

AI for an Open Society

Lecture 2: Modeling individual decision making

Dr. Leendert van Maanen

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	Modeling individual decision making	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	Grimmelikhuijsen + Liefoghe
7	Wed 13-05	13:15 – 15:00	20	Medical AI / Digital Twins	7	Van Rooij + Bontje

Progressively
higher level
of description



Application-oriented



*The important role of cognitive psychology in
formal models of thinking and AI*

Herbert Simon
(1916-2001)



1956 Dartmouth workshop on Artificial Intelligence

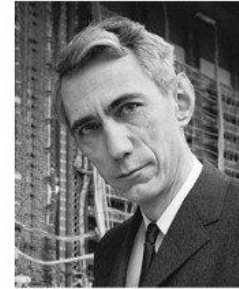
1956 Dartmouth Conference: The Founding Fathers of AI



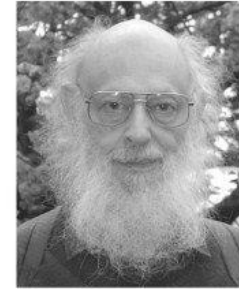
John MacCarthy



Marvin Minsky



Claude Shannon



Ray Solomonoff



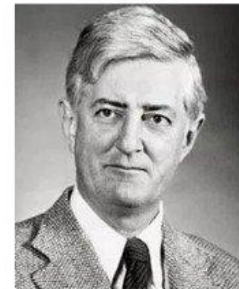
Alan Newell



Herbert Simon



Arthur Samuel



Oliver Selfridge



Nathaniel Rochester



Trenchard More

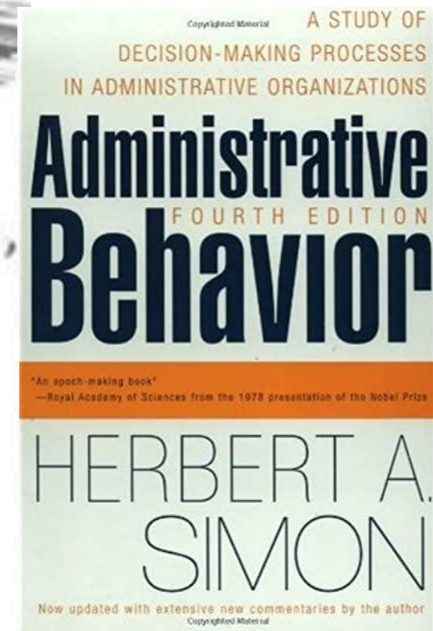


Herbert Simon (1916-2001)

- Organizational Psychology
- Models of cognition → AI
- Nobel Laureate 1978 "for his pioneering research into the decision-making process within economic organizations"



Cf Future of Work hub →

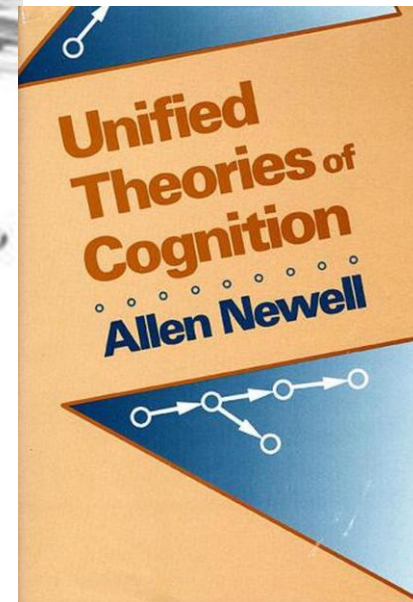


Allen Newell (1927 - 1992)

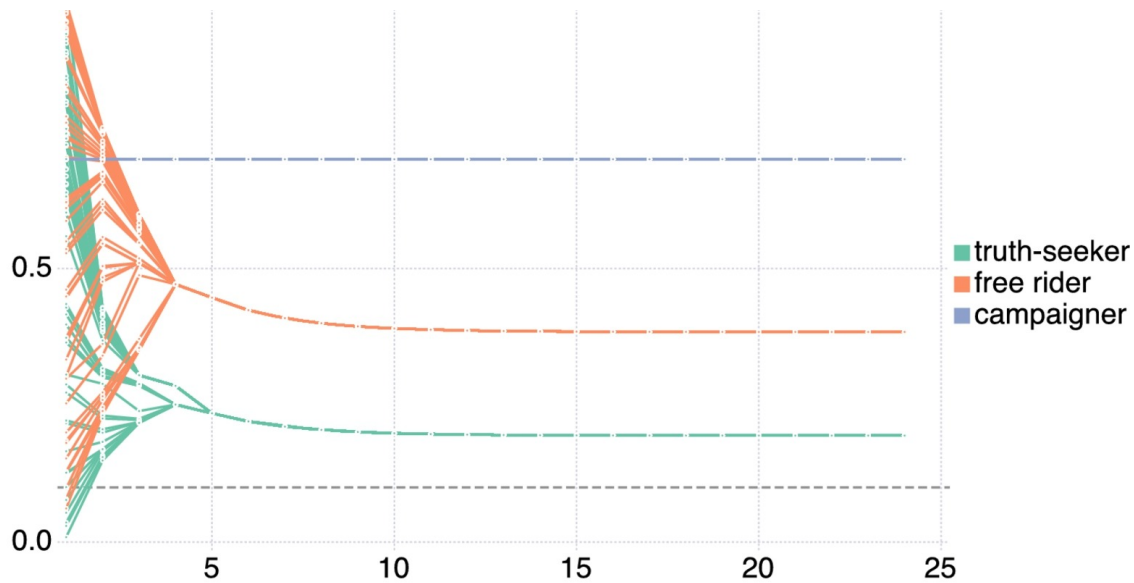
- Mathematics/Computer science
- Time scales of human action



<u>Scale</u> (sec)	<u>Time Units</u>	<u>System</u>	<u>World</u> (theory)
10^7	months		SOCIAL BAND
10^6	weeks		
10^5	days		
10^4	hours	Task	RATIONAL BAND
10^3	10 min	Task	
10^2	minutes	Task	
10^1	10 sec	Unit task	COGNITIVE BAND
10^0	1 sec	Operations	
10^{-1}	100 ms	Deliberate act	
10^{-2}	10 ms	Neural circuit	BIOLOGICAL BAND
10^{-3}	1 ms	Neuron	
10^{-4}	100 μ s	Organelle	

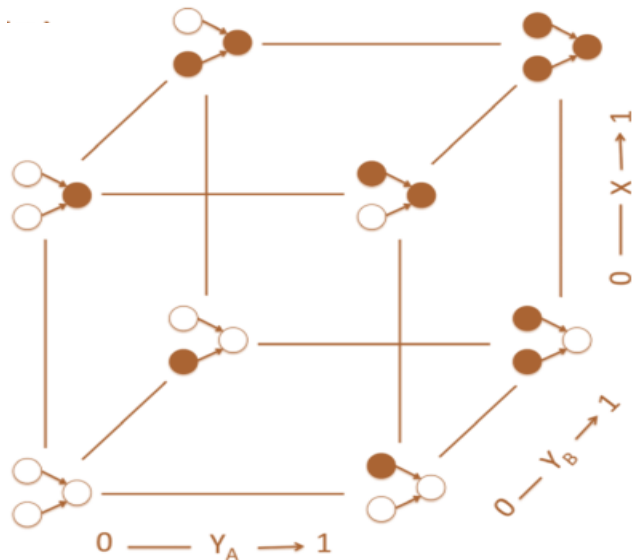
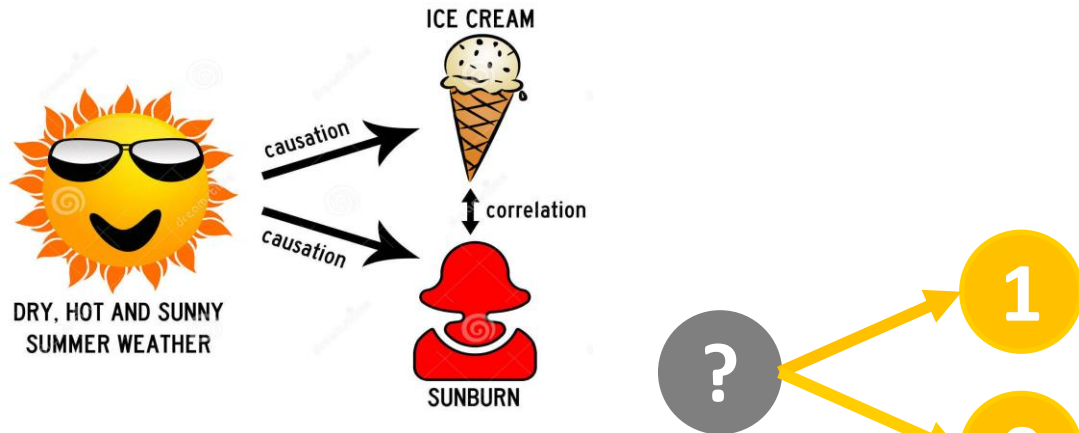


time scales of human action: social band



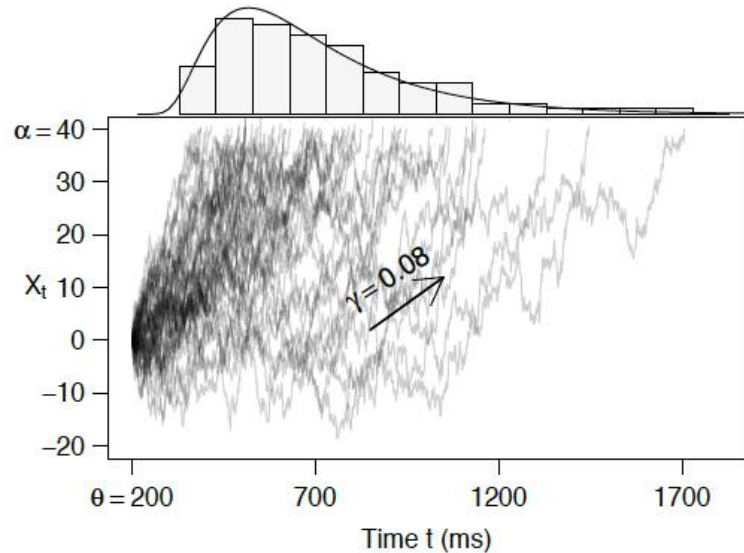
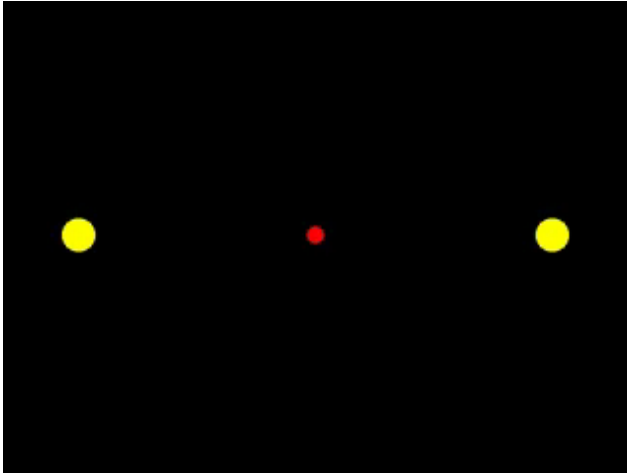
<u>TIME SCALE OF HUMAN ACTION</u>			
<u>Scale (sec)</u>	<u>Time Units</u>	<u>System</u>	<u>World (theory)</u>
10^7	months		SOCIAL BAND
10^6	weeks		
10^5	days		
10^4	hours	Task	RATIONAL BAND
10^3	10 min	Task	
10^2	minutes	Task	
10^1	10 sec	Unit task	COGNITIVE BAND
10^0	1 sec	Operations	
10^{-1}	100 ms	Deliberate act	
10^{-2}	10 ms	Neural circuit	BIOLOGICAL BAND
10^{-3}	1 ms	Neuron	
10^{-4}	100 μ s	Organelle	

time scales of human action: rational band



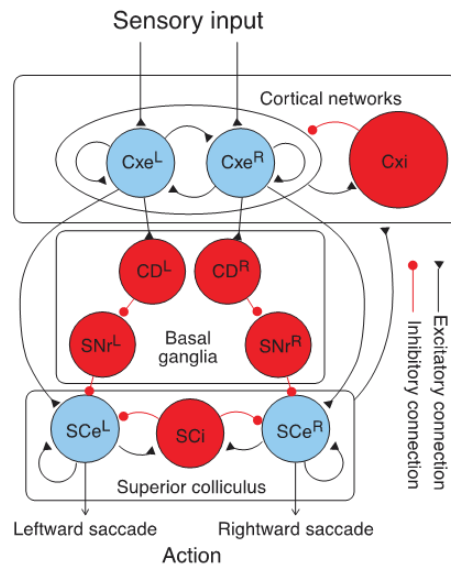
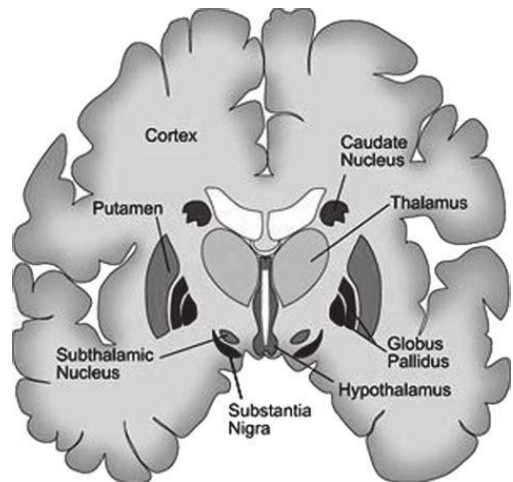
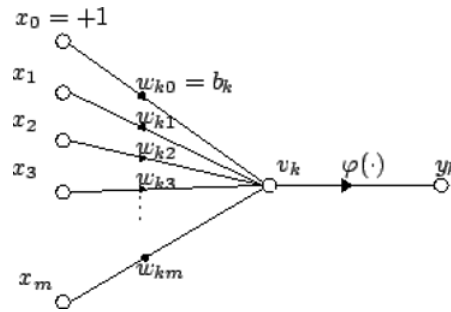
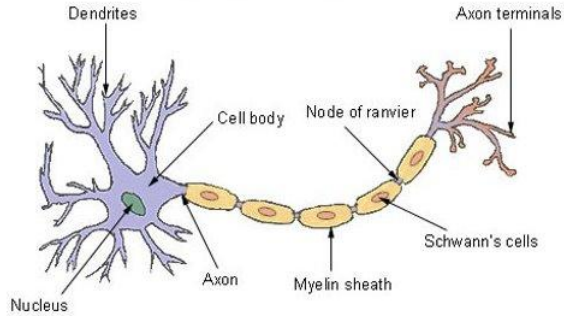
TIME SCALE OF HUMAN ACTION			
<u>Scale (sec)</u>	<u>Time Units</u>	<u>System</u>	<u>World (theory)</u>
10^7	months		SOCIAL BAND
10^6	weeks		
10^5	days		
10^4	hours	Task	RATIONAL BAND
10^3	10 min	Task	
10^2	minutes	Task	
10^1	10 sec	Unit task	COGNITIVE BAND
10^0	1 sec	Operations	
10^{-1}	100 ms	Deliberate act	
10^{-2}	10 ms	Neural circuit	BIOLOGICAL BAND
10^{-3}	1 ms	Neuron	
10^{-4}	100 μ s	Organelle	

time scales of human action: cognitive band



TIME SCALE OF HUMAN ACTION			
<u>Scale</u> (sec)	<u>Time Units</u>	<u>System</u>	<u>World</u> (theory)
10^7	months		
10^6	weeks		SOCIAL BAND
10^5	days		
10^4	hours	Task	
10^3	10 min	Task	RATIONAL BAND
10^2	minutes	Task	
10^1	10 sec	Unit task	
10^0	1 sec	Operations	COGNITIVE BAND
10^{-1}	100 ms	Deliberate act	
10^{-2}	10 ms	Neural circuit	
10^{-3}	1 ms	Neuron	BIOLOGICAL BAND
10^{-4}	100 μ s	Organelle	

time scales of human action: biological band

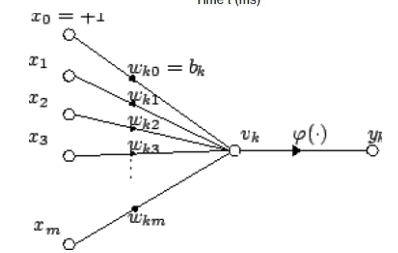
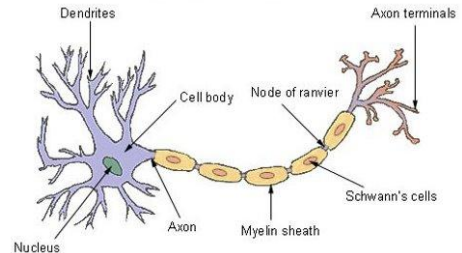
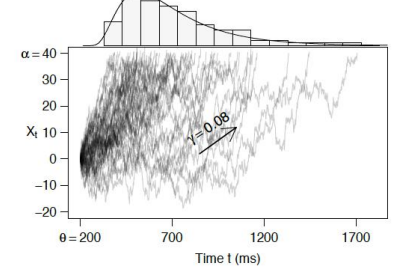
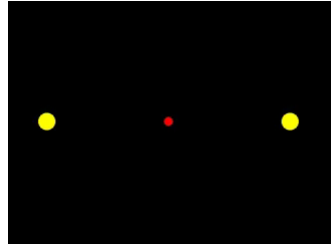
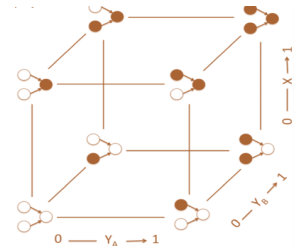
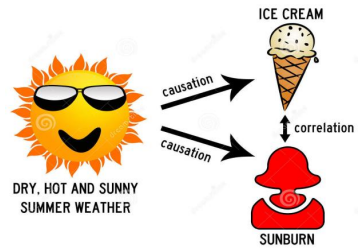
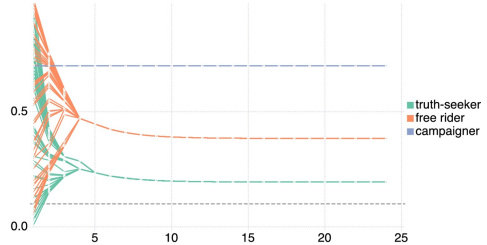


