Terrorism*, 18(2), 231-239.
6. ~~Grimmelikhuijsen, S., & Meijer, A. (2022). Legitimacy of algorithmic decision-making: Six threats and the need for a calibrated institutional response. *Perspectives on Public Management and Governance*, 5(3), 232-242.~~
7. Wang, W., He, F., Li, Y., Tang, S., Li, X., Xia, J., & Lv, Z. (2023). Data information processing of traffic digital twins in smart cities using edge intelligent federation learning. *Information Processing & Management*, 60(1), 103171

Setup of this course: Midterm Exam

Friday May 29d

11:00 – 13:00

EDUC - BETA

Rupert D (extra time)

Laptops are provided

Mock Exam

5 questions about the 7 first lectures and assigned reading

Used to practice your insight into lecture topics and questions

Setup of this course: schedule part 2

8	Fri 05-06	11:00 – 15:00	23	Workgroup	Van Rooij, Van Maanen, Liefoghe, v/d Vegt, Klein
9	Fri 12-06	11:00 – 15:00	24	Pitches + Workgroup	Van Rooij, Van Maanen
10	Fri 19-06	11:00 – 15:00	25	Workgroup	Van Rooij, Van Maanen
	Fri 26-06		26	Workgroup	Homework group, no location
	Sun 28-06		26	Deadline proposal 23:59	

← We make groups,
think about an RQ

← Check your research
question!

← Present your Pitch!

Grant assignment:

Write a grant proposal for a new AI centred research project featuring one of the topics of the UU IOS-Platforms.

Where the **IOS-platforms** must provide the theme and goals of the new research project, the **lectures** within the AIOS course should provide the basis for the method and application of the AI that will be applied in this project.

Learning goals: After this assignment you are familiar with the basic structure of a grant proposal and you can apply this structure to a specific topic within AI centred research in Open Societies.

Grant assignment:

1. Title of the proposal
2. Summary of the proposed research
3. Keywords
4. Main body
 - Background
 - Objective
 - Scientific relevance & Novelty
 - Approach / Experimental design
5. Budget and Feasibility of your research
6. Proof of Concept

Data simulation showing how your algorithm or model would work

An experimental setup, testing your AI solution on a small number of test subjects.

Preliminary data analysis of an existing dataset using your proposed method

A prototype algorithm demonstrating its functionality

7. Impact of the research

Lectures:

Summary and potential implementation in the assignment

Lectures:

Lecture 2, Dr. van Maanen:

