

Methods in AI Research

Chris' live lecture 1:

Cognitive Modeling & Experimentation

Chris Janssen

c.p.janssen@uu.nl
www.cpjanssen.nl

Today's structure

- 1. Interesting events coming up**
- 2. Cognitive Modeling**
 - a. Quiz questions that were harder for most**
 - b. Your questions**
- 3. Experimentation**
 - a. Quiz questions that were harder for most**
 - b. Your Questions**
- 4. Group assignment: experimentation**
- 5. Remainder: available for 1-on-1 questions
(I will stay on this call longer)**

Let's make this a conversation..

1. Interesting events

- **AI colloquia and symposia in HAI Team**
 - **2 November 16:00 Dario Salvucci (Drexel University)**
 - From distract-R
 - **14 January 15:00 Shamsi Iqbal (Microsoft Research)**
 - Attention research (discussed opening lecture); worked on design paper next week; models of attention in applications; upgrades to MS Teams
 - **January or February (t.b.d.) Rick Cooper (Birkbeck)**
 - From the paper!
 - **Even more coming up!!**

2A. Quiz questions CogMod

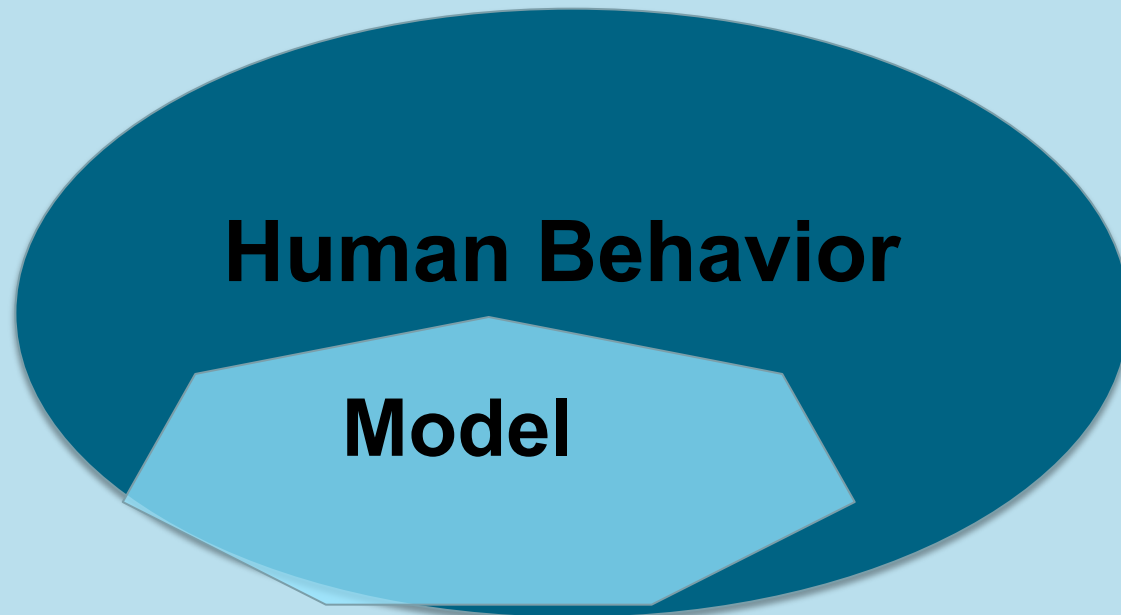
- **General remark:**
 - Might not be asked as multiple choice
 - If “at home” exam, essay question might be a bit harder
(less “looking up terms”, as you could cheat.. So, more interpretation/evaluation/application oriented)

(we hope to decide on exact format of exam very soon)

2A. Quiz questions CogMod

Q3: Below is a Venn diagram that shows a model relative to human behavior. The model is an example of a