TIME SCALE OF HUMAN ACTION

<u>Scale</u> (sec)	<u>Time Units</u>	<u>System</u>	<u>World</u> (theory)
10^7	months		
10^6	weeks		SOCIAL BAND
10^5	days		
10^4	hours	Task	
10^3	10 min	Task	RATIONAL BAND
10^2	minutes	Task	
10^1	10 sec	Unit task	
10^0	1 sec	Operations	COGNITIVE BAND
10^{-1}	100 ms	Deliberate act	
10^{-2}	10 ms	Neural circuit	
10^{-3}	1 ms	Neuron	BIOLOGICAL BAND
10^{-4}	100 μ s	Organelle	

time scales of human action

	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	Modeling individual decision making	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 Grimmelikhuisen + Liefoghe
7	Wed 13-05	13:15 - 15:00	20	Medical AI / Digital Twins	7 Van Rooij + Bontje



TIME SCALE OF HUMAN ACTION

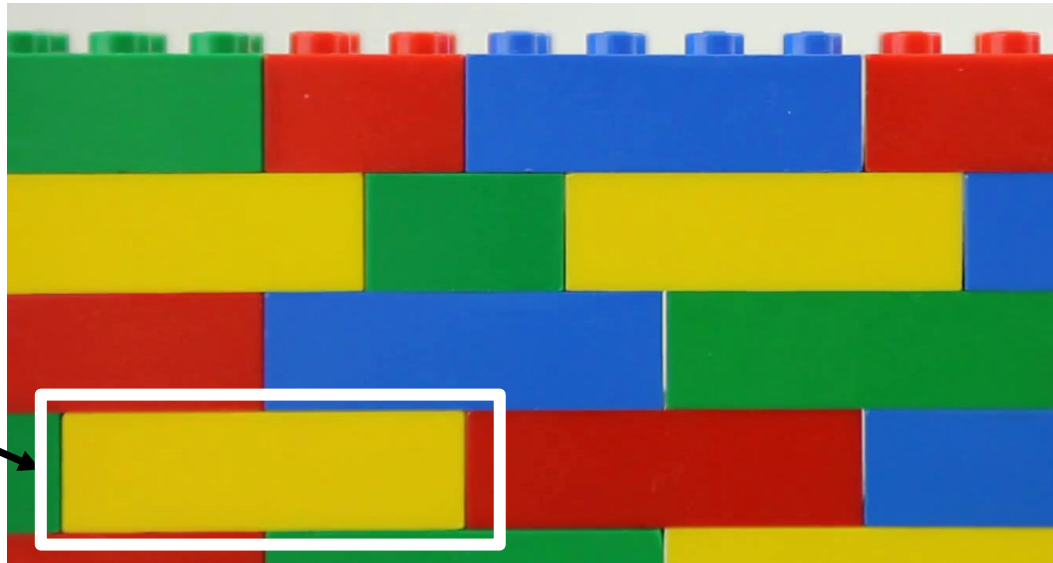
Scale (sec)	Time Units	System	World (theory)
10^7	months		SOCIAL BAND
10^6	weeks		
10^5	days		
10^4	hours	Task	RATIONAL BAND
10^3	10 min	Task	
10^2	minutes	Task	
10^1	10 sec	Unit task	COGNITIVE BAND
10^0	1 sec	Operations	
10^{-1}	100 ms	Deliberate act	
10^{-2}	10 ms	Neural circuit	BIOLOGICAL BAND
10^{-3}	1 ms	Neuron	
10^{-4}	100 μ s	Organelle	

Cognitive band

- Actions are performed on the order of seconds
- Cognitive operations
→ Perception, memory, attention allocation



What I typically study



Roadmap for today

- Allan Newell and Herbert Simon: personal scientific heroes
 - Time scales of human action
- Levels of integration of cognitive theory
 - The algorithmic level: fact learning
- Models of decision making
 - How can these cognitive models be useful? Effect of time pressure
 - Other uses (in the context of open societies)

Levels of integration of cognitive theory:

*The importance of psychology for interactive
Artificial Intelligence*

Or: Why should YOU care??



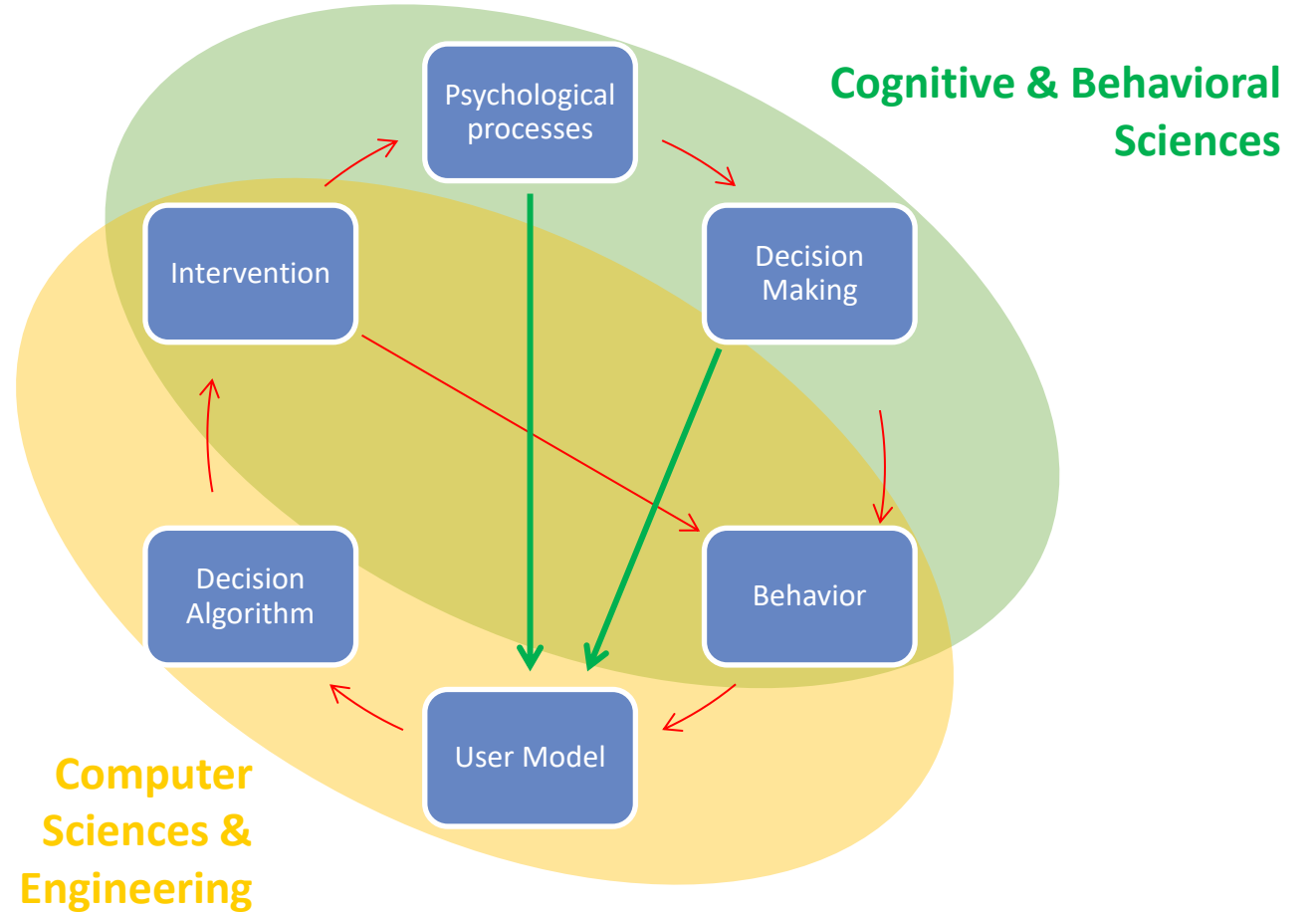
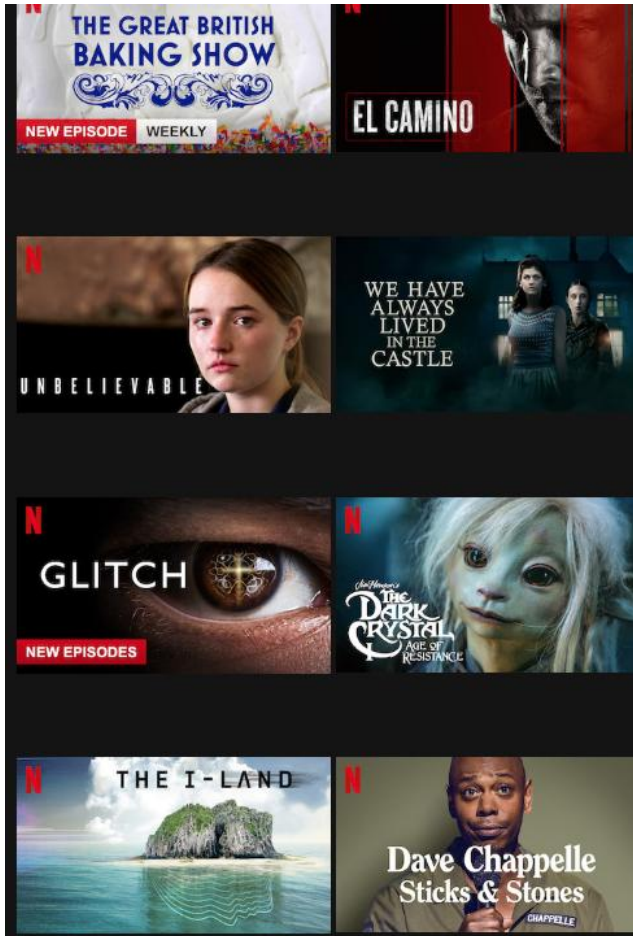
NETFLIX

Interactive AI



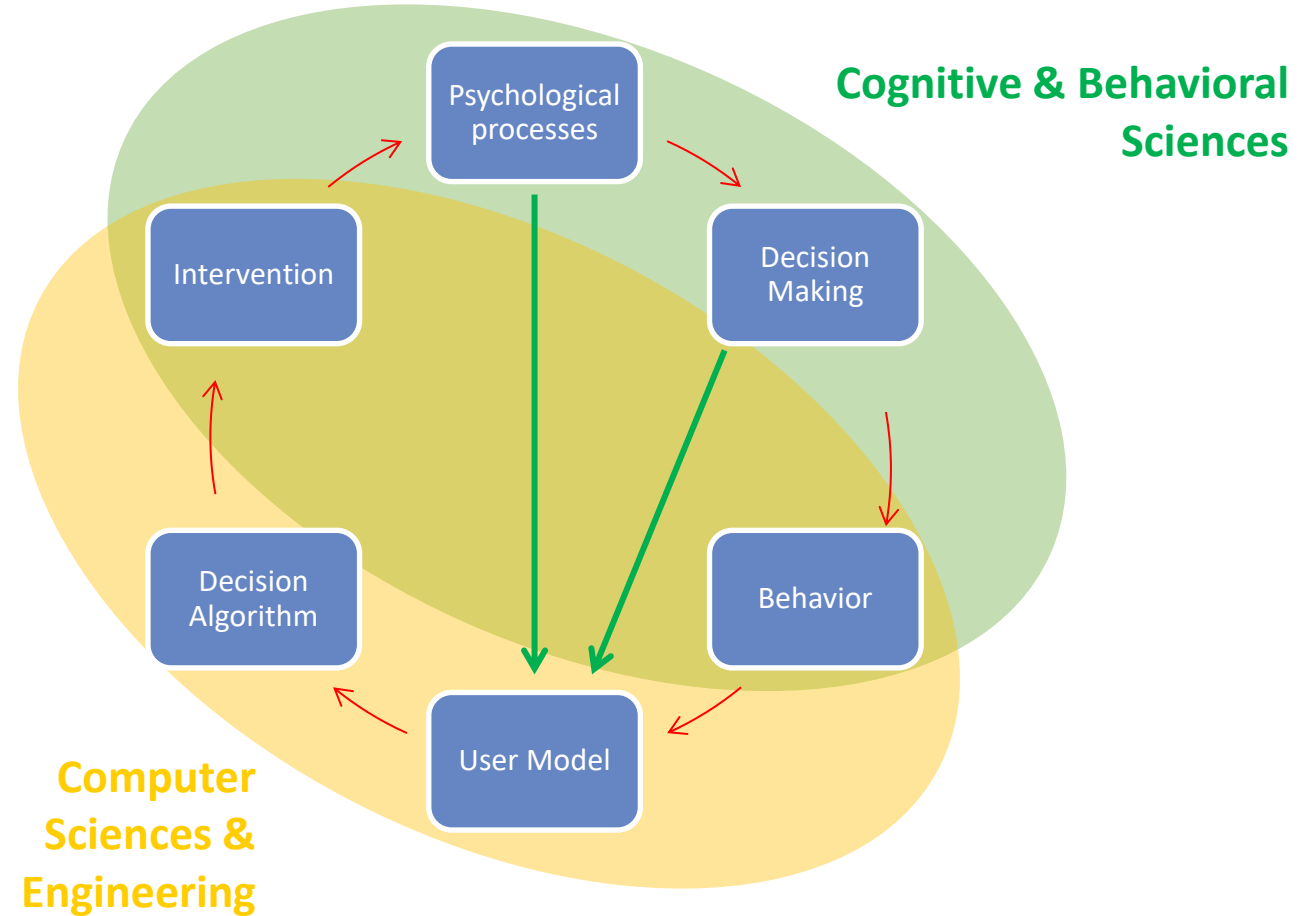
Remember me?





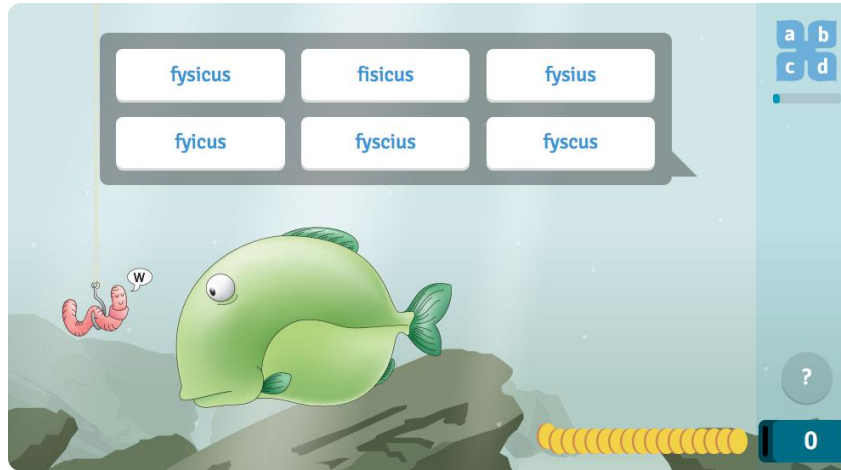
Levels of integration of cognitive theory

- Anecdotal
- **general knowledge of cognition**
- Computational
- **considering hypothesized computations that individuals perform**
- Algorithmic
- **considering hypothesized cognitive processes**



Example: smart fact learning

- Support of fact learning (speed up) using interactive AI



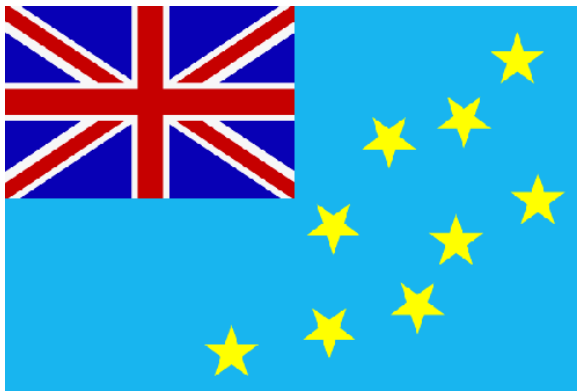
Example: smart fact learning



Kiribati



Saint Lucia



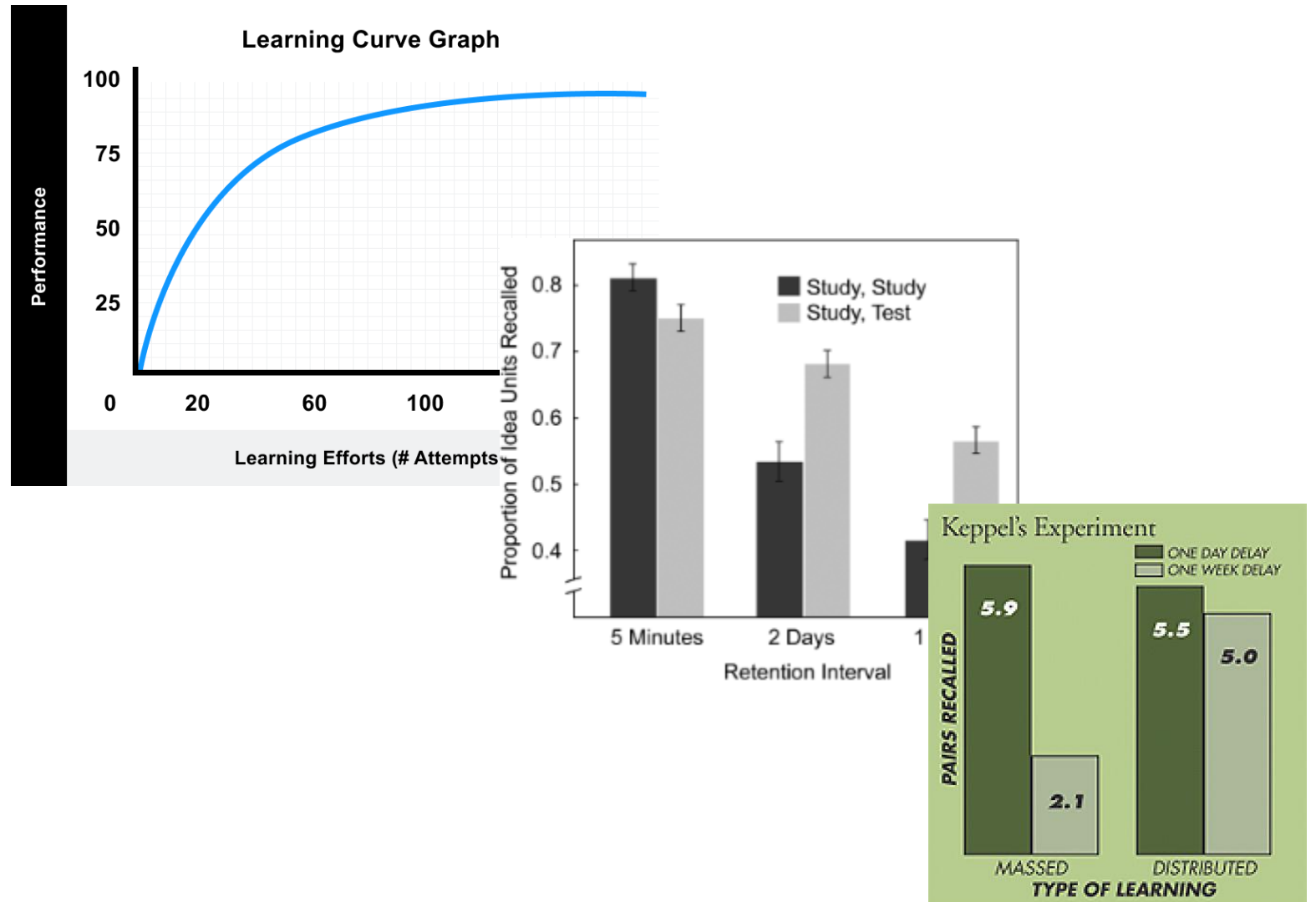
Tuvalu



Åland

Example: smart fact learning

- Support of fact learning (speed up) using interactive AI
- Clearly it is important to understand cognitive processes (how does learning work?)
- **Practice effect**
- **Testing effect**
- **Spacing effect**