Models of perceptual decision making

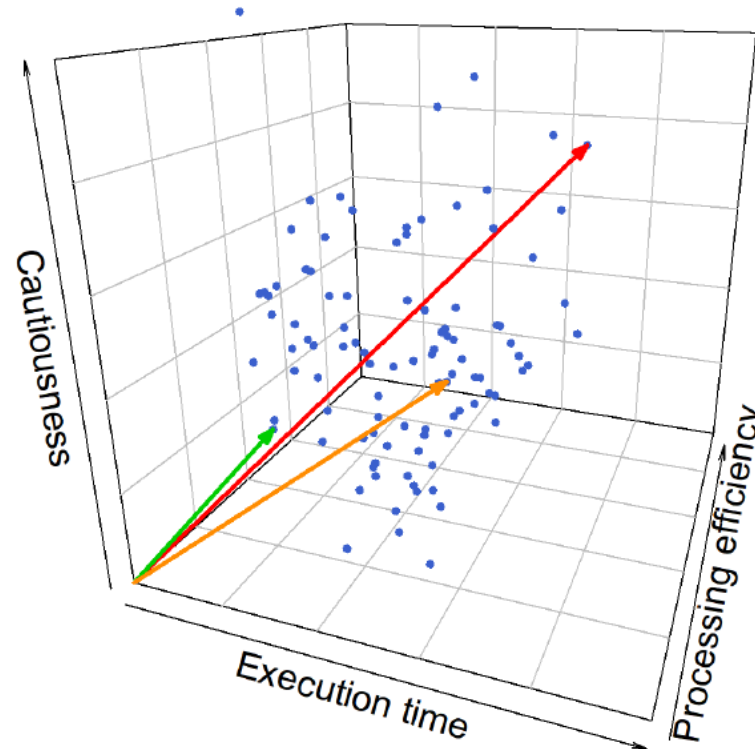
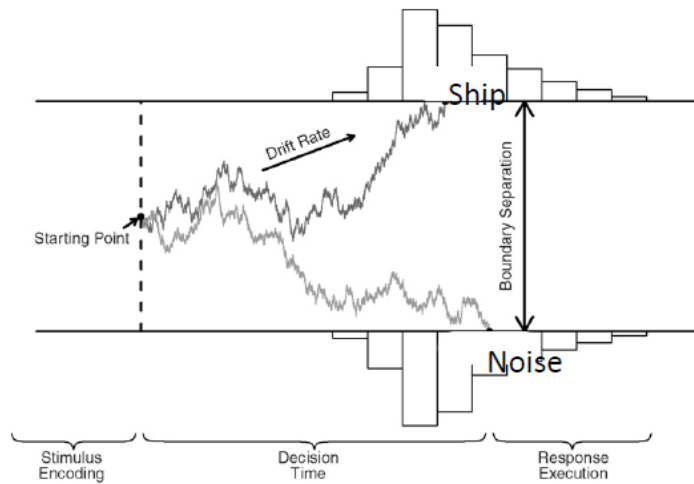
- How cognitive models can be used to explain decision making in humans
- Influence of AI on an individual level: how do different people make decisions?

Potential topics:

- Potential relevance for all topics like selection and evaluation of employees. E.g. Future of work, security and society

Lecture 2, Dr. van Maanen: potential topics

Individual differences



Good decision maker

Medium decision maker

Poor decision maker

Example question: Using (AI) enhanced image processing, which factors decrease decision time and/or increase decision accuracy for radiologists examining medical scans? Can we predict which cues would work for which doctors?

e.g. Future of work

Lectures:

Lecture 3, Dr. Hortensius:

Psychology of interaction with AI systems

- How can our interaction with online AI systems be used/abused?

Potential topics:

- Potential relevance for all topics dealing with user online data profiles, usability and ethics around using this data, and how people are influenced by interaction with AI systems.

Lectures:

Example:

- Investigating the bias in Youtube algorithm

(Future of Democracy, Behavior and Institutions)

Lectures:

Lecture 4, Dr. Klein:

