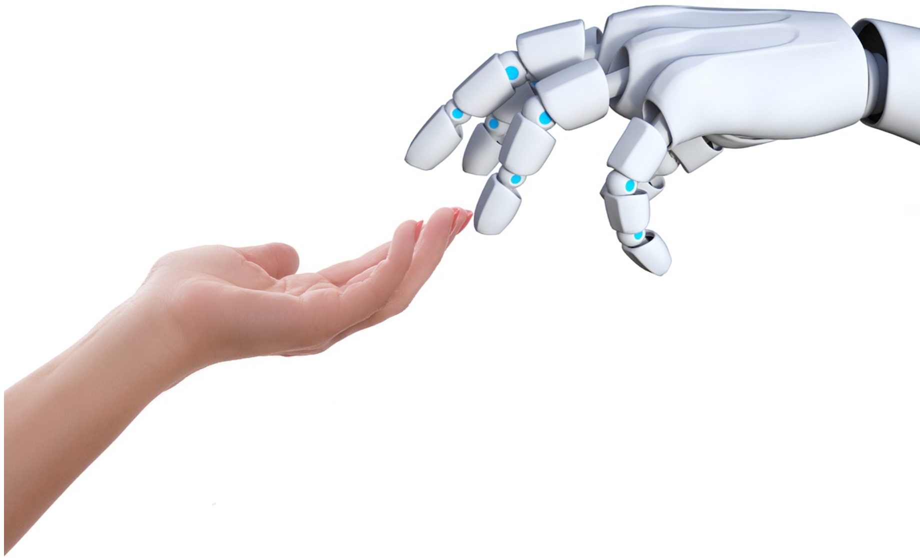
CHI2023 Course (C28)
Cognitive Modelling: From GOMS to Deep Reinforcement Learning

CHI23
Hamburg, Germany | Hybrid
April 23-28, 2023
reCHInnecting



Introduction to Cognitive Modelling

Jussi Jokinen, Antti Oulasvirta, Andrew Howes

“What I cannot create, I do not understand.”
— Richard Feynman

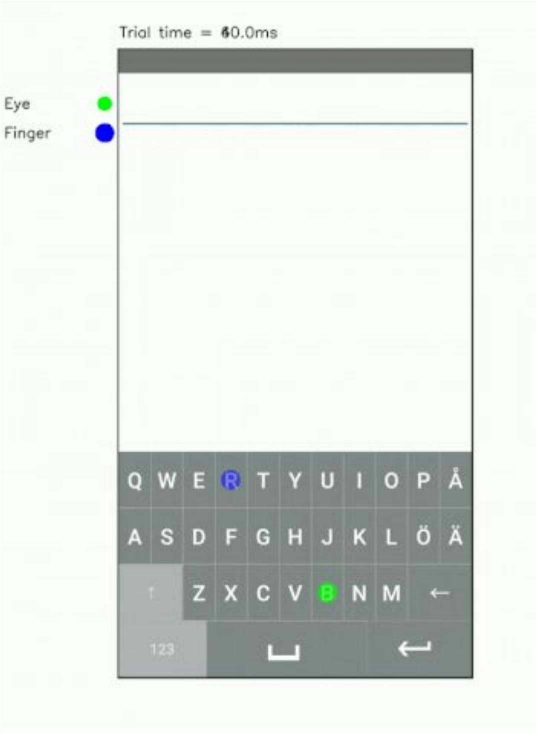
“Science is what we understand well enough to explain to a computer; art is everything else.”
— Donald E. Knuth

“Trying to understand human cognition is analogous to that of trying to discover how a computer has been programmed.”
—Ulrich Neisser