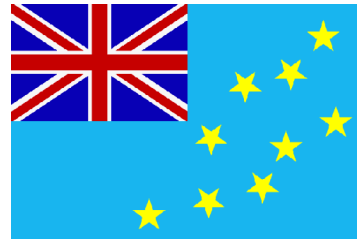


The anecdotal level of integration

- Flash cards (Leitner, 1972)
- Learn the facts:
- **If you are correct, move to next deck**
- **Incorrect, move back**
- **Frequency of learning decks decreases**
- Repetition → Practice effect
- Recall → Testing effect
- Mix → Spacing effect



Kiribati



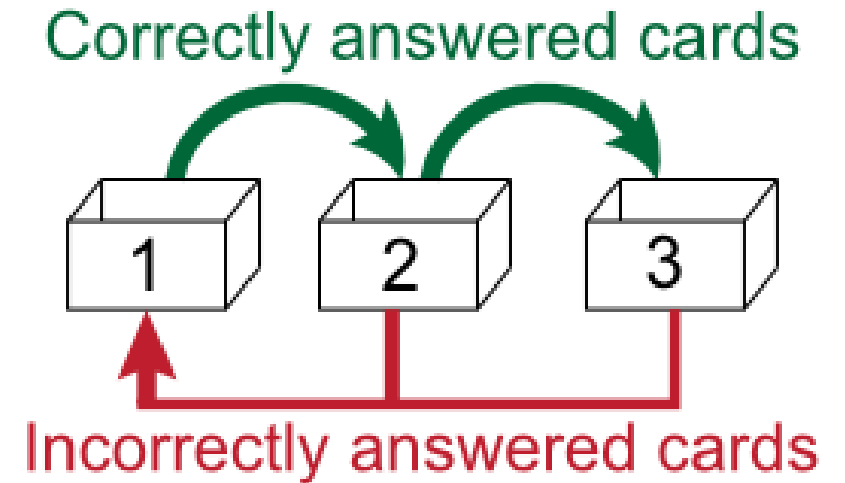
Tuvalu



Saint Lucia



Åland



The algorithmic level of integration

1. Formalize/model how memory recall works
2. Use that formal model in a edutech application

Anderson & Schooler (1991). Reflections of the environment in memory, Psychological Science (and previous work):

Rational Analysis of Memory

Forgetting if $p[\text{Activation}] * \text{Gain} < \text{Cost}$?

p: "Need" probability

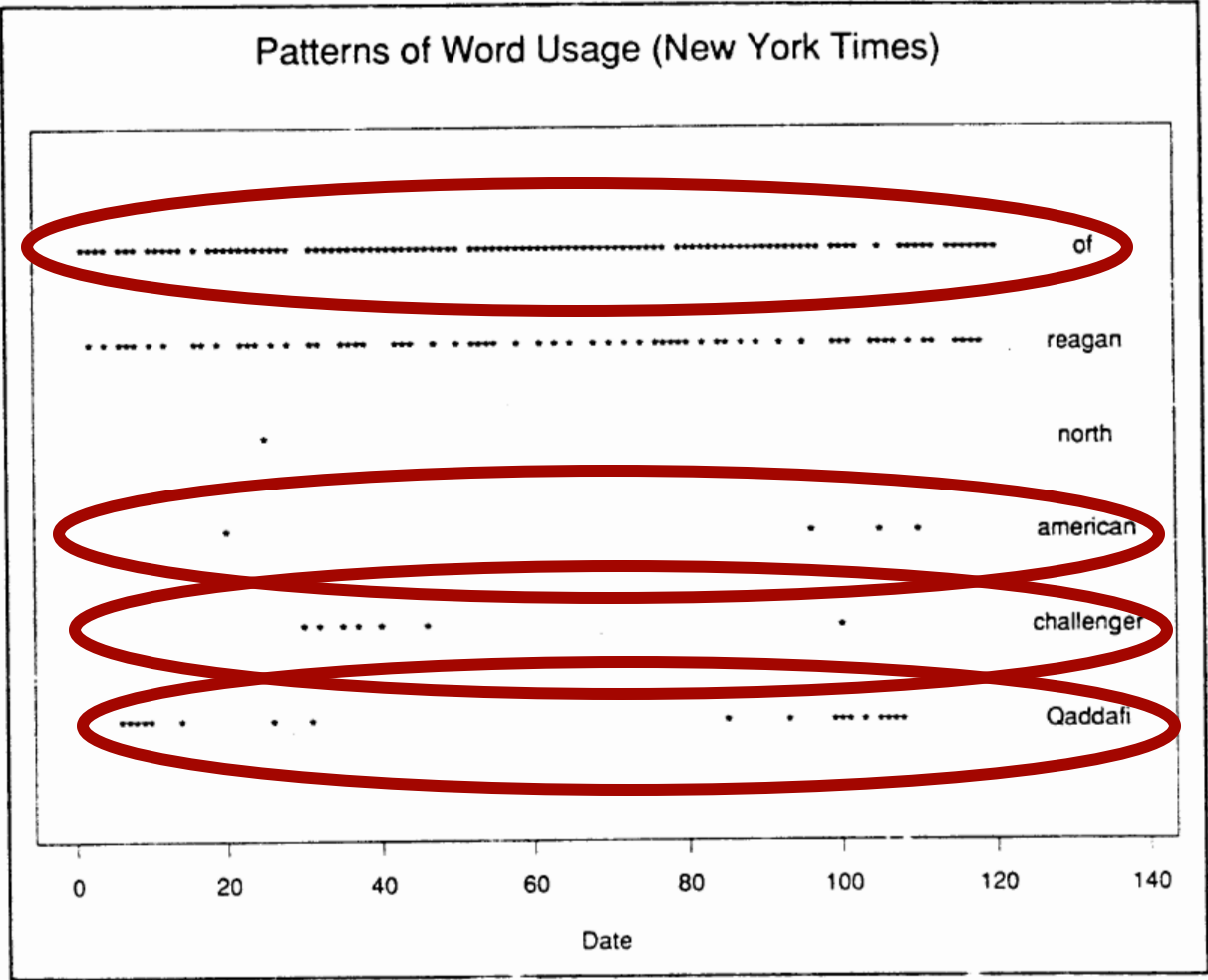
How likely is it that a particular piece of information is required in the near future?

Activation function

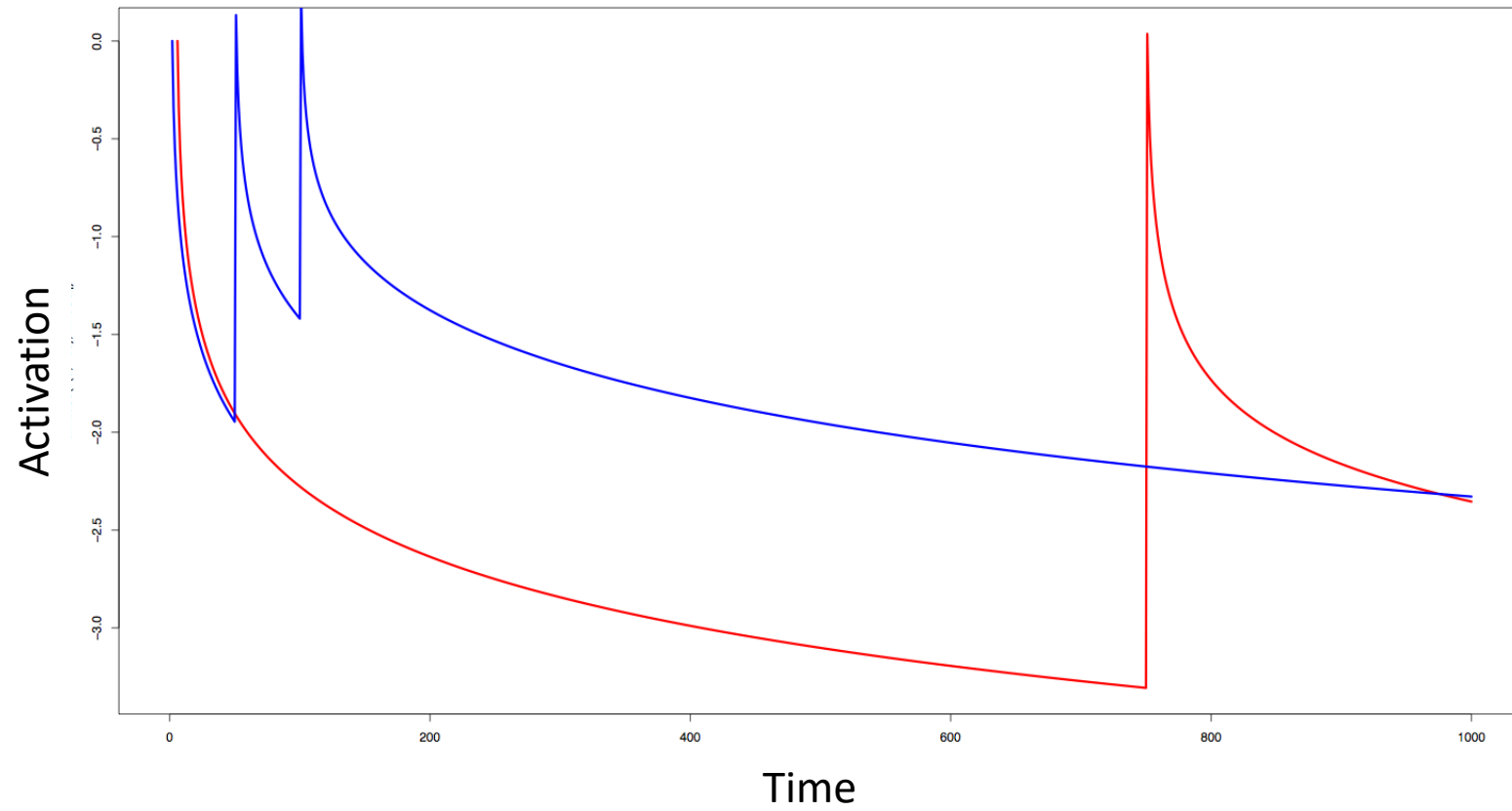
$$B_i = \ln\left(\prod_{j=1}^n t_j^{-d}\right)$$

- Activation (B)
- Frequency (n)
- Recency (t_j)

- Decay (d)

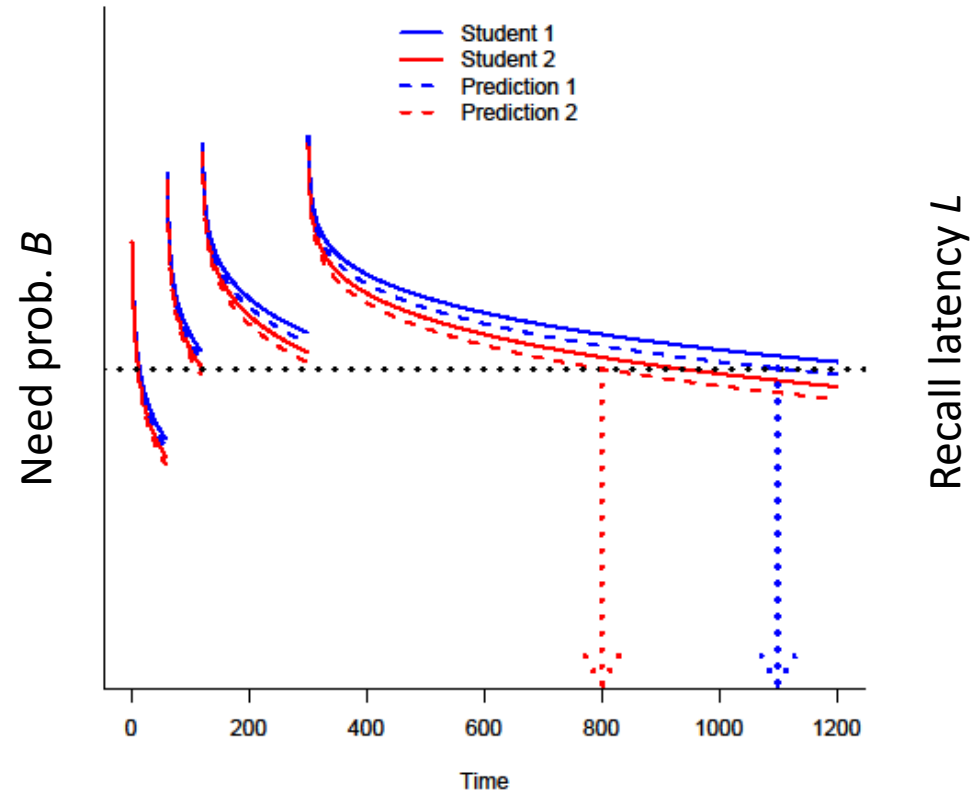


Frequency vs recency



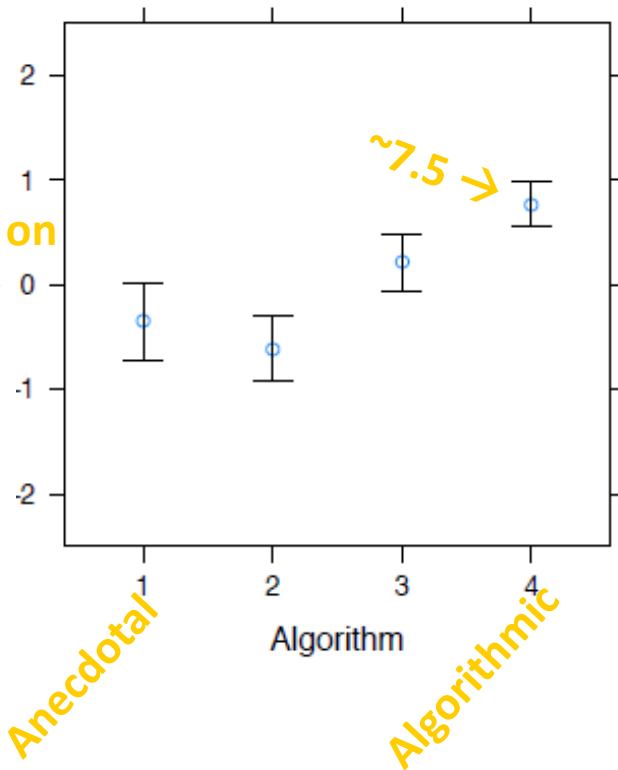
The algorithmic level of integration

- Rational analysis of memory (Anderson & Schooler, 1991)
- Cognitive model
- **indicating the “need probability” of items in memory**
- By student, predict recall time
- Iteratively calibrate model parameters

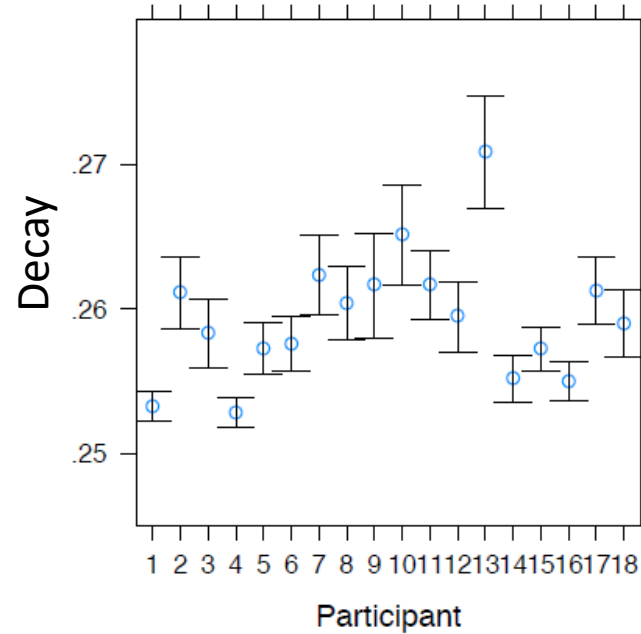


Level comparison

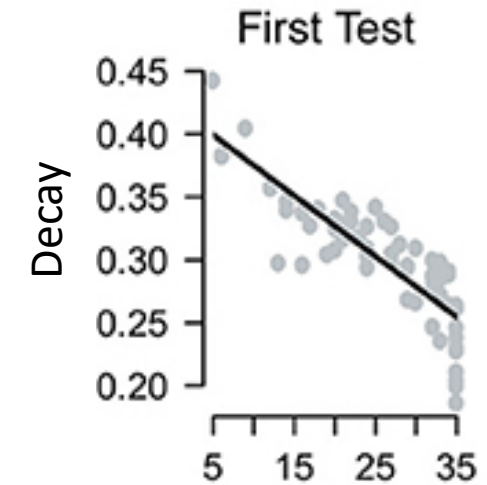
Performance increase



Explanation of individual differences



Prediction of success



Roadmap for today

- Allan Newell and Herbert Simon: personal scientific heroes
 - Time scales of human action
- Levels of integration of cognitive theory
 - The algorithmic level: fact learning
- Models of decision making
 - How can these cognitive models be useful? Effect of time pressure
 - Other uses (in the context of open societies)



**Utrecht
University**

Security in Open Societies

Bottom-up Initiatives
for Societal Change

Behaviour and Institutions

**Democracy and
good governance**

Contesting Governance

Futures of Democracy

The Transactional State
as an Institution for Good

Gender, Diversity and Global Justice

In/Equality

Equality and diversity

Future of Work

Open Cities

Transitions and well-being

Fair Transitions

Longtermism and Institutional Change

Markets and Corporations

Openness challenged: the university at risk?