All possible behavior



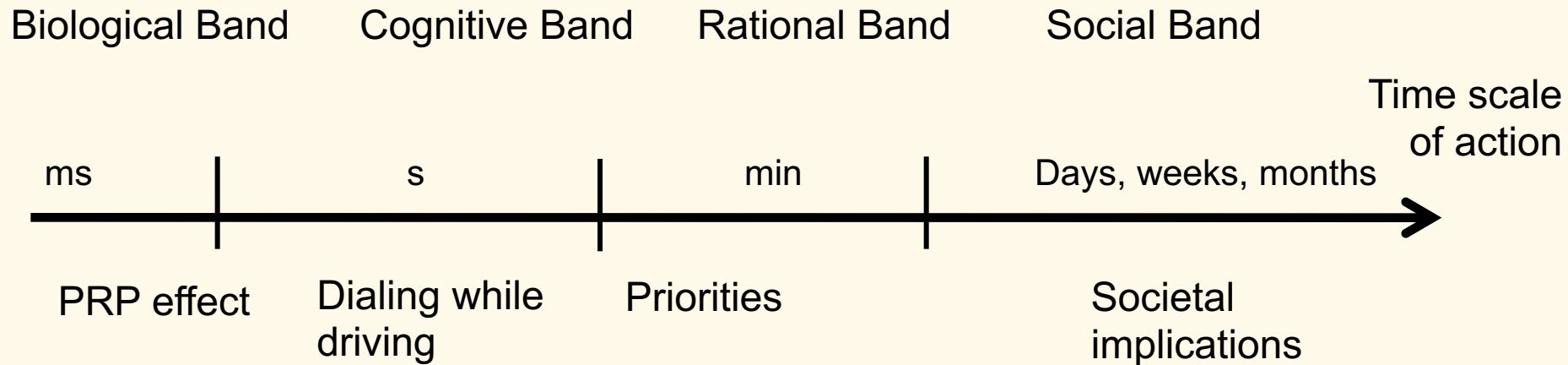
2A. Quiz questions CogMod

- Q5: Consider this case study: A researcher wants to build an “intelligent e-mail notification system”. This system dynamically tracks how “busy” a person is using a cognitive model. Based on the model, the system finds opportune moments to notify the person of incoming e-mails. For example, only when the person is judged to be not too busy (an example would be when they are checking Facebook). The researcher tests the system with a pilot dataset of 20 users that perform various tasks on the computer such as inserting data in a spreadsheet, typing e-mails, and checking facebook. She wants to find the most opportune moments to notify the user of e-mails in the future. The researcher has access to the following data (all measured in 20 ms accuracy):
 - what tasks the person/user works on when
 - eye-movement data: where do they look when?
 - mouse clicks: what do they click when?
 - key presses: what do they type when?If they were to model this, which of **Newell's bands** would be most appropriate?

Answer distribution

- **04: Marr's levels (implementation, algorithmic)**
- **12: biological**
- **54: cognitive**
- **12: rational**
- **01: social**