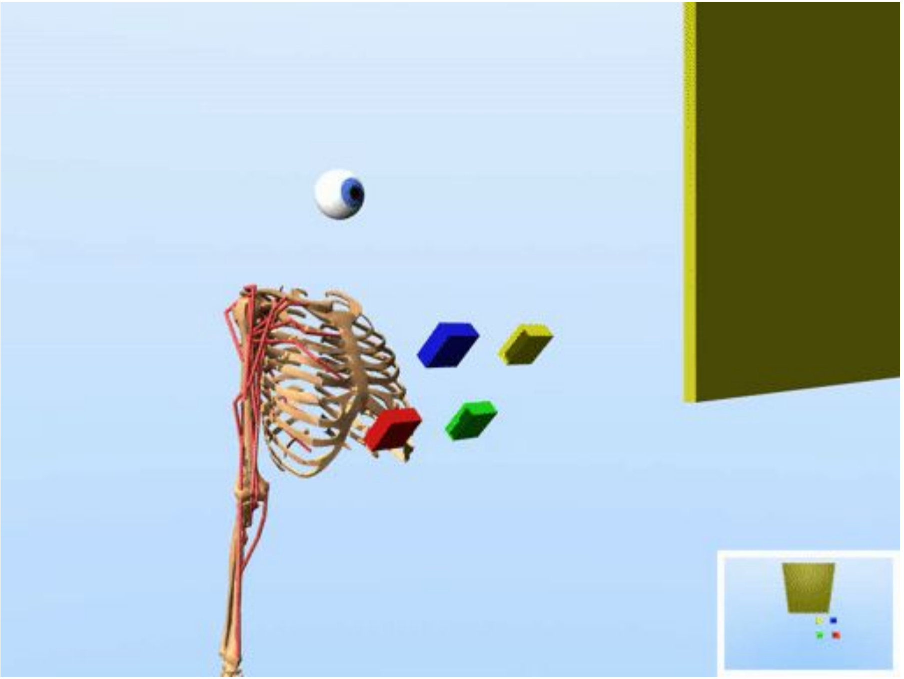
7 reasons to be excited about cognitive modeling in 2023

1. Models help us understand users and explain their behavior.
2. Models help us predict behavior and experience.
3. Models drive algorithmic design and adaptation of user interfaces.
4. Modern models are converging with machine learning (ML) theory.
5. Modern models are generative. They can learn from data and experience.
6. Models provide a strong prior for ML models to partner with humans.
7. Models may be needed for human-centric AI.

Emerging cognitive models are generative



Jokinen et al. CHI'21



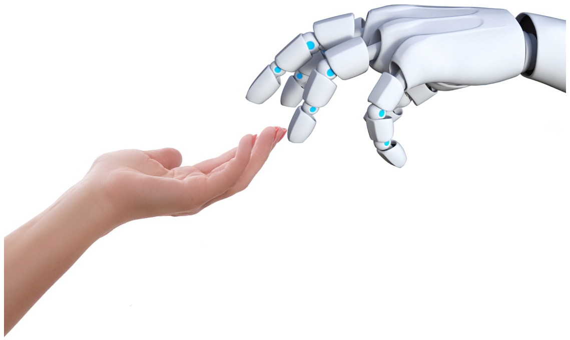
Ikkala et al. UIST'22

Course overview

Cognitive models are computer programs that model how users perceive, think, and act. Modeling offers a powerful approach for understanding interactive tasks and improving designs.

1. Learn to understand, create, and utilize computational cognitive models in HCI contexts.
2. Use Python notebooks to easily simulate interactive task behaviour in HCI environments.
3. Topics include deep learning, computational rationality, reinforcement learning, model design