**IOS-Themes:
15 platforms**

Roadmap for today

- Allan Newell and Herbert Simon: personal scientific heroes
 - Time scales of human action
- Levels of integration of cognitive theory
 - The algorithmic level: fact learning
- Models of decision making
 - How can these cognitive models be useful? Effect of time pressure
 - Other uses (in the context of open societies)

What's the first decision you had to make today?

How many decisions have you made today?

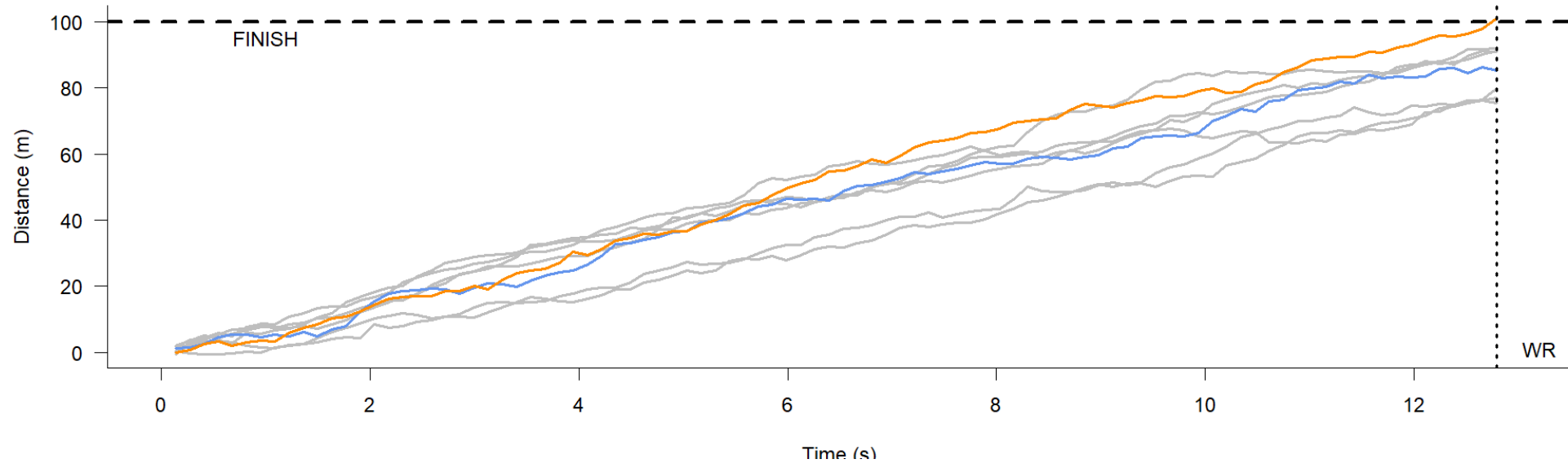
Noisy decision making

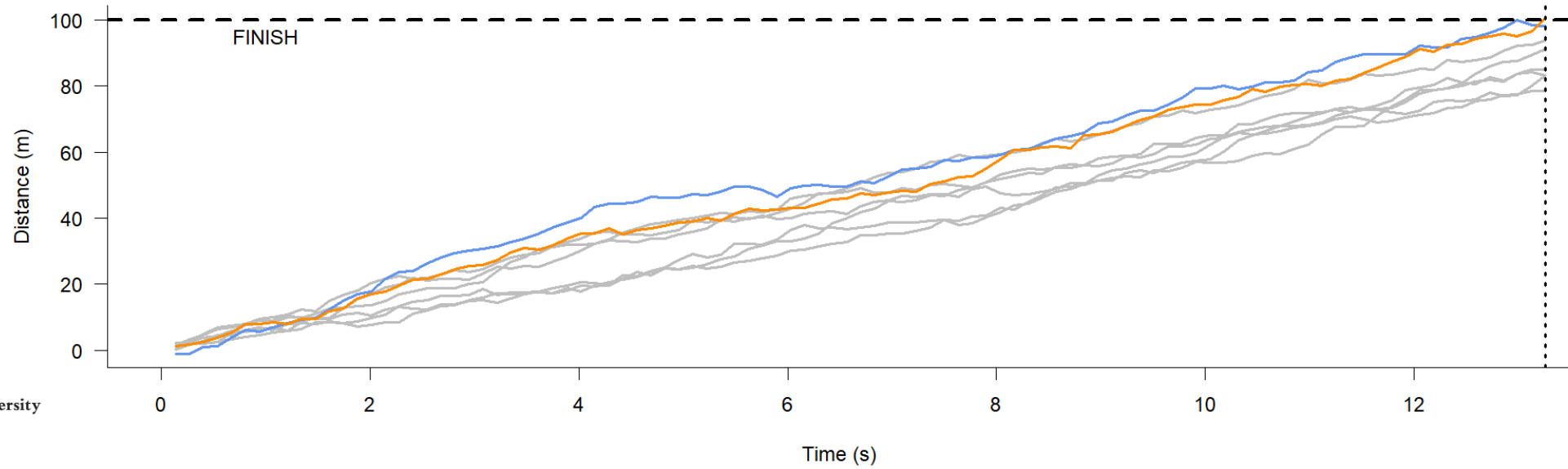
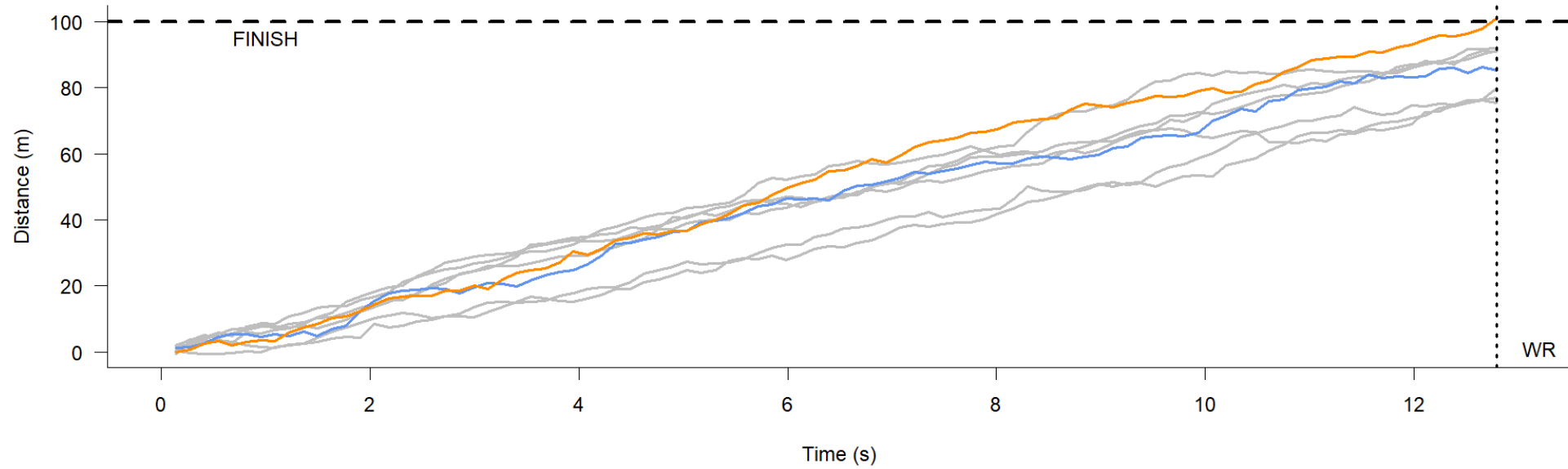
- Actions are performed on the order of seconds
- Cognitive operations
→ Perception, memory, attention allocation, but also *simple decision making*





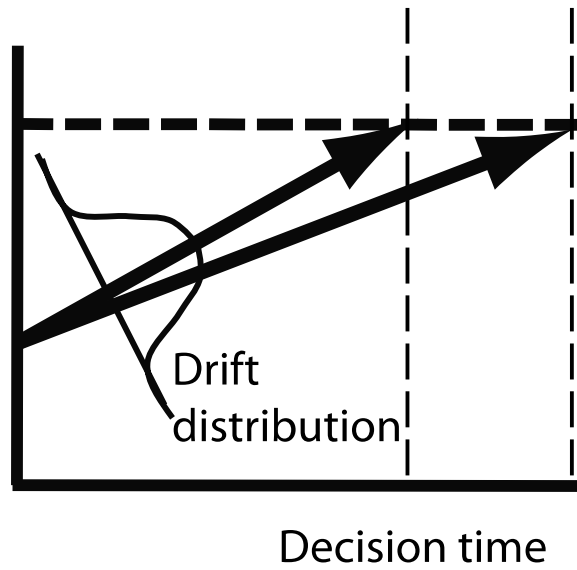
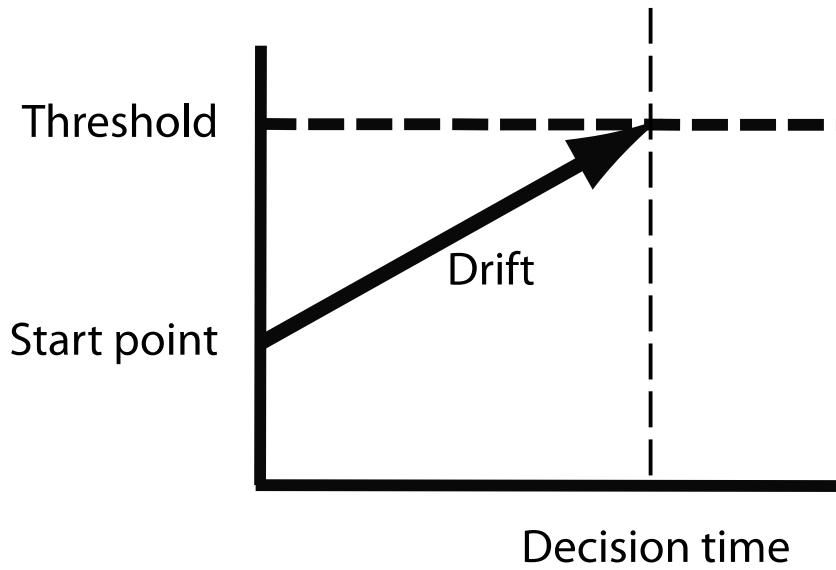


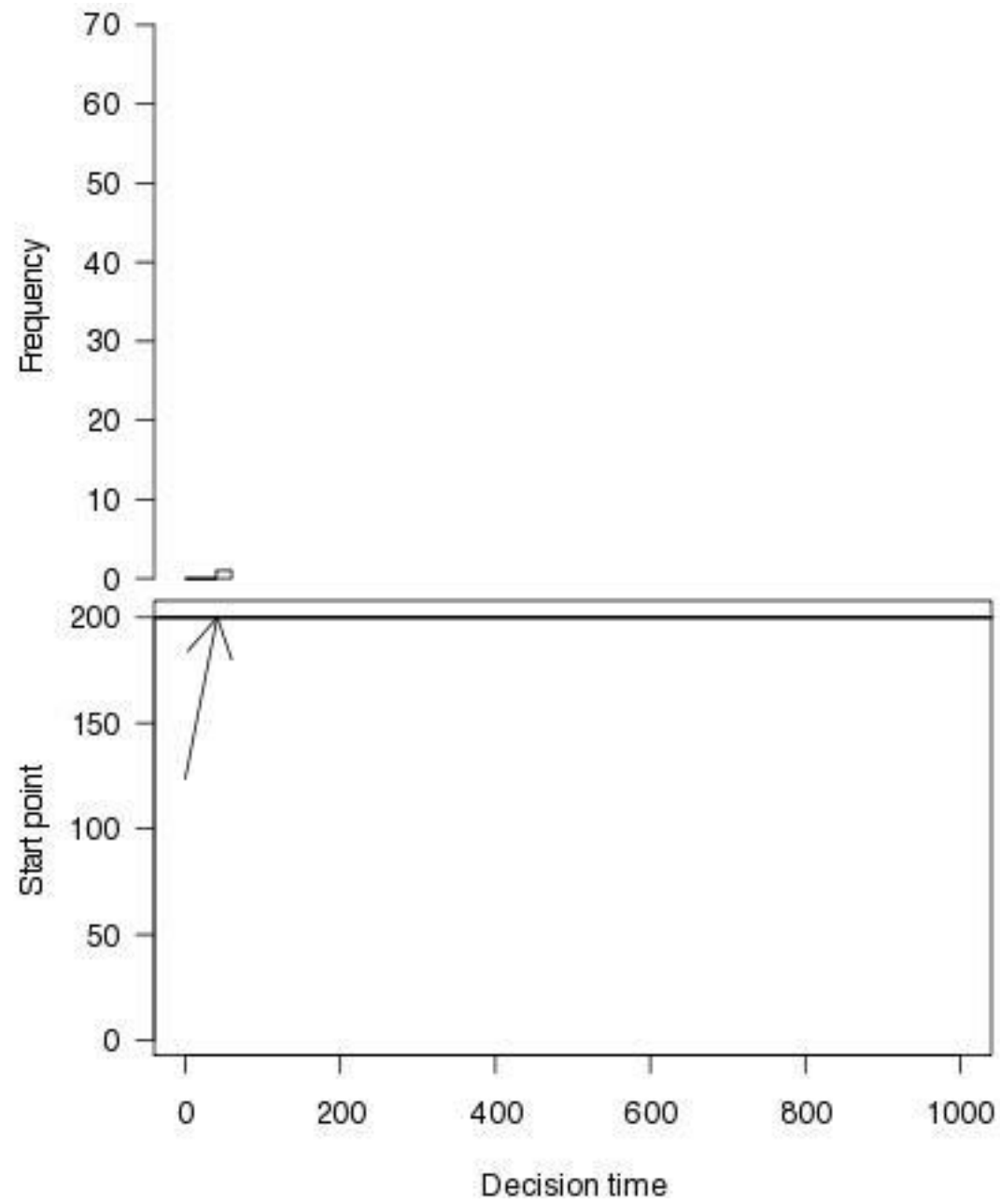


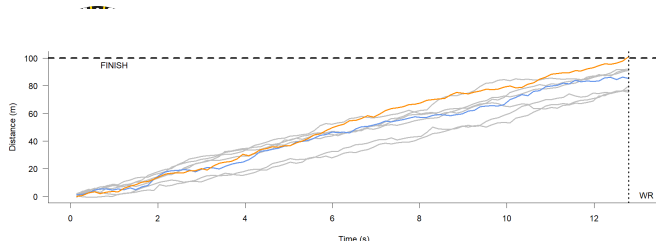




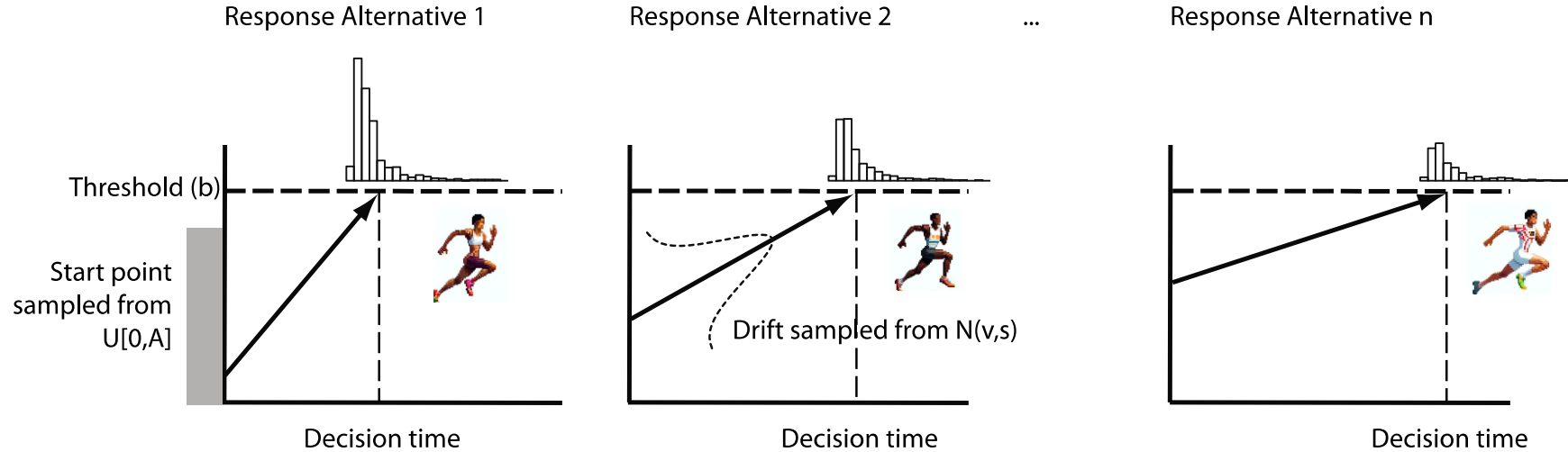
Linear Ballistic Accumulator







Linear Ballistic Accumulator



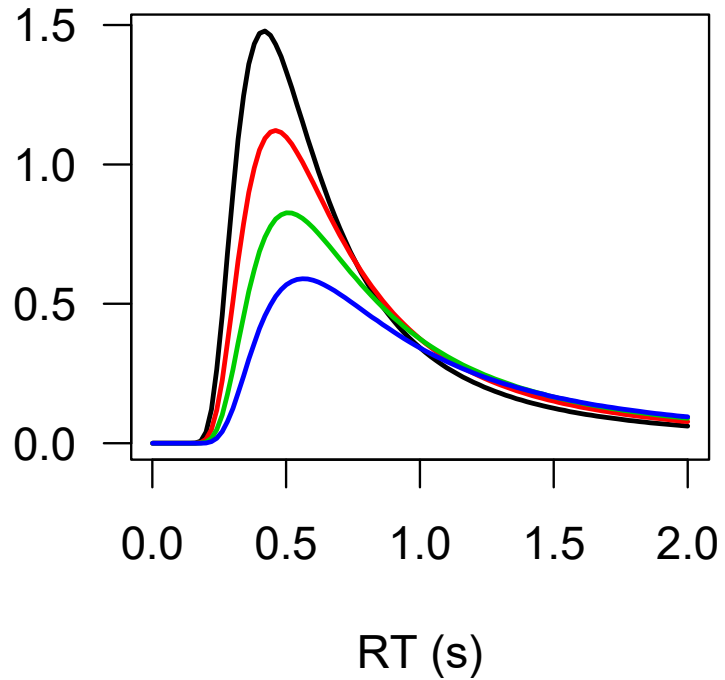
Every option has its own accumulator

Fastest accumulator wins (that option gets selected)