Abstraction continuum; time scale of action



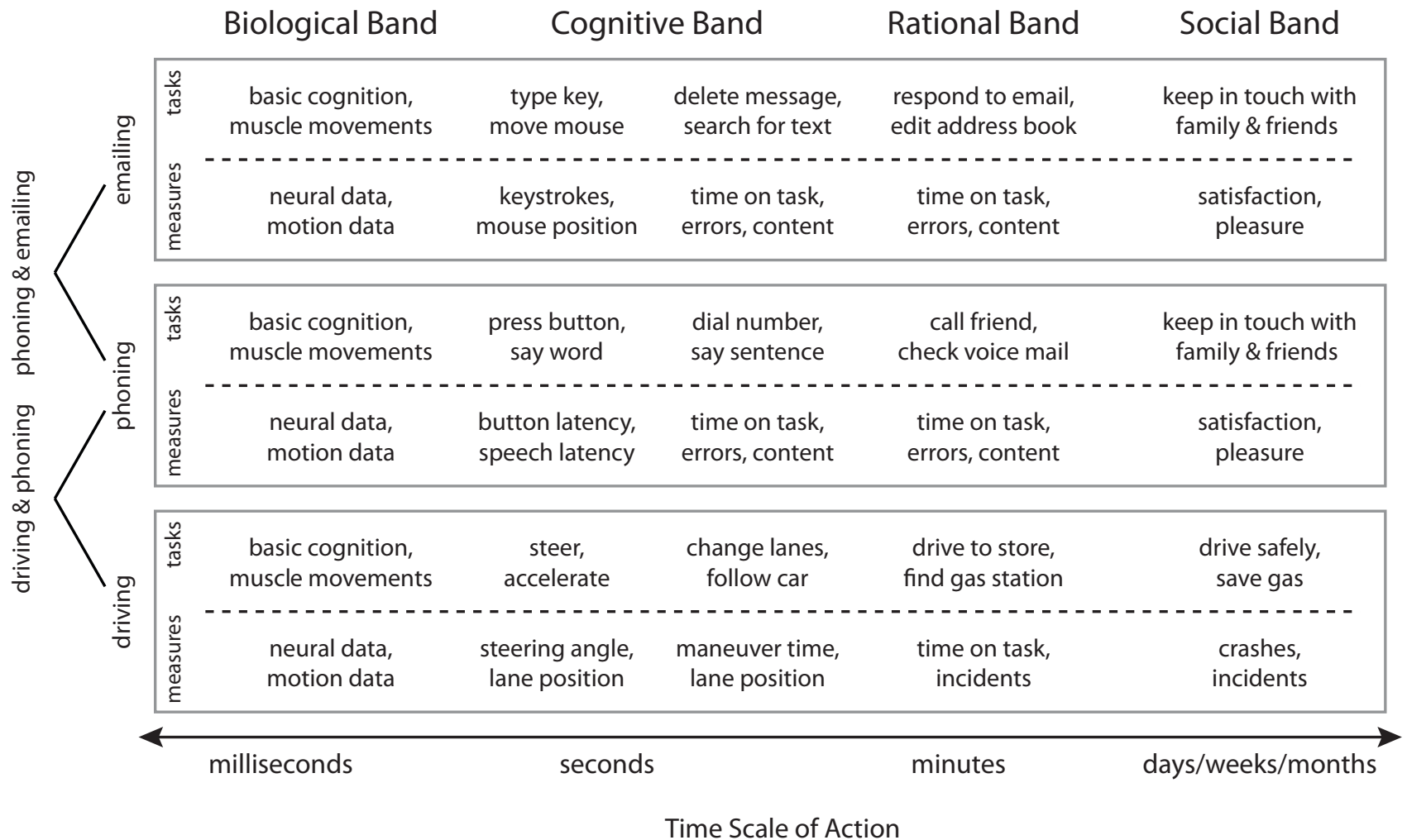


Figure 1.3: The Abstraction Continuum (derived from Newell, 1990).

(From Salvucci & Taatgen (2011) The multitasking mind)

2A. Quiz questions CogMod

- Q5: Consider this case study: A researcher wants to build an “intelligent e-mail notification system”. This system dynamically tracks how “busy” a person is using a cognitive model. Based on the model, the system finds opportune moments to notify the person of incoming e-mails. For example, only when the person is judged to be not too busy (an example would be when they are checking Facebook). The researcher tests the system with a pilot dataset of 20 users that **perform various tasks on the computer such as inserting data in a spreadsheet, typing e-mails, and checking facebook**. She wants to find the **most opportune moments** to notify the user of e-mails in the future. The researcher has access to the following data (all measured *in 20 ms accuracy*):
 - **what tasks the person/user works on when**
 - **eye-movement data: where do they look when?**
 - **mouse clicks: what do they click when?**
 - **key presses: what do they type when?**If they were to model this, which of **Newell's bands** would be most appropriate?

Why cognitive?

- **Actions (related to task) take place over ms- s**
- **Measurements within typical range of band (ms-s)**

Model is not:

- **rational or social: decision needs to be made faster than minutes**
- **biological: not so much about the underlying biological process; more about the action that needs to be taken (type of tasks done)**