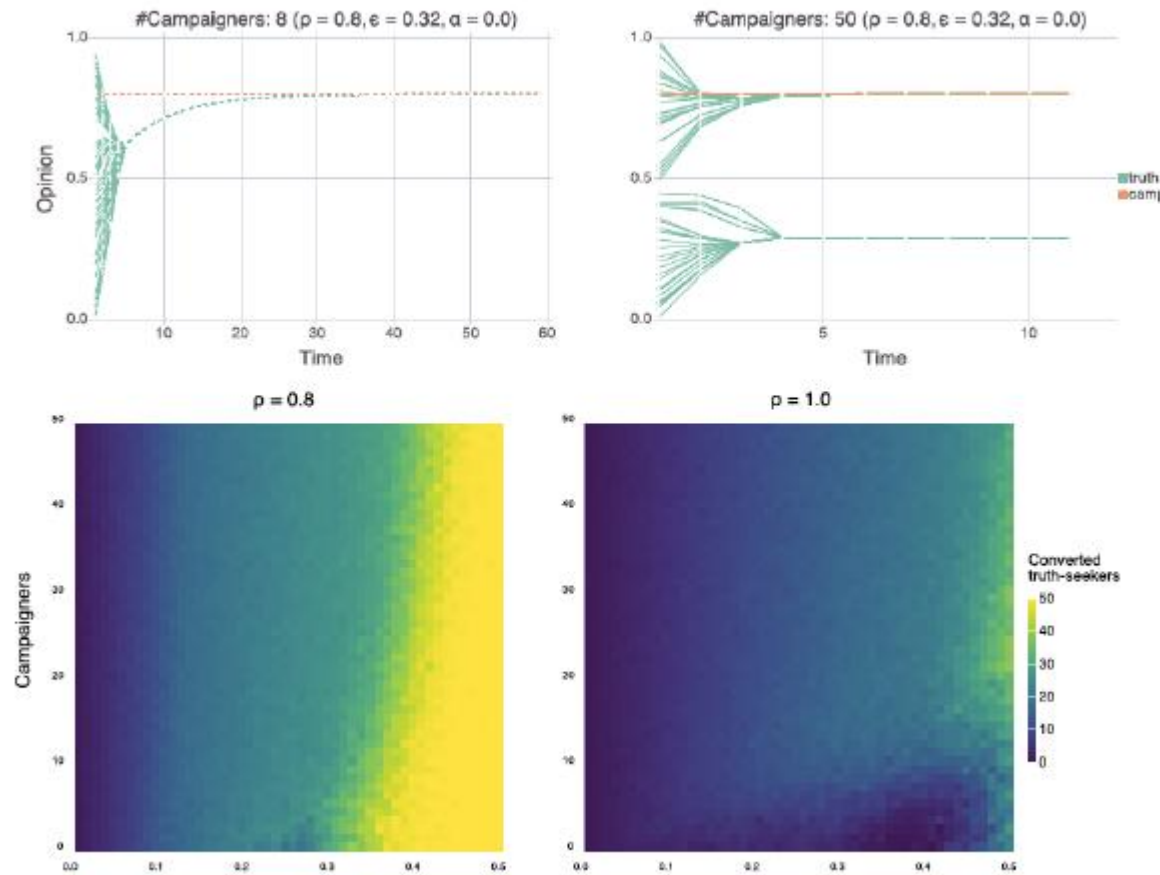
Emergence of collective patterns

- Using models like the bounded confidence model to explain behavior of groups of people.

Potential topics:

- Security: crowd control, emergent crowd behavior. Can you model tipping points in groups of people using netlogo?

Lecture 4, Dr. Klein: potential topics



Example question: How might the results of the US presidential election have changed with altered campaigning compositions and stances? How does the use of disinformation alter voting behavior in the US presidential elections?

E.g. Future of citizen based initiatives, Future of Democracy, Open Cities

Lectures:

Lecture 5, Dr. van der Vegt:

Attitudes & NLP

- Using NLP/AI tools to predict human behavior on societal level.

Potential topics:

- Security: can we predict different types of violent attitudes and behavior using publicly available data and AI tools?

Lecture 5, Dr. van der Vegt: potential topics

Example from previous year:

Predicting Riots Surrounding Football Matches with Trained AI

- Publicly available data on football matches, coded by event timing
- Long short-term memory classification model

Predicting violent events based on language use on far-right forums

- Publicly available language data used on far-right forum
- NLP model predicting far-right terror attacks based on language use

Lectures:

Lecture 6, Dr. van Rooij & Dr. Bontje: AI in medicine & Digital Twins

- How does the employment of AI in medicine change medical decision making of the future?
- How do digital twins impact changes in society?

Lectures:

Lecture 7, Dr. van Rooij & Dr. Bontje: AI in medicine & digital twins

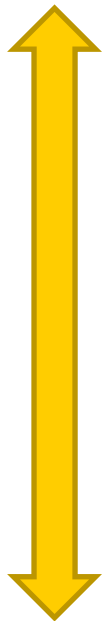
Potential topics:

- How do changes in medical AI use alter the openness, fairness, security or democratic nature of our future society?
- How can digital twins be used to shape evolution of institutions? (link with modelling approaches from lecture 4).