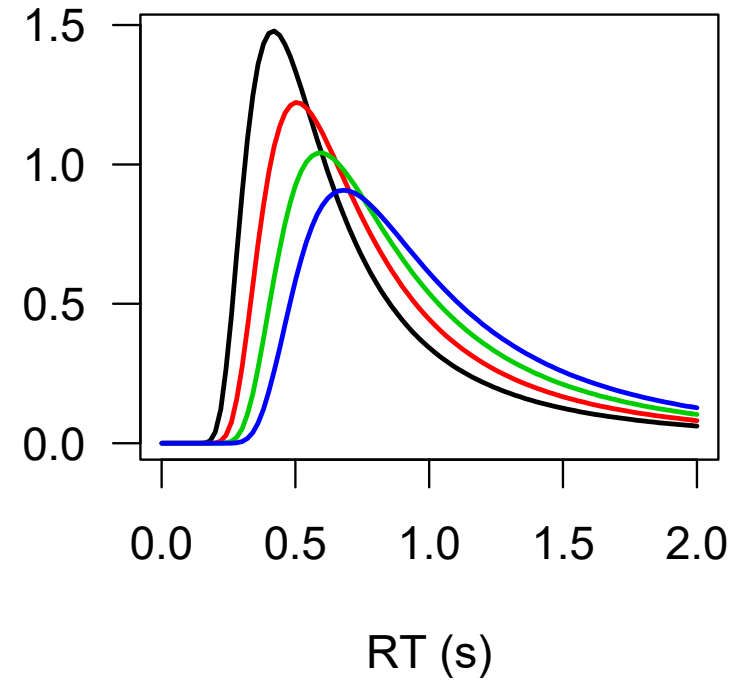
$$f_i(t) = \frac{1}{A} \left[-v_i \Phi \left(\frac{b - A - tv_i}{ts} \right) + s \phi \left(\frac{b - A - tv_i}{ts} \right) + v_i \Phi \left(\frac{b - tv_i}{ts} \right) - s \phi \left(\frac{b - tv_i}{ts} \right) \right]$$



Drift change



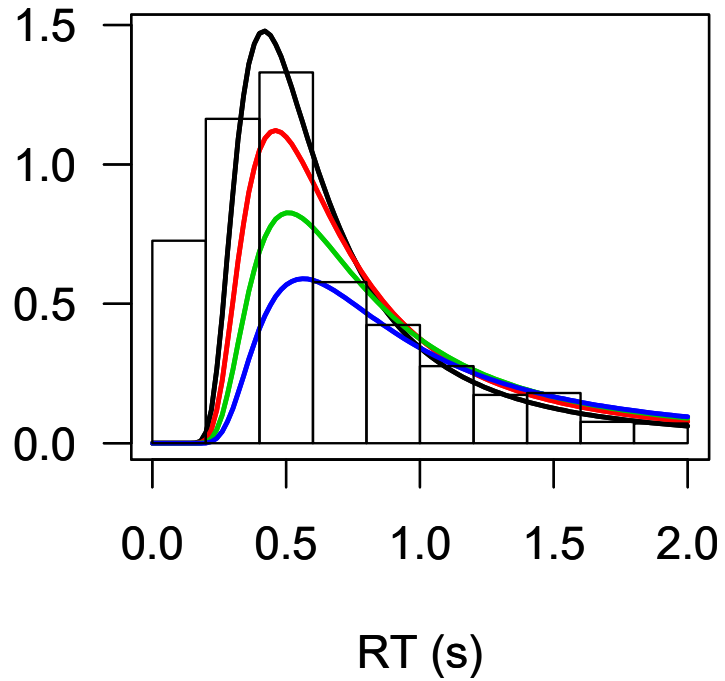
Threshold change



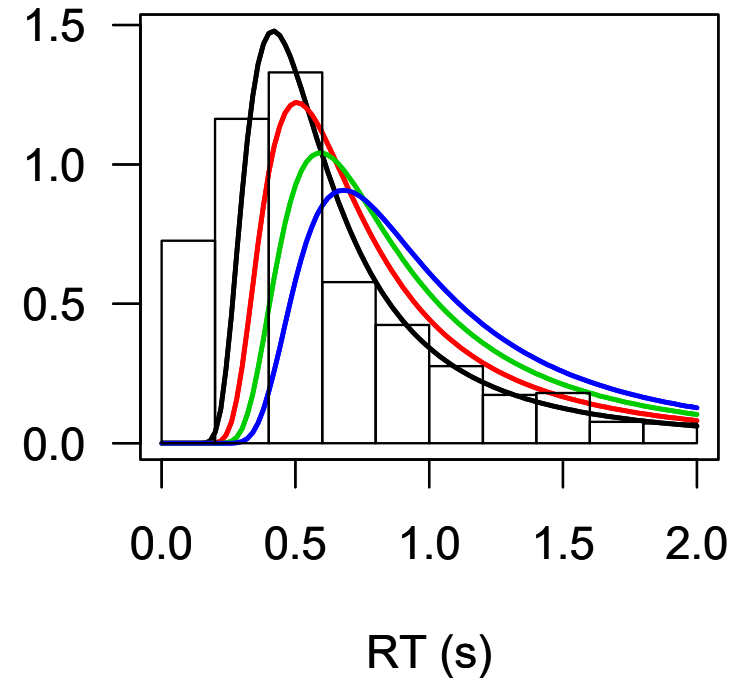
$$f_i(t) = \frac{1}{A} \left[-v_i \Phi \left(\frac{b - A - tv_i}{ts} \right) + s \phi \left(\frac{b - A - tv_i}{ts} \right) + v_i \Phi \left(\frac{b - tv_i}{ts} \right) - s \phi \left(\frac{b - tv_i}{ts} \right) \right]$$



Drift change



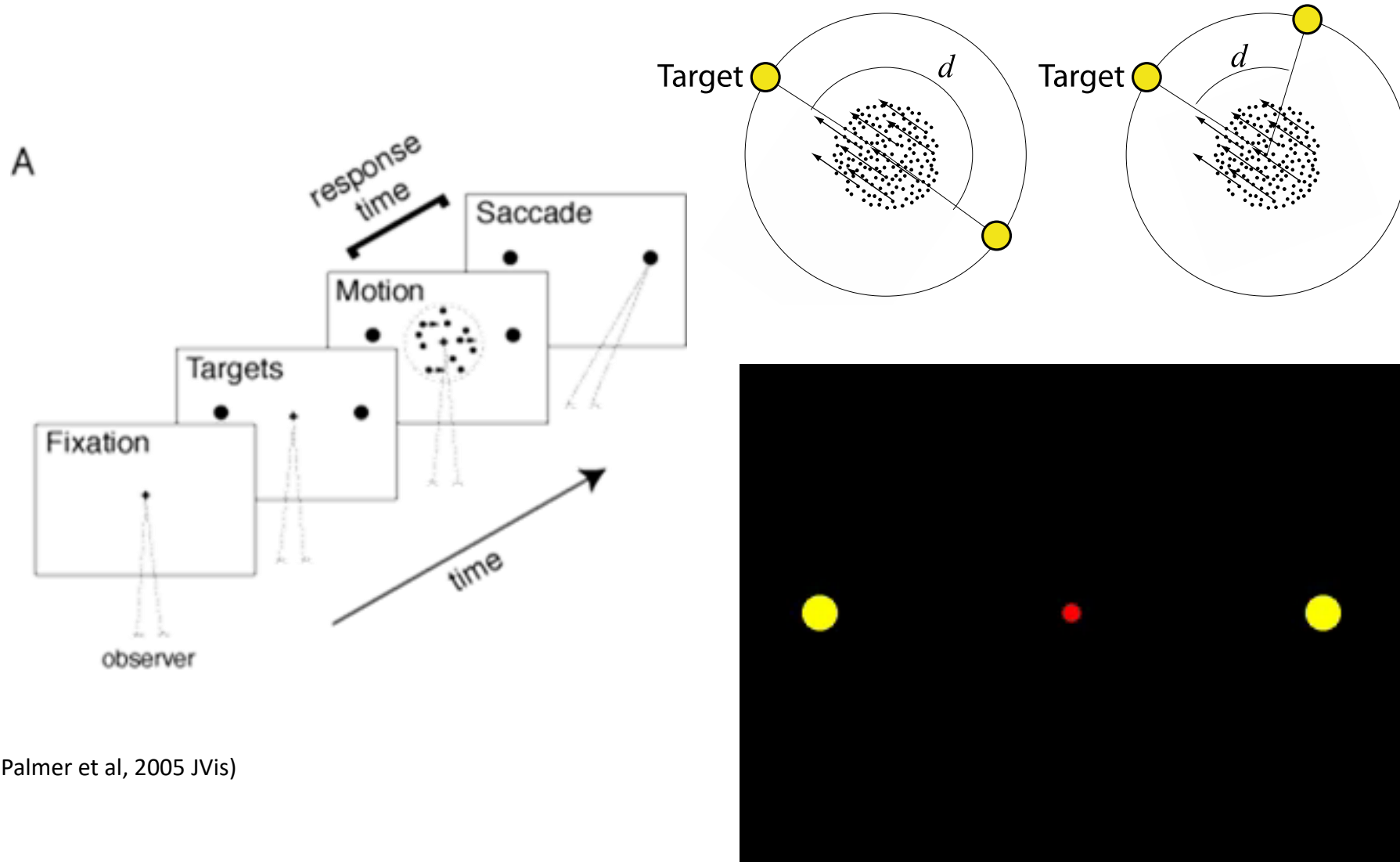
Threshold change



Lines: different model predictions
Histogram (bars): Observed data



Example: effect of target location in 2AFC

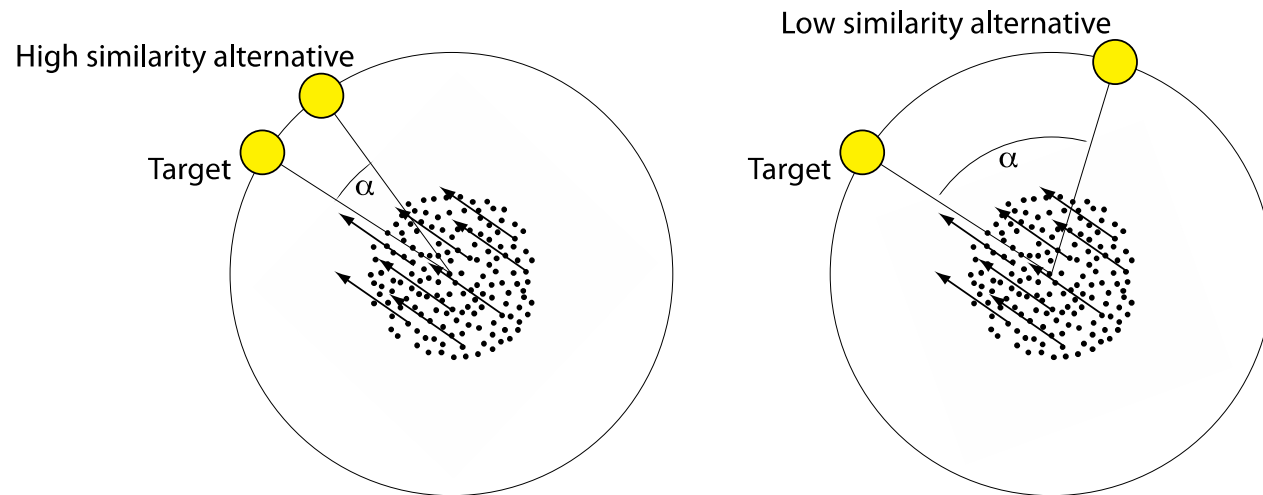


(Palmer et al, 2005 JVis)



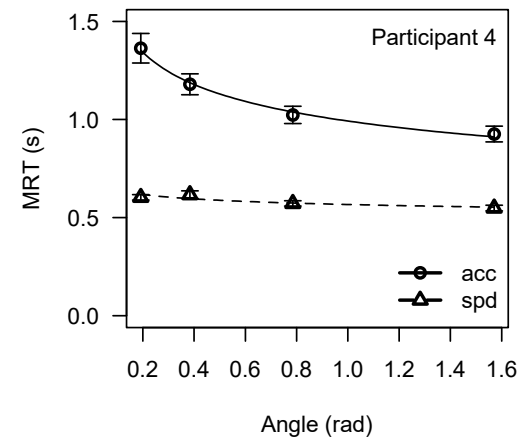
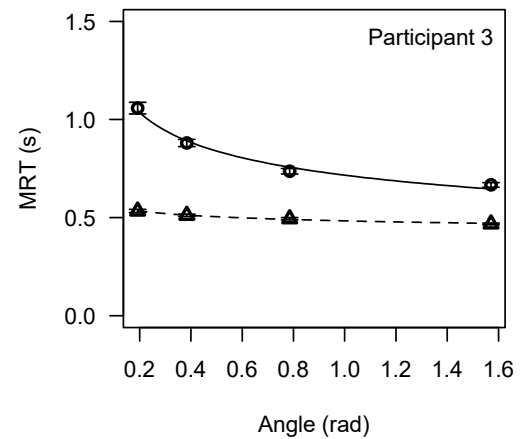
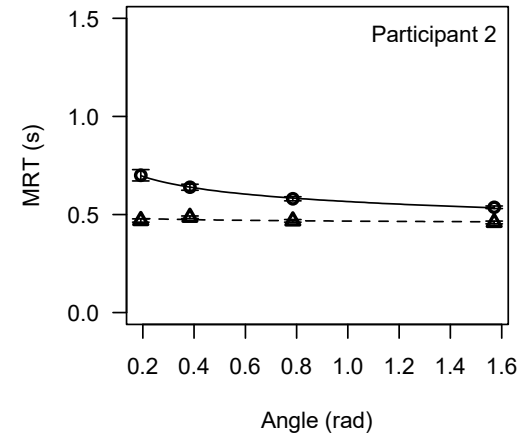
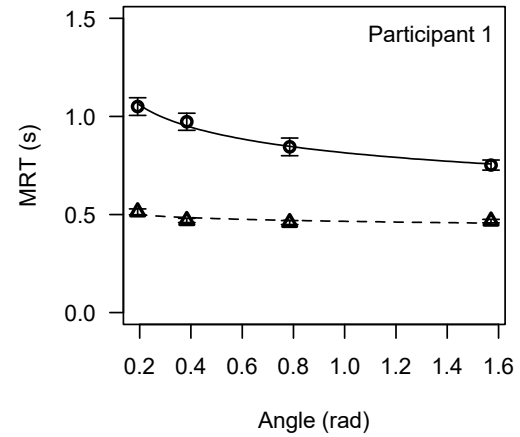
Example: effect of target location in 2AFC

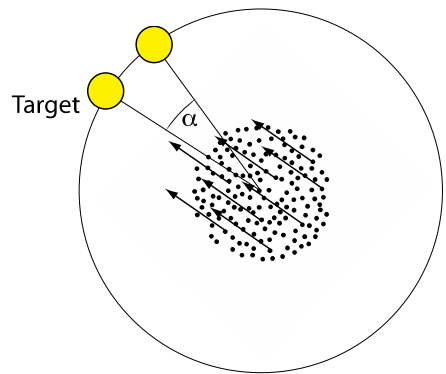
- Binary choice
- Two instructions (blocks): "Focus on speed" vs "Focus on Accuracy"
- Target location varied