2A. Quiz questions CogMod

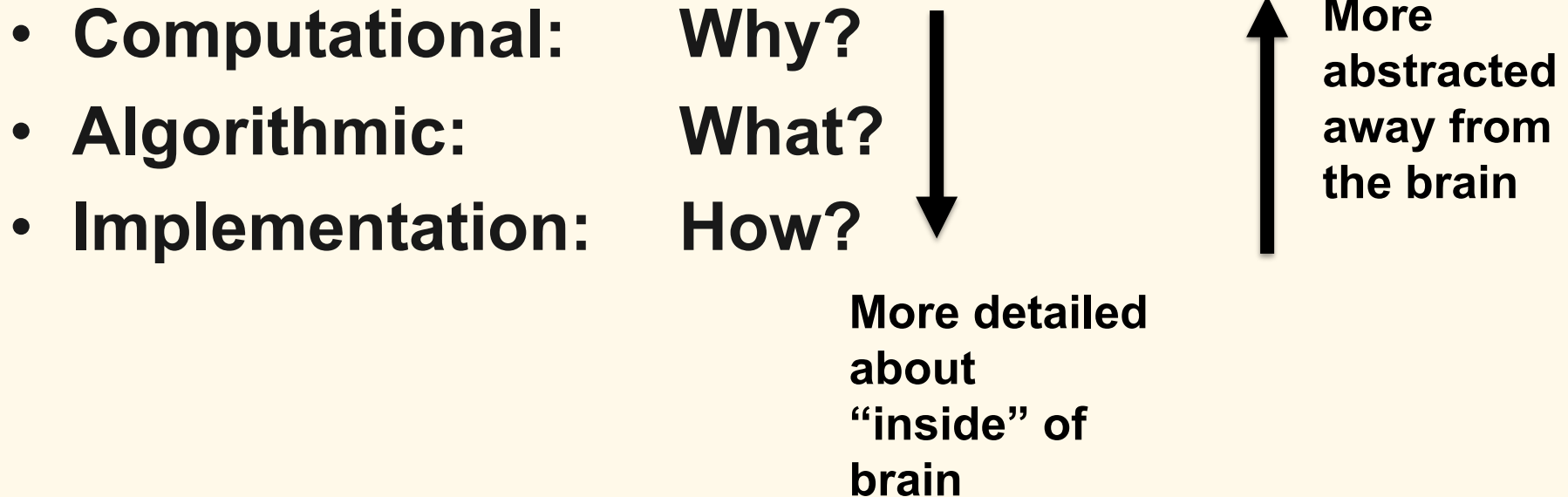
Q6: Jussi made a computer simulation of how the human visual system, memory, and manual actions interact to type letters on a keyboard. What type of model would this be in **Marr's** terms?

Answer distribution

- **07: computational**
- **32: algorithmic**
- **33: implementation**
- **09: cognitive (Newell level!)**

What to build? For humans

Levels of Abstraction - David Marr:



David Marr (1982) Vision: A computational investigation

Marr: Algorithmic level

- **Goal :**
 - What type of algorithm/strategy solves this problem?
 - What is the “input”, what is “output”
- **What they do not answer:**
 - *Why* is implementation like this? What is purpose (computational)
 - What is physical implementation? How brain achieves this (implementation)
- **Common characteristics:**
 - Detailed algorithms
 - That describe aspects of the *process*

But isn't a model always...

- **Computational?**
(because it is on a computer)
- **Algorithmic?**
(Because it is an algorithm)
- **Implementation?**
(because it is “implemented” in code)

NO!



Solution

- Let go of thinking about the computer, think about the *research question* you want to answer.
(even if this concerns humans and not models)
- Do I want to answer:
 - *Why* behavior occurs -> computational
 - *What* are information processing steps -> algorithmic
 - *How* physically realized in the brain -> implementation

2B Your questions about CogMod

- **Procedural questions:**
 - What to focus on when learning
 - Article on Marr is hard: what to focus on?
 - Read paper first, or watch lecture first?

Study questions Cooper & Peebles (not bullet proof)

- **Describe each of Marr's 3 levels**
- **What was Marr's criticism of cognitive systems that were not rooted in CL?**
- **What are limitations of CL?**
- **What are limitations of IL?**
- **What is the value of tying the CL or IL to the ARL?**
- **What is the value of specifying a model in a cognitive architecture (or with interacting subfunctions)?**

All questions: Be able to apply these principles/concepts/ideas to case studies

2B Your questions about CogMod

- **Bands and levels – hopefully tackled:**
 - **Can I explain Newell's bands once more**
 - **Marr's levels were confusing and harder to grasp compared to Newell bands**
 - **What is the difference between the Computational and Algorithmic Level?**

2B Your questions about CogMod

- **Bands and levels – still open:**
 - **The approach to link the CL and ARL levels mentioned in the 4th part of the Cooper and Peebles paper is not totally clear for me**
 - **What model to choose according to Marr or Newell is not always clear to me, is there always a clear “correct” answer?**

2B Your questions about CogMod

- **Bigger picture (at end)**
 - Can we “merge” architectures or can they be “linked” when we encounter a problem that needs more than one?
 - What are ACT-R and SOAR?
 - Why is there a problem with falsifying Computational level explanations? (*detail from article*)
 - Has neurological research made some progress since paper came out?