Today’s schedule



09:00-10:30 Module 1: Introduction

10:30-11:00 Break

11:00-12:30 Module 2: Deep neural nets

12:30-14:00 Lunch

14:00-15:55 Module 3: Computational rationality

15:55-16:35 Break

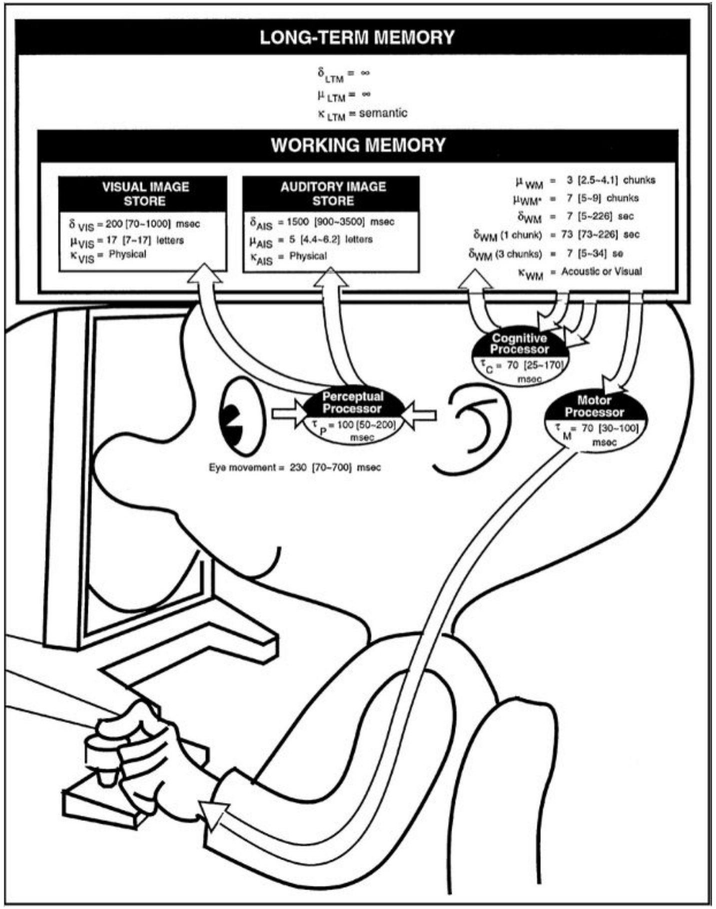
16:35-18:00 Module 4: Building a model

Module 1: Introduction

Slides

1. Cognition
2. Scientific modeling
3. Cognitive modeling
4. Discussion

Antti Oulasvirta

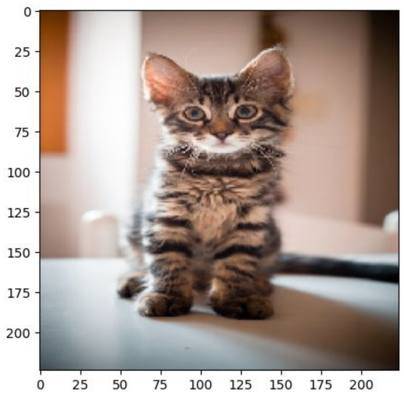


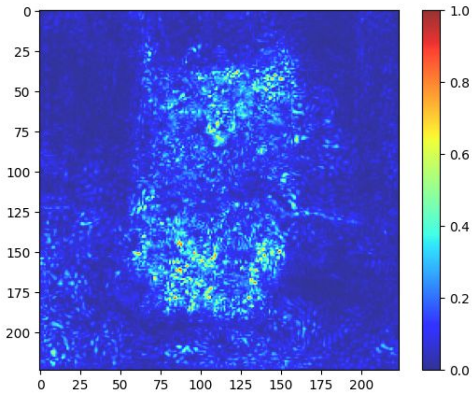
Module 2: Deep neural nets

A Colab notebook

1. Deep learning primer
2. DNNs as models of cognition
3. Visual saliency as an example

Antti Oulasvirta

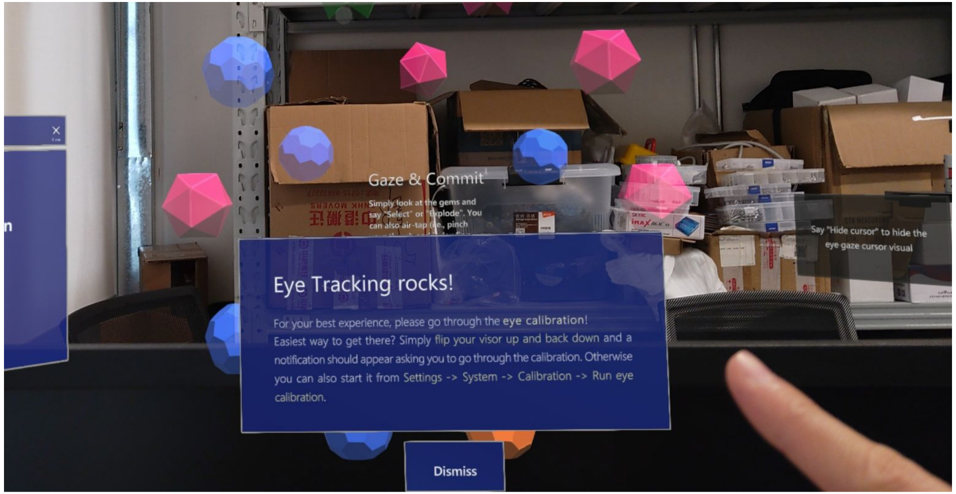




Module 3: Computational rationality

A Colab notebook

1. Reinforcement learning
2. Computational rationality
3. Case: Gaze-based interaction