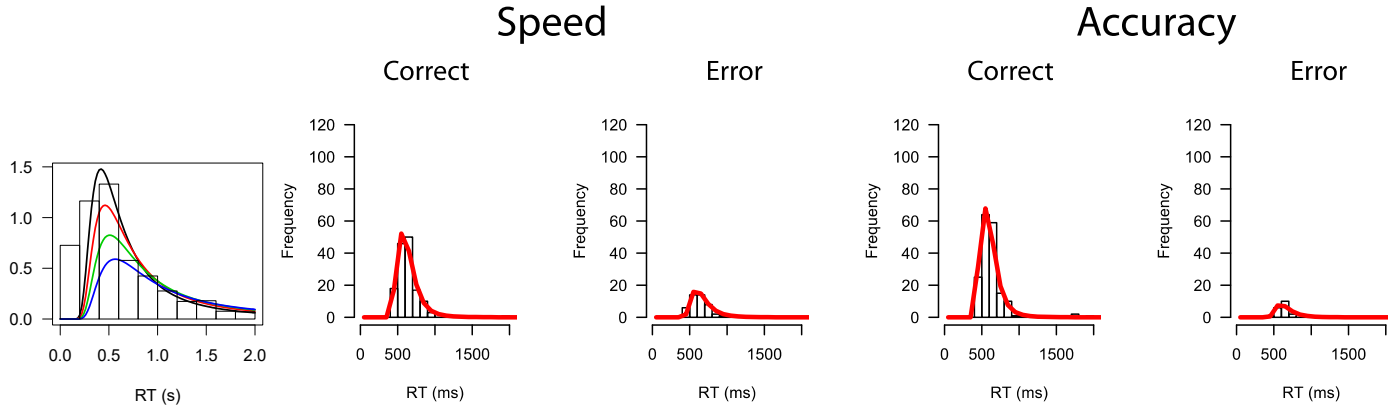


Mean response times

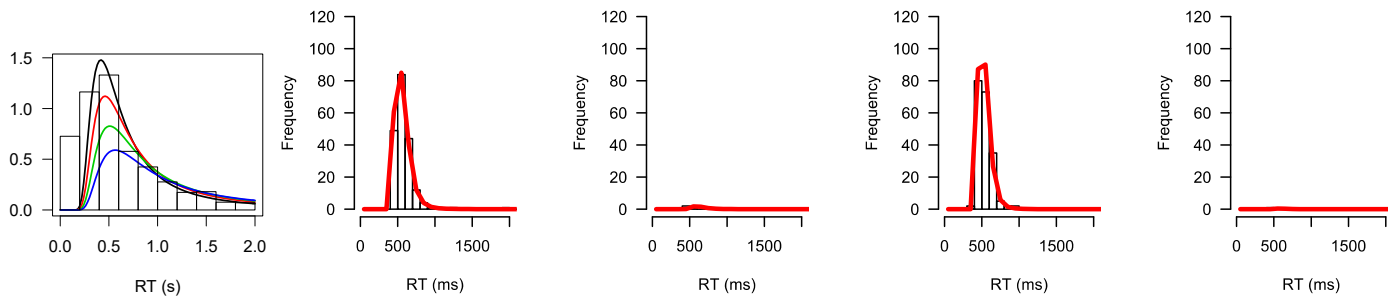




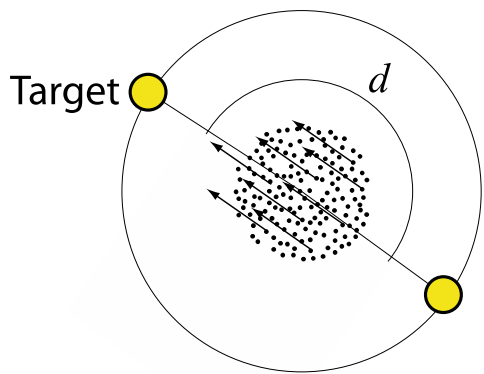
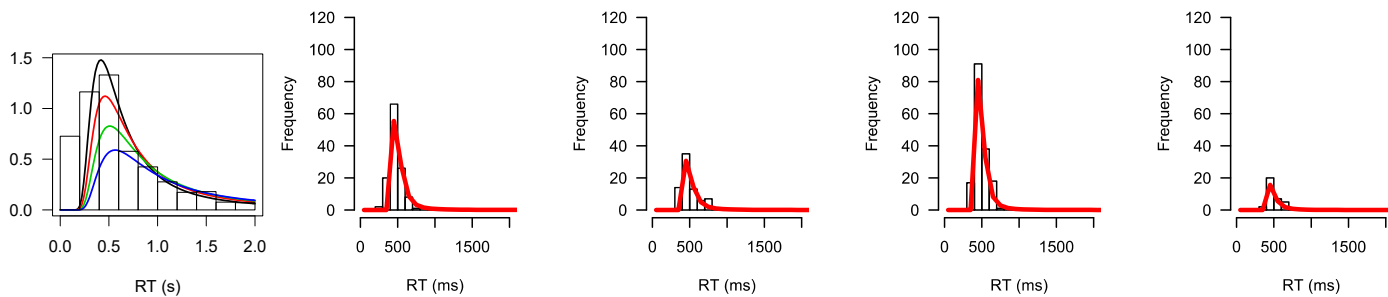
Difficult



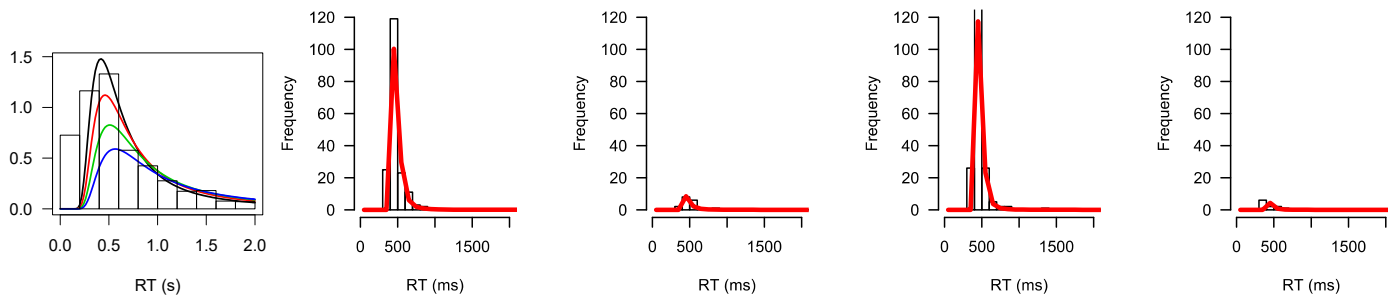
Medium



Easy



Very Easy



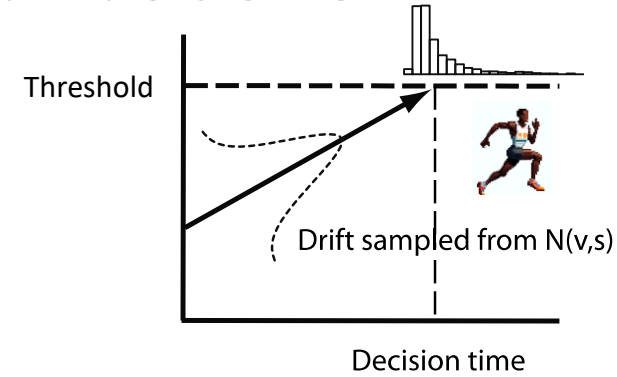


Accumulator model parameters

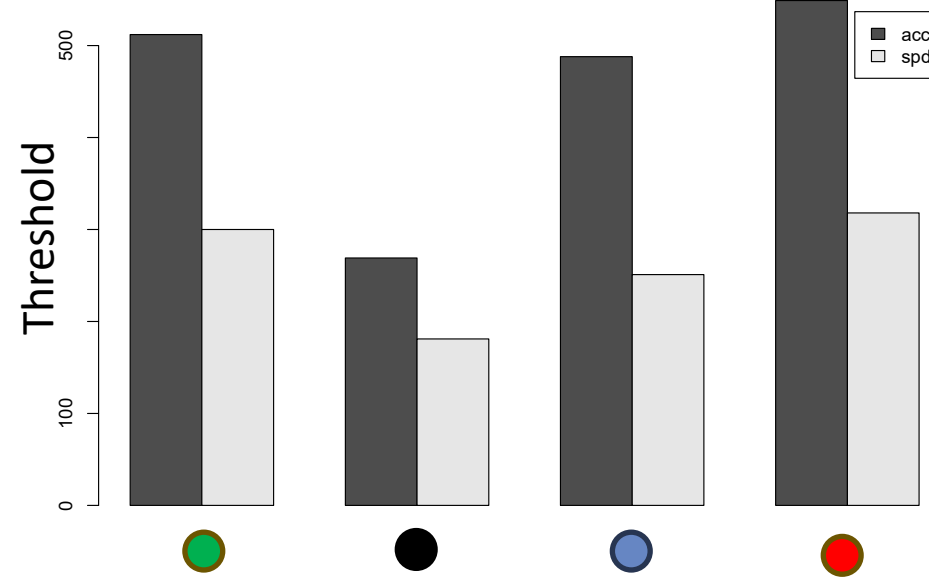
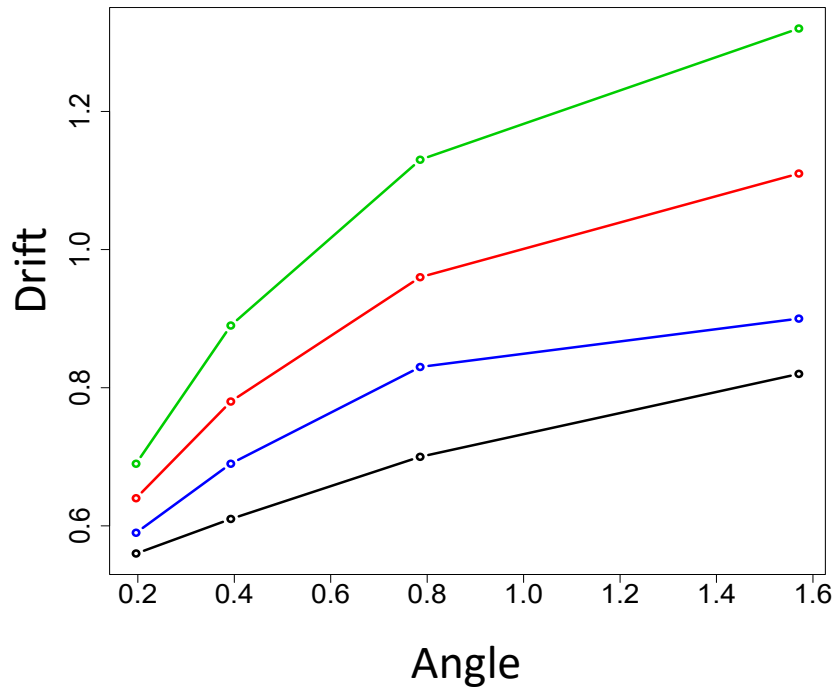
Drift rate influenced by **difficulty**

Threshold influenced by instruction (**speed/accuracy**)

Other parameters did not vary across conditions



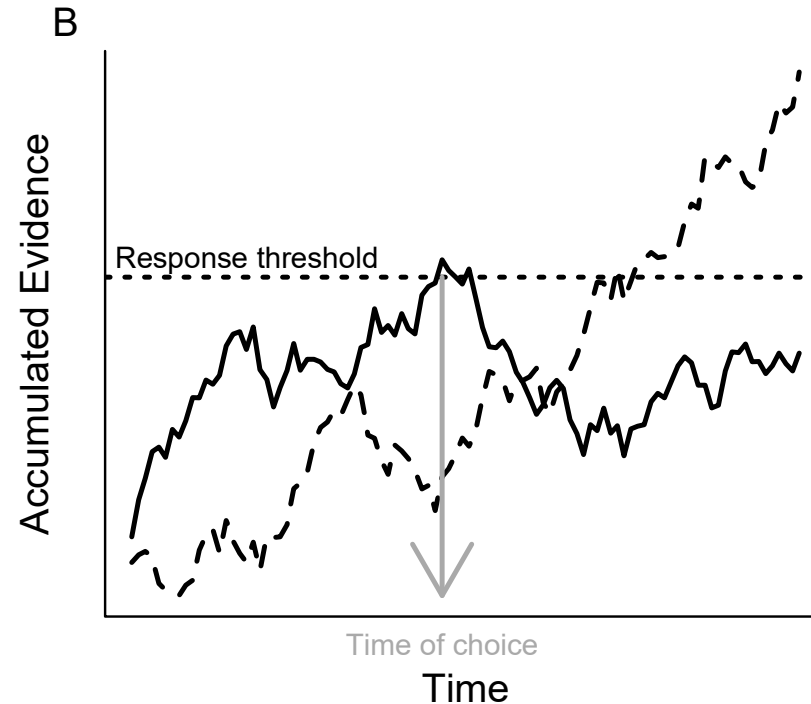
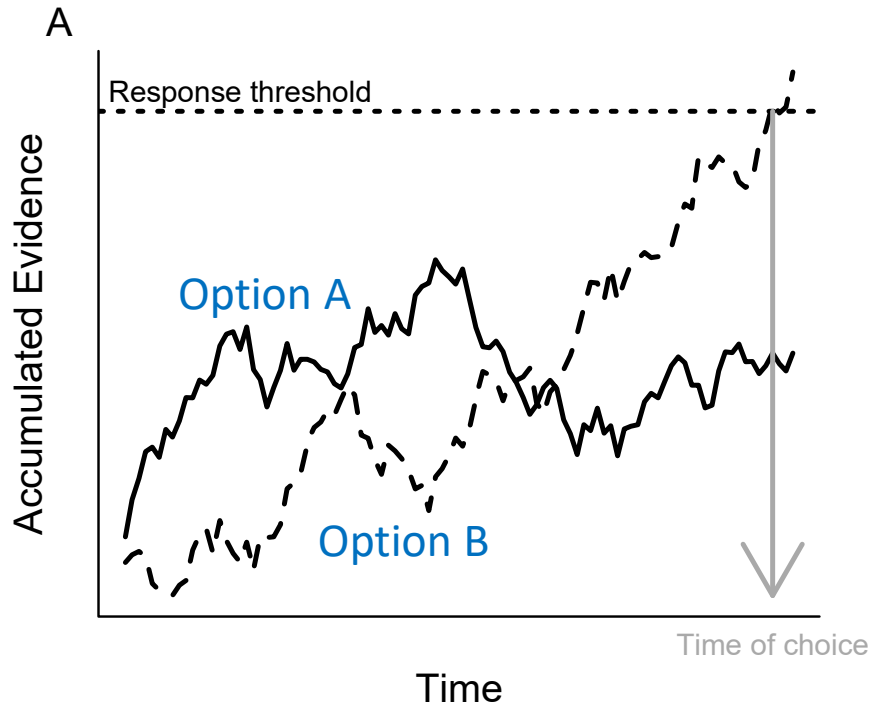
-  Participant 1
-  Participant 2
-  Participant 3
-  Participant 4





Speed instruction: Lower threshold

- Fast instructions lower thresholds
- Why does this actually affect accuracy?



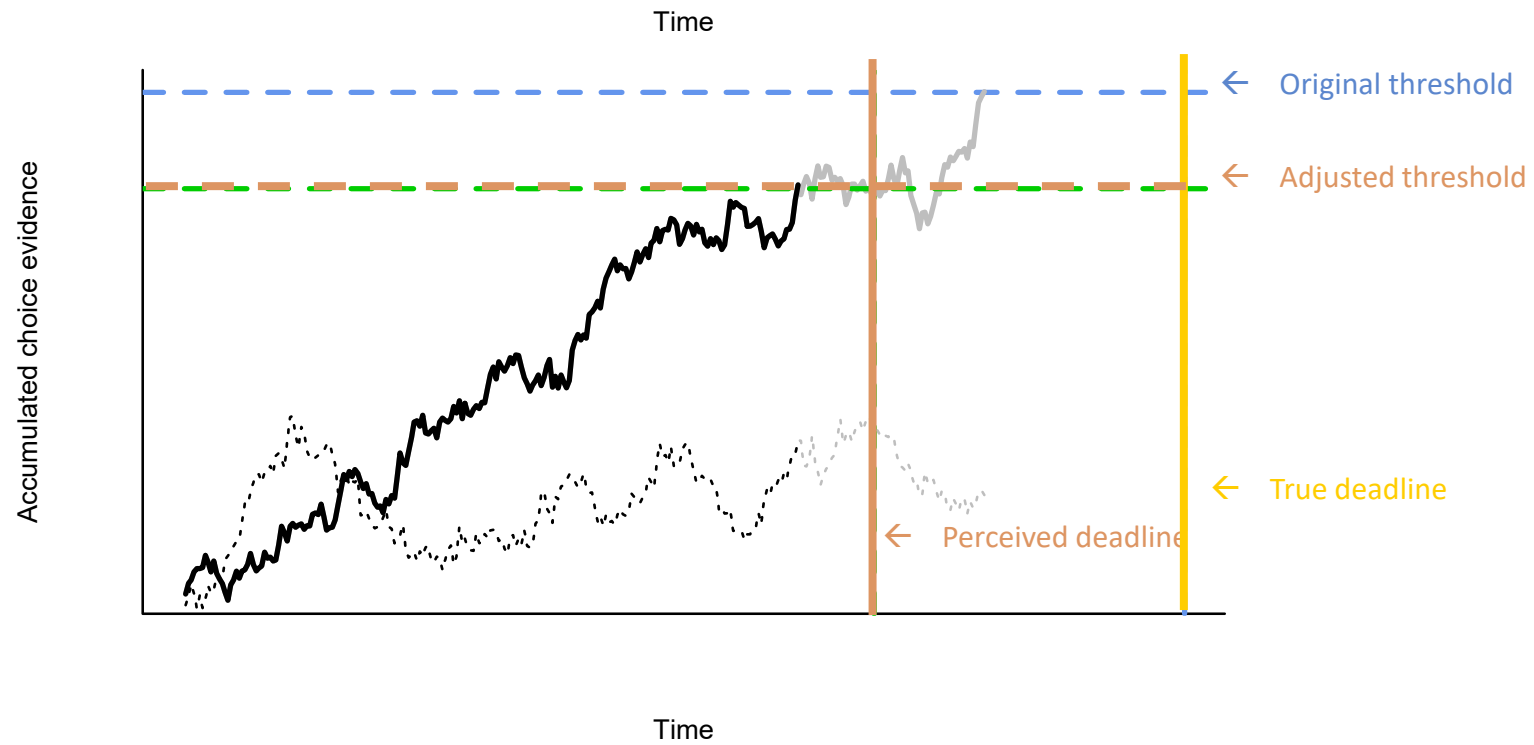
Roadmap for today

- Allan Newell and Herbert Simon: personal scientific heroes
 - Time scales of human action
- Levels of integration of cognitive theory
 - The algorithmic level: fact learning
- Models of decision making
 - How can these cognitive models be useful? Effect of time pressure
 - Other uses (in the context of open societies)

What is the effect of body temperature on decision making?



Hypothesis



The experiment

- Time production task + Expanded judgment task with deadline
- Core temperature manipulated through full body immersion

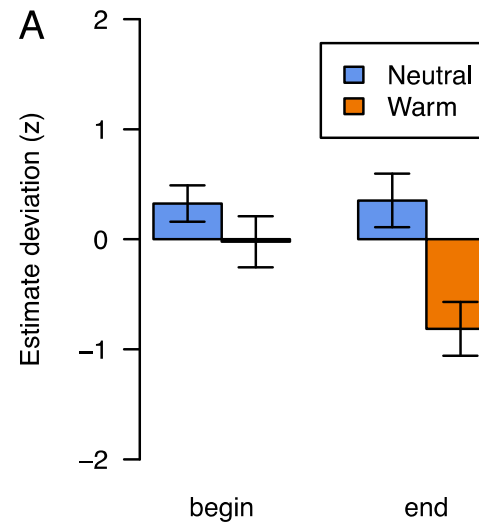
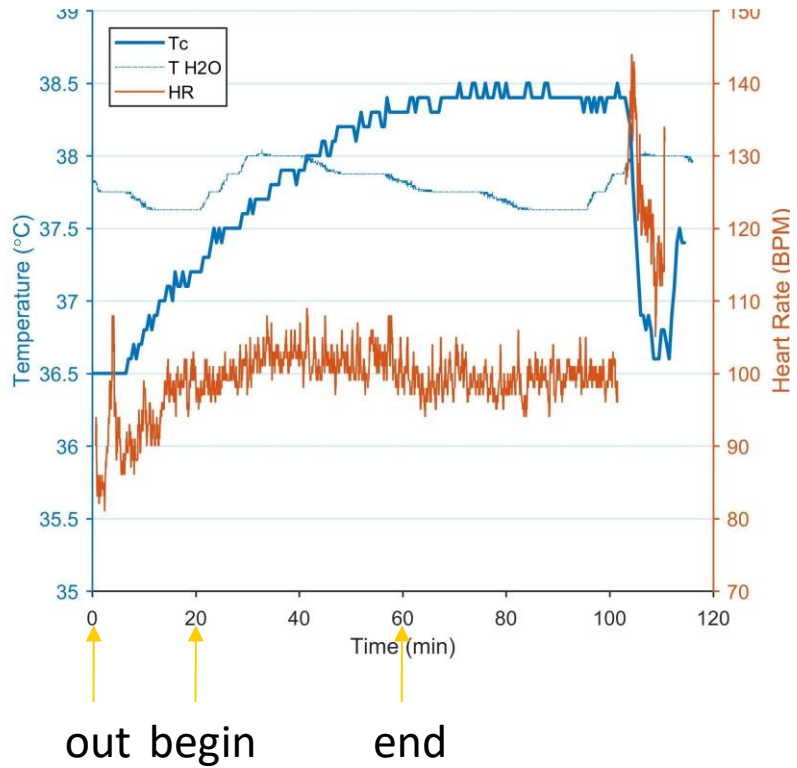


Neutral tub

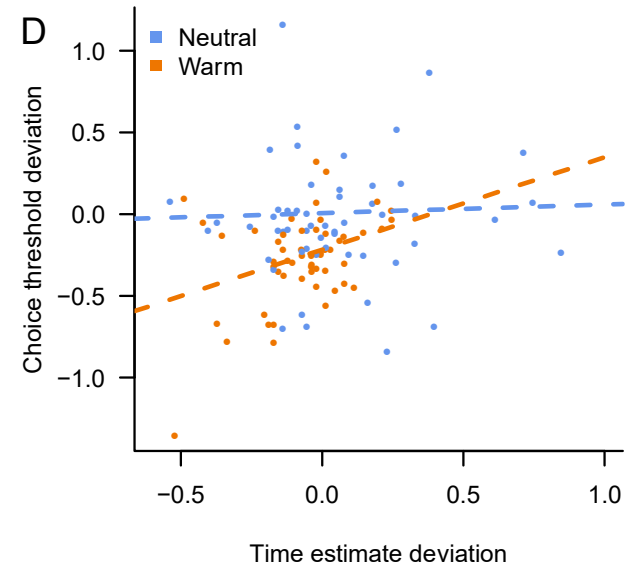
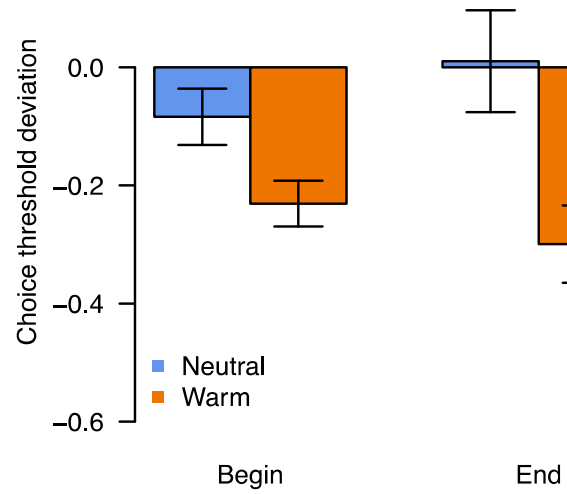


Hot tub

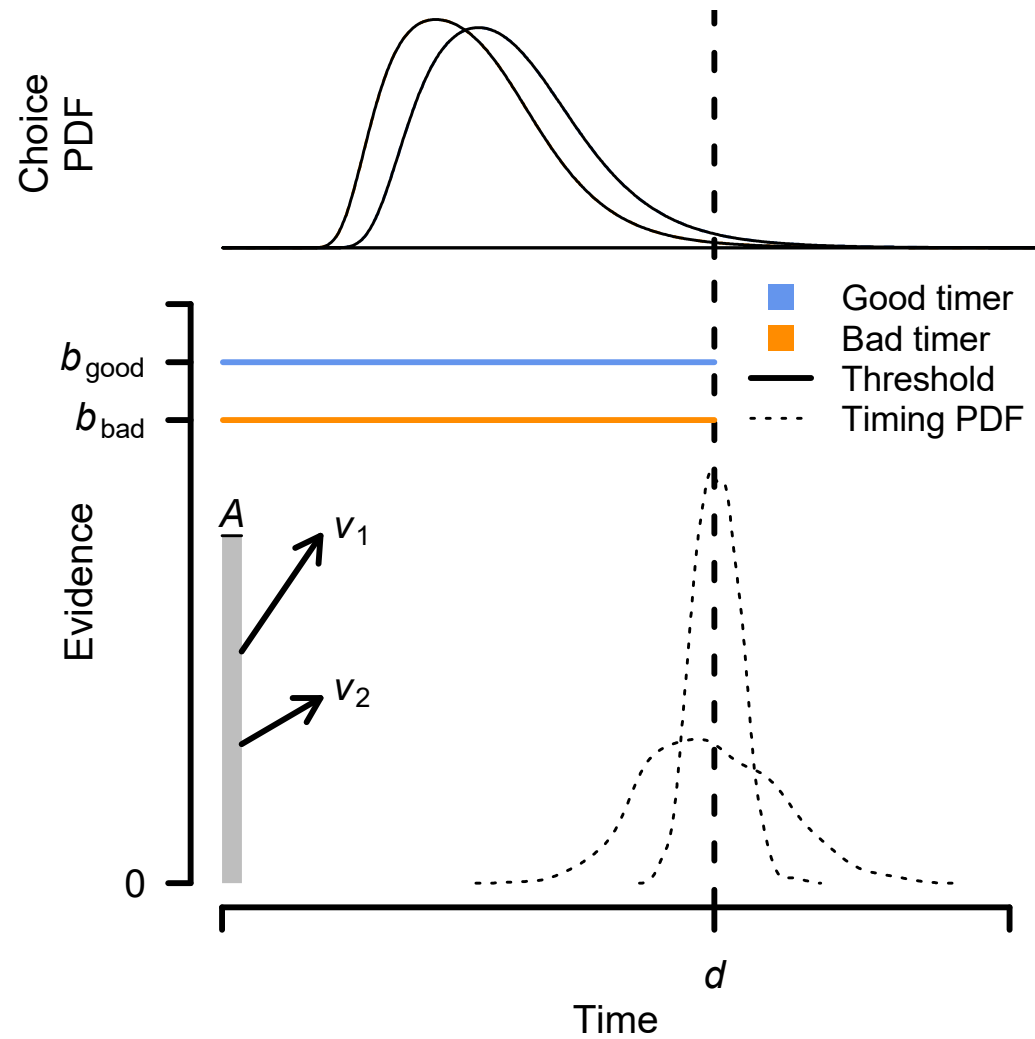
Results



LBA Results



Precise time perception influences speed-accuracy trade-off



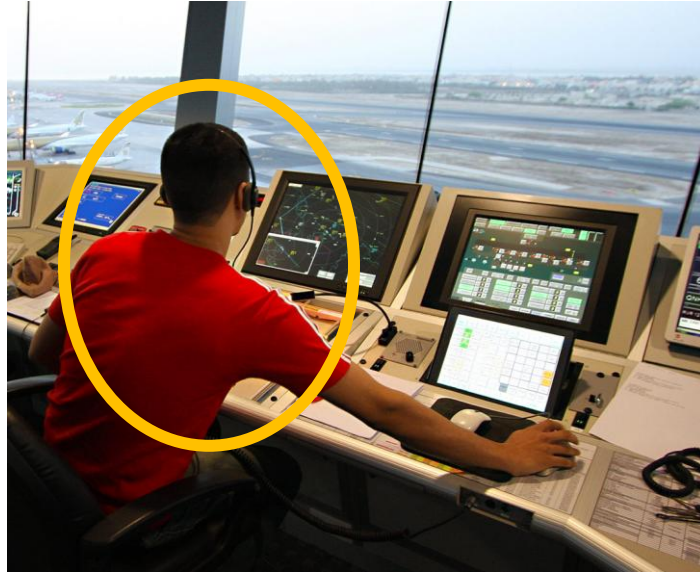
Roadmap for today

- Allan Newell and Herbert Simon: personal scientific heroes
 - Time scales of human action
- Levels of integration of cognitive theory
 - The algorithmic level: fact learning
- Models of decision making
 - How can these cognitive models be useful? Effect of time pressure
 - Other uses (in the context of open societies)

User Models

Applied in many domains,
but *not speeded actions*

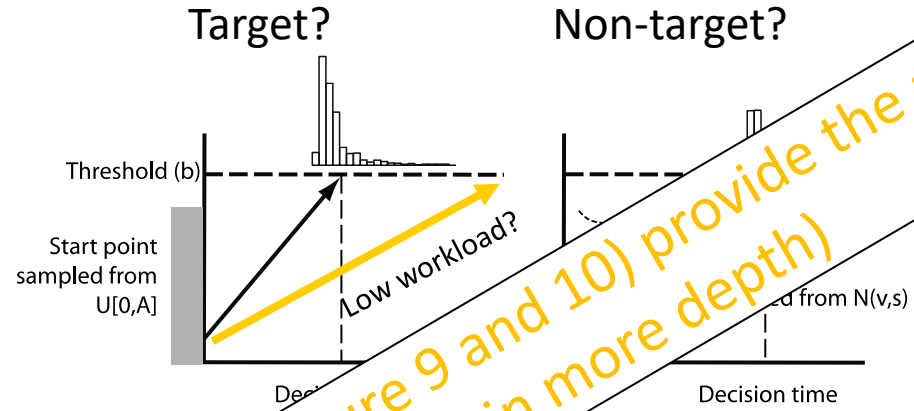
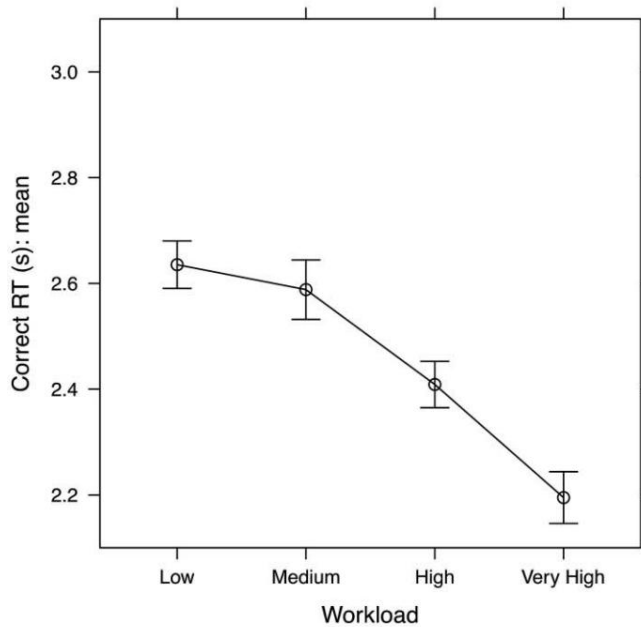
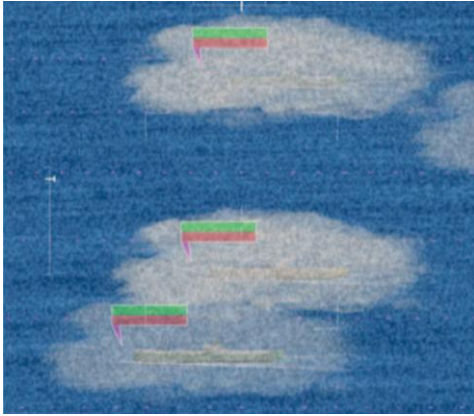
1. Many decisions/actions
2. Time pressure
3. (High stakes)



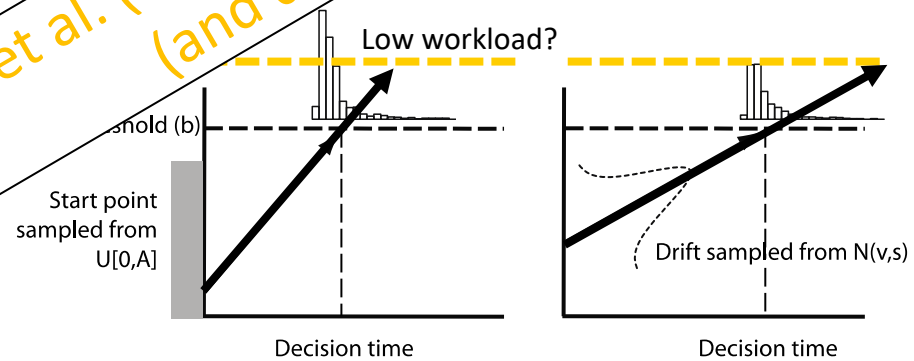
Palada et al. (2016) (see the student manual)



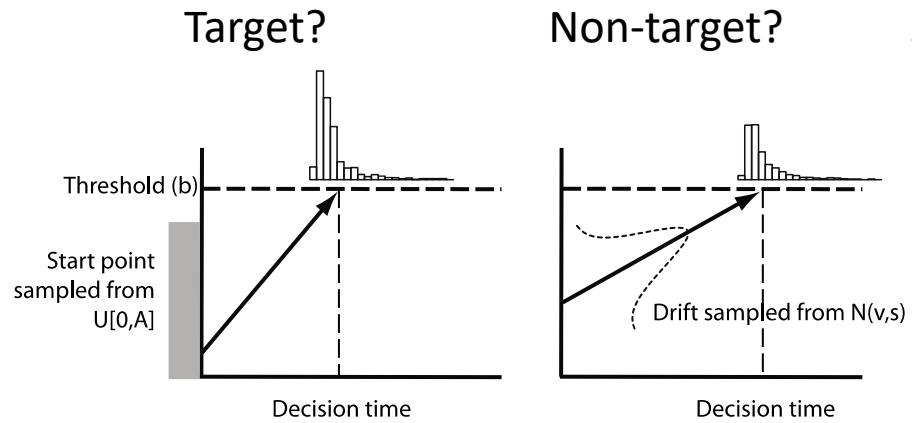
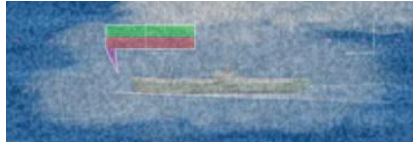
Modeling real-life tasks



Palada et al. (2016, Figure 9 and 10) provide the answer (and discusses in more depth)