Today's structure

1. Interesting events coming up

2. Cognitive Modeling

- a. Quiz questions that were harder for most
- b. Your questions

3. Experimentation

- a. Quiz questions that were harder for most
- b. Your Questions

4. Group assignment: experimentation

5. Remainder: available for 1-on-1 questions
(I will stay on this call longer)

Let's make this a conversation..

3A Quiz questions experimentation

- **Q4: Jane tests how coffee dosage affects the reaction time in a driving experiment with 24 women. She tests four levels of coffee dosage: no coffee, 1 cup, 2 cups, or 5 cups. What is (or are) the independent variable(s)?**

3A Quiz questions experimentation

- Q4: Jane tests how coffee dosage affects the reaction time in a driving experiment with 24 women. She tests **four levels of coffee dosage**: no coffee, 1 cup, 2 cups, or 5 cups. What is (or are) the independent variable(s)?

3A Quiz questions experimentation

- **Q3: Cairns (2016) discusses 4 types of validity. For the following example, please indicate which validity type is breached “John tests the immersion of his new game ‘AI RULEZZ’. He develops his own immersion questionnaire, and finds that players rate the immersion of his game higher than the immersion of another game called ‘LOOZAHS’.”**

Answer distribution:

- **20: Construct**
- **19: Internal**
- **11: External**
- **11: Ecological**

3A Quiz questions experimentation

- Q3: Cairns (2016) discusses 4 types of validity. For the following example, please indicate which validity type is breached “John tests the immersion of his new game ‘AI RULEZZ’. He **develops his own immersion questionnaire**, and finds that players rate the immersion of his game higher than the immersion of another game called ‘LOOZAHS’.”

Answer distribution:

- **20: Construct**
- 19: Internal
- 11: External
- 11: Ecological

3B Your questions about Experim.

- **Process:**

- Is there a possibility to access the Cairns paper without paying for it?
- how comprehensively should we read the Cairns paper?

3B Your questions about Experim.

- **Already (partially) covered:**
 - Elaborate a bit on question 3 in the live lecture.
 - Validity was a bit confusing for me, I did note the general ideas, but linking it to practical situation was little confusing for me
 - Validity is hardest even with having read the paper it is still really vague. Especially the difference between ecological and external validity.
 - External : Whether the results have a wider application outside of the science lab.*
 - Ecological : Whether the results have relevance in the real world.*
- Maybe it's just me, but these just sound like different ways of saying the same thing?

3B Your questions about Experim.

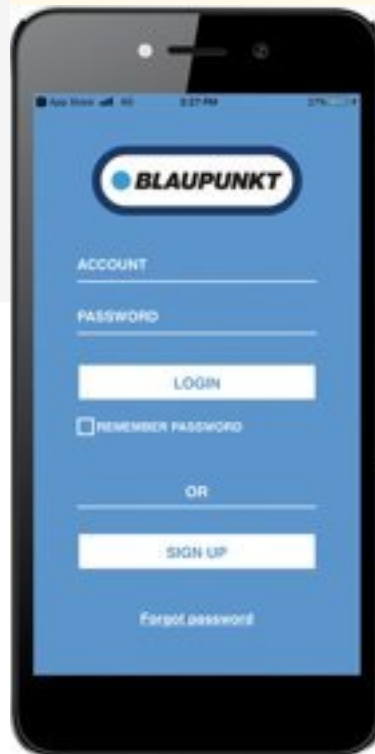
- **Bigger picture:**
 - Can one never use the empirical cycle when researching an existing dataset?

3B Your questions about Experim.