Lectures:

Summary and potential implementation in the assignment

Scaling:



Human cognition - computational modelling, behavioral prediction

Human/AI psychology - interaction with AI systems, trust and ethics

Human/AI networks - groups of people, systems

Human/AI society - governments, societal groupings

Mock exam:

Question 1 (Lecture 1: intro)

During the introductory lecture, the IOS platforms defined by Utrecht University were discussed.

Name two of the fifteen IOS platforms.

Describe the aim of each topic you named in one sentence each.

Mock exam:

Question 2 (Lecture 3, modelling individual decision making)

What parameter of the Linear Ballistic Accumulator Model is typically associated with time pressure?

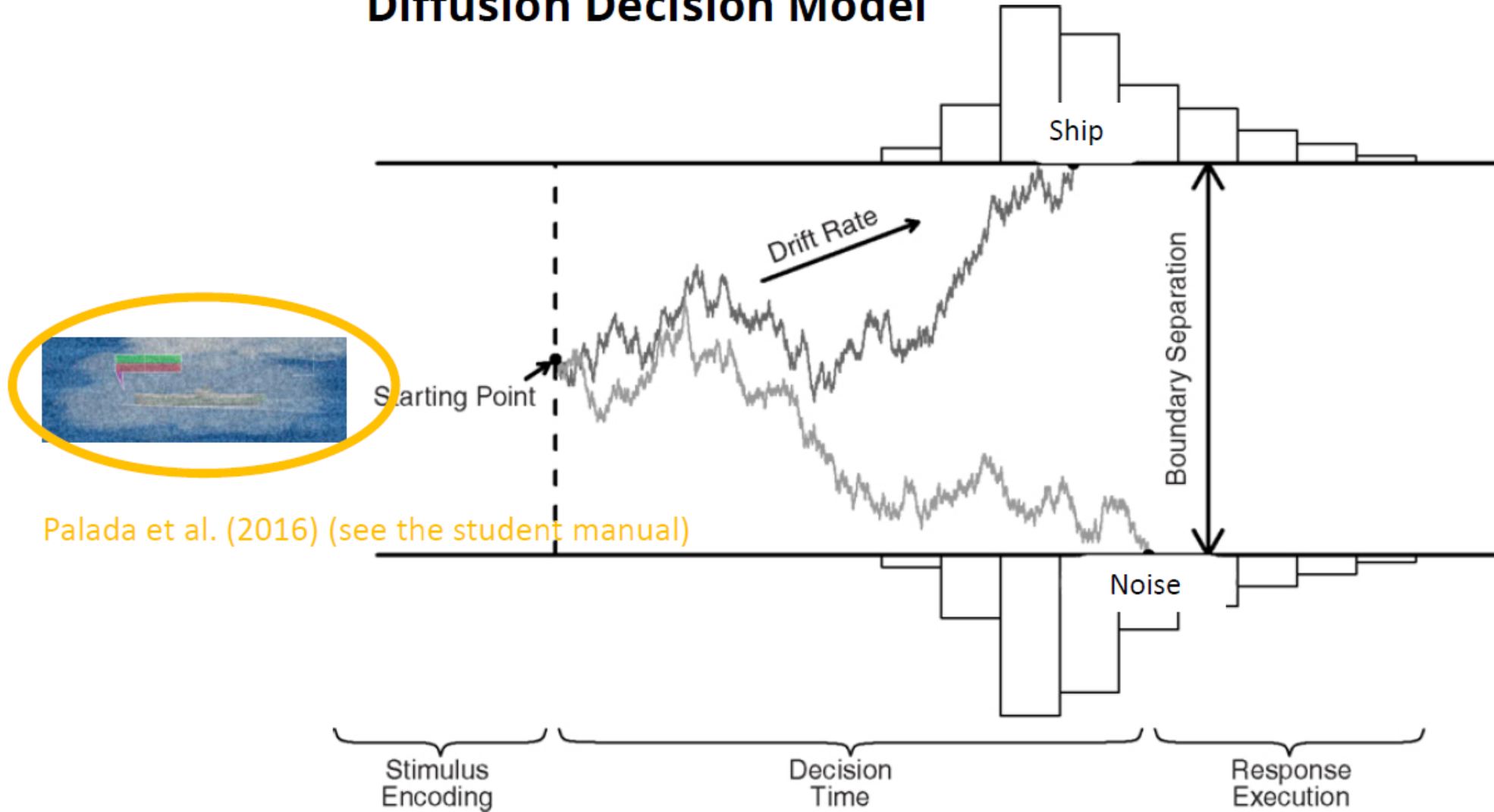
response caution/threshold/boundary/B

According to Palada et al. (2016), which diffusion model parameter explains why Unmanned Aerial Vehicle operators miss more target ships when these are obscured by clouds than when these are not? Explain your answer.

drift rate/evidence accumulation/clouds make it more difficult to extract all information, increasing difficulty, decreasing rate.

Mock exam:

Diffusion Decision Model



Mock exam:

Question 3: (lecture 4 emergence of collective patterns)

What is a tipping point of an emergent system? Explain and illustrate with an example.

first part: small changes in one parameter have drastic effects on the output

example part : (should really convince us of a system where this happens)

According to Douven and Hegselmann (2021) how can the presence of free riders impact the performance of truth seekers? Explain this finding.

first part: Several aspects: they don't reduce the amount of truth seekers reaching truth, but can have slightly negative impact on truth seekers performance (measured in mean square error from the truth), but only if campaigners are subtle.

Explanation: Free riders, when also influenced by the campaigners, can prevent truth seekers from reaching the truth (dragging them away from it)

Mock exam:

Question 4: (lecture 5, Attitude & NLM's by Dr. vd Vegt)

In the lecture on Attitude & NLM's by dr. Van der Vegt, we talked about using AI to predict events based on human group behavior. Explain one of the key problems associated with using AI to predict rare events, like terrorist attacks