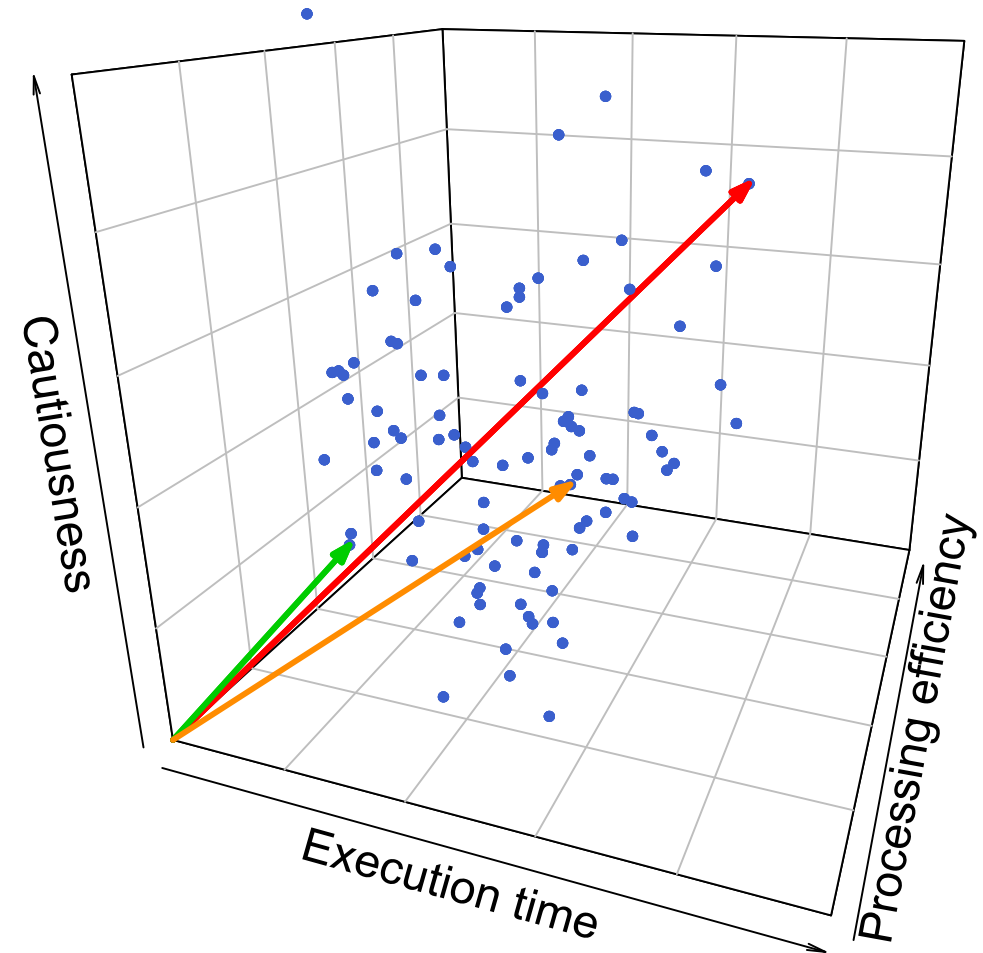
Individual differences



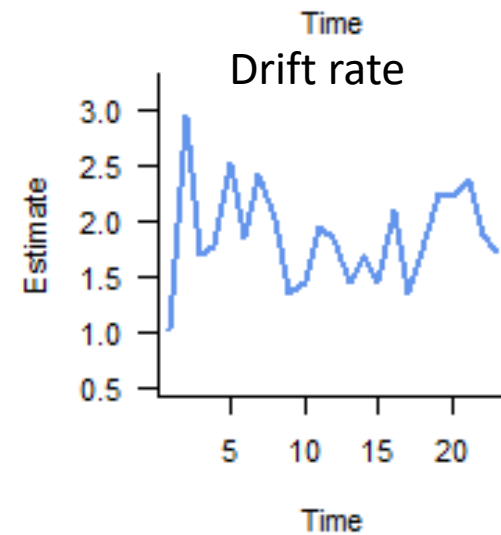
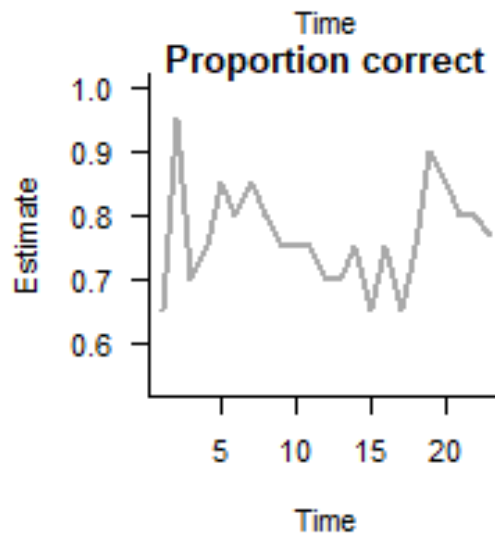
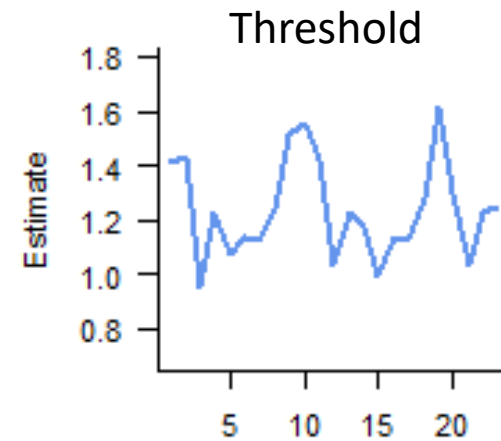
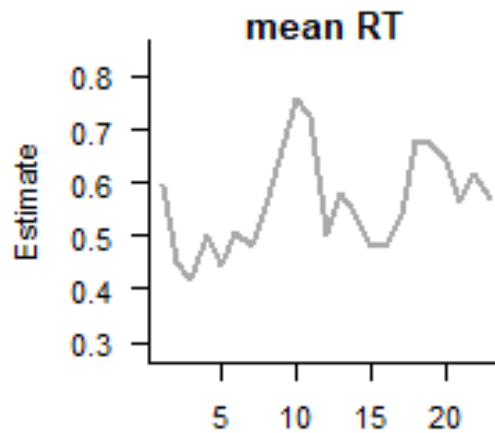
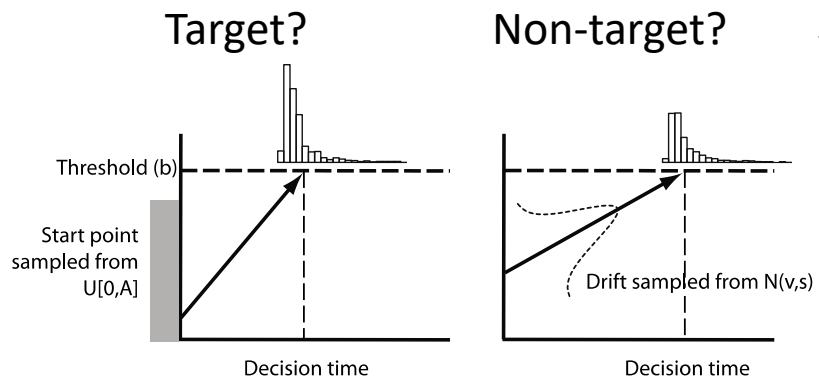
Good decision maker

Medium decision maker

Poor decision maker



Monitoring performance over time



Take home

- Formalization of Cognitive/Psychological Theory (still) has a major role in AI development
- The level of integration of the theory with AI matters
 - In terms of prediction of the user state
 - In terms of eXplainable AI
 - In terms of user acceptance and trust (?)

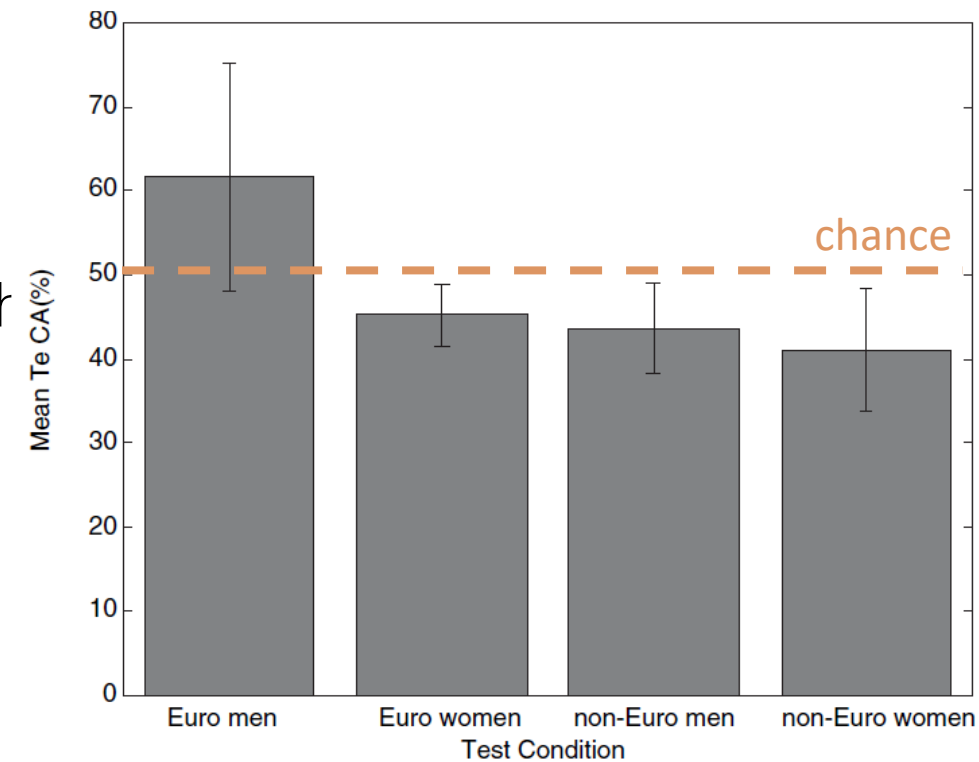
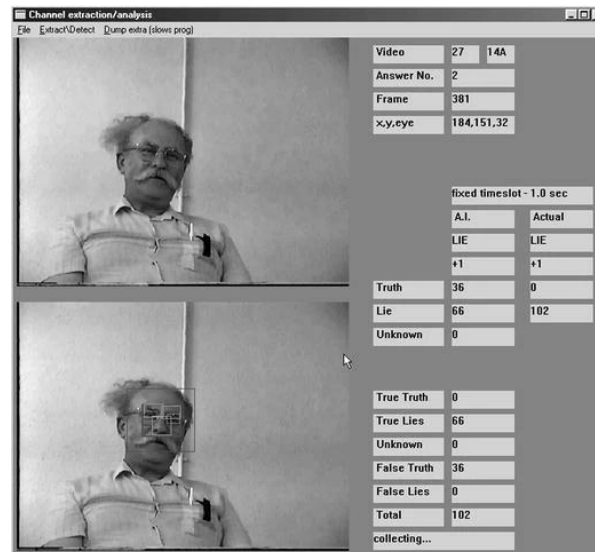
IOS-Themes:

SWOT analysis of cognitive models for user modeling

	Positive	Negative
Internal	<p>Strengths</p> <ul style="list-style-type: none">• Better model of human cognition• More feeling of autonomy	<p>Weaknesses</p> <ul style="list-style-type: none">• Model may be wrongly specified
External	<p>Opportunities</p>	<p>Threats</p>

Example of incorrect model specification: iBorderCtrl

- Aimed to predict deceit via “micro-expressions”
 - ~50 ms
 - Massive individual variability
 - Doubts over validity → Is there even a relationship between micro-expressions and deceit?



IOS-Themes:

SWOT analysis of current topic

	Positive	Negative
Internal	<p>Strengths</p> <ul style="list-style-type: none">• Better model of human cognition• More feeling of autonomy	<p>Weaknesses</p> <ul style="list-style-type: none">• Model may be wrongly specified
External	<p>Opportunities</p> <ul style="list-style-type: none">• More precise user model	<p>Threats</p>

IOS-Themes:

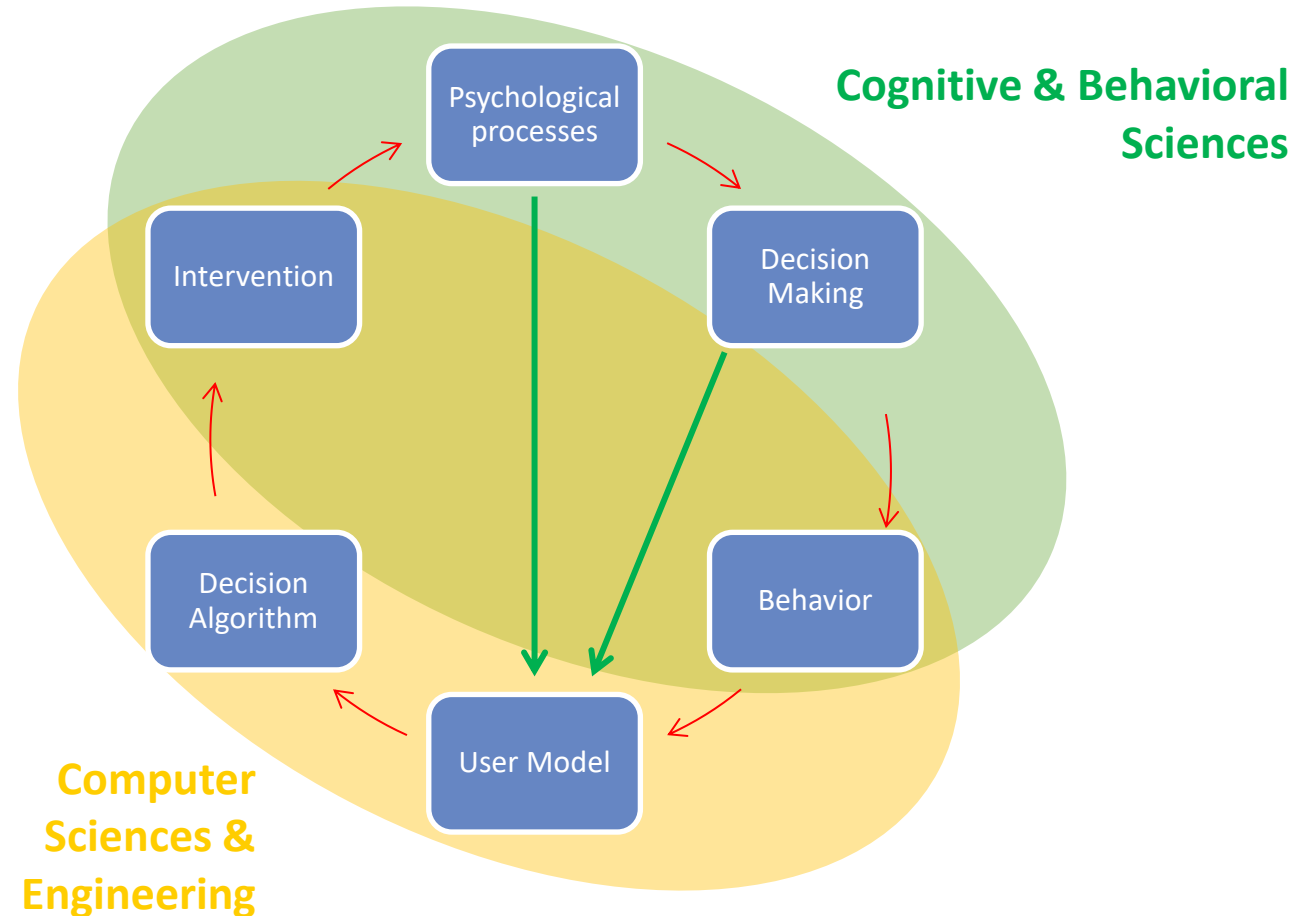
SWOT analysis of current topic

	Positive	Negative
Internal	<p>Strengths</p> <ul style="list-style-type: none">• Better model of human cognition• More feeling of autonomy	<p>Weaknesses</p> <ul style="list-style-type: none">• Model may be wrongly specified
External	<p>Opportunities</p> <ul style="list-style-type: none">• More precise user model	<p>Threats</p> <ul style="list-style-type: none">• Overspecification• Less autonomy



Levels of integration of cognitive theory

- Anecdotal
- **general knowledge of cognition**
- Computational
- **considering hypothesized computations that individuals perform**
- Algorithmic
- **considering hypothesized cognitive processes**





Utrecht University

Security in Open Societies

Bottom-up Initiatives for Societal Change

Future of Work

Open Cities

Behaviour and Institutions

Democracy and good governance

Transitions and well-being

Fair Transitions

IOS-Themes: 15 platforms

Contesting Governance

Longtermism and Institutional Change

Futures of Democracy

Markets and Corporations

The Transactional State as an Institution for Good

Gender, Diversity and Global Justice

Openness challenged: the university at risk?

In/Equality

Equality and diversity



Do people associate AI chatbots with consciousness?

Project team

Uwe Peters, Anastasia Apeiron, and Baptist Liefoghe

Folk psychological attributions of consciousness to large language models

Clara Colombatto ^{1 2}, Stephen M Fleming ^{1 3 4}

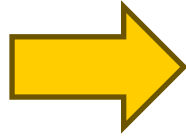
Affiliations + expand

PMID: 38618488 PMCID: [PMC11008499](#) DOI: [10.1093/nc/niae013](#)

Abstract

Technological advances raise new puzzles and challenges for cognitive science and the study of how humans think about and interact with artificial intelligence (AI). For example, the advent of large language models and their human-like linguistic abilities has raised substantial debate regarding whether or not AI could be conscious. Here, we consider the question of whether AI could have subjective experiences such as feelings and sensations ('phenomenal consciousness'). While experts from many fields have weighed in on this issue in academic and public discourse, it remains unknown whether and how the general population attributes phenomenal consciousness to AI. We surveyed a sample of US residents ($n = 300$) and found that a majority of participants were willing to attribute some possibility of phenomenal consciousness to large language models. These attributions were robust, as they predicted attributions of mental states typically associated with phenomenality-but also flexible, as they were sensitive to individual differences such as usage frequency. Overall, these results show how folk intuitions about AI consciousness can diverge from expert intuitions-with potential implications for the legal and ethical status of AI.

“the majority (67%) attributed some possibility of consciousness” to ChatGPT



But they used only **self-report measure**

Research question

Do people also **implicitly associate** chatbots with consciousness?

Use Go/no-go task

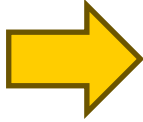
sentient



Practice round

You will first complete a **short practice round**.

The rule



The task is to press the **space bar** on your keyboard when the word on the screen either refers to **AI chatbots** (e.g., *ChatGPT, AI*) or describes **having consciousness** (e.g., *feeling, aware, conscious*).

Notice:

Do **not** press any key for words that describe **a lack of consciousness** (e.g., *unaware, unconscious*), or any other words.

The trials **will be fast**
(to elicit implicit
processing) (650 ms)

Please respond as **quickly and accurately** as possible.

Feedback will be shown after each response to help you learn the task.
Making mistakes during practice is normal.

Press **Next** to begin.

Next