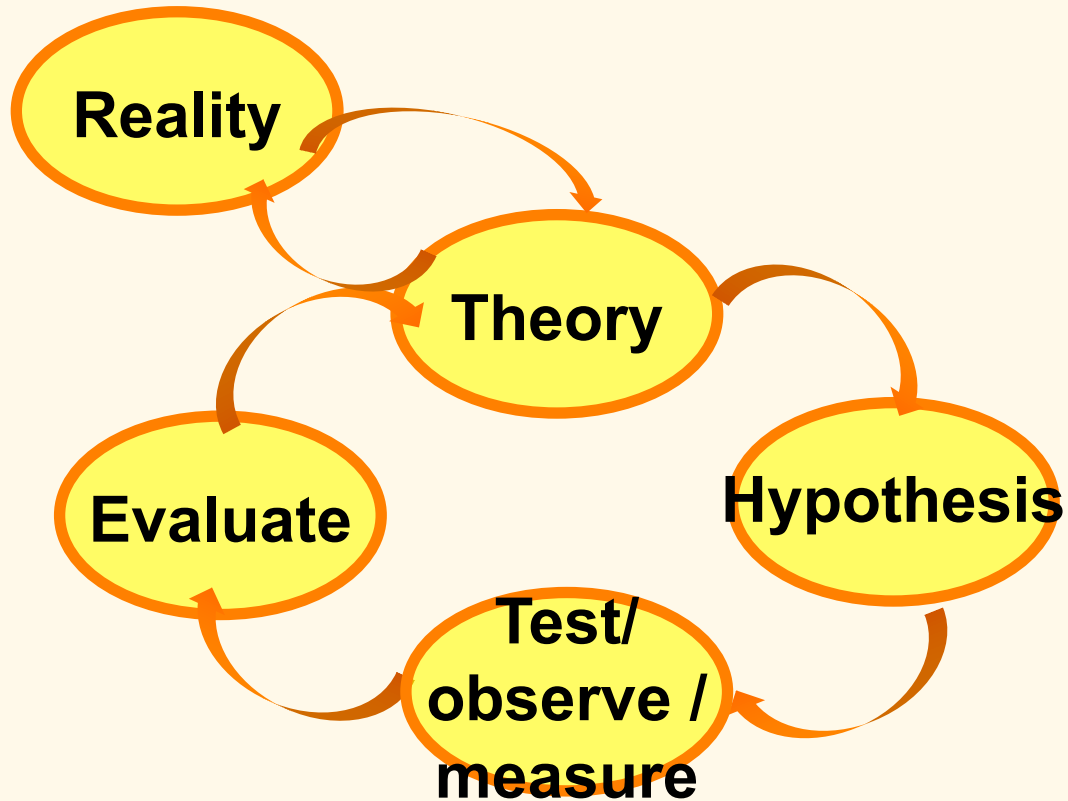
- **Bigger picture:**
 - Can one never use the empirical cycle when researching an existing dataset?
 - I would like to see more problems or violations of experiments addressed! More examples what you should not do or what might go wrong, even if you think it is right.



Map-based robot



Which vacuum is best: world-based (Chris) or map/model-based (Remo)?



- Participants
- Materials/Stimuli
- Design
- Procedure
- Measures

- **Terms to also consider...**

- **Manipulation**
- **Causality**
- **validity (4 types)**
- **confounds & control**
- **(in-)dependent variable**
- **Factor**
- **Condition**
- **Level**
- **within- and between-subjects**
- **counterbalancing....**

Discuss

- **Discuss with your peer-group for X minutes, then return**
- **Be prepared to report back!**

2B Your questions about CogMod

- **Bigger picture (at end)**
 - Can we “merge” architectures or can they be “linked” when we encounter a problem that needs more than one?
 - What are ACT-R and SOAR?
 - Why is there a problem with falsifying Computational level explanations? (*detail from article*)

2B Your questions about CogMod

- **Bigger picture**
 - **Has neurological research made some progress since paper came out? Context: recent Neuralink developments, and was wondering whether that implies that they were able to put the IL and neuroscience evidence closer together, since it looked like they managed to read information from parts of the brain? Or am I seeing that wrong since information gathering isn't the same as analyzing cognition?**

Next week

- **Topics**
 - (Scientific) writing
 - Designing responsible AI (lectures by Dong and me)
- **Preparation**
 - 2 quizzes
 - 2 preparatory assignments

Today's structure

- 1. Interesting events coming up**
- 2. Cognitive Modeling**
 - a. Quiz questions that were harder for most**
 - b. Your questions**
- 3. Experimentation**
 - a. Quiz questions that were harder for most**
 - b. Your Questions**
- 4. Group assignment: experimentation**
- 5. Remainder: available for 1-on-1 questions
(I will stay on this call longer)**

Let's make this a conversation..