Challenge: The base rate fallacy

- Imagine: we have 99% accurate system
- 100 million individuals
- Base rate: 0.01% is violent (10,000)

		Prediction		
		<u>Violence</u>	<u>No violence</u>	Total
Reality	<u>Violence</u>	9,900	100 (false negatives)	10,000
	<u>No violence</u>	999,900 (false positives)	98,990,100	99,990,000
Total				

Mock exam:

Question 5: (lecture 3, interaction with AI systems, by Dr. Hortensius)

In the lecture on interaction with AI systems by Dr. Hortensius, we talked about digital traces. Digital traces are sometimes seen as the goldmine of data for psychological science. Which of the following statements about digital traces is/are *true* (multiple possible)?

•(B)

The 'digital dossier' can capture only explicit behaviours of individuals, thus limiting its usefulness in reducing experimental bias.

•(D)

The use of digital traces can potentially raise ethical questions around privacy and data usage (e.g., consent and ownership).

•(E)

Digital traces have a stable meaning and are not affected by self-report or awareness biases.

•(C)

Digital traces can provide a detailed recording and measurement of behaviours that include contextual factors.

•(A)

Digital traces allow for larger and more varied sample populations, potentially moving beyond traditionally WEIRD samples.

Mock exam:

Question 5: (lecture 3, interaction with AI systems, by Dr. Hortensius)

In the lecture on interaction with AI systems by Dr. Hortensius, we talked about digital traces. Digital traces are sometimes seen as the goldmine of data for psychological science. Which of the following statements about digital traces is/are *true* (multiple possible)?

•(B)

The 'digital dossier' can capture only explicit behaviours of individuals, thus limiting its usefulness in reducing experimental bias.

•(D)

The use of digital traces can potentially raise ethical questions around privacy and data usage (e.g., consent and ownership).

•(E)

Digital traces have a stable meaning and are not affected by self-report or awareness biases.

•(C)

Digital traces can provide a detailed recording and measurement of behaviours that include contextual factors.

•(A)

Digital traces allow for larger and more varied sample populations, potentially moving beyond traditionally WEIRD samples.

Questions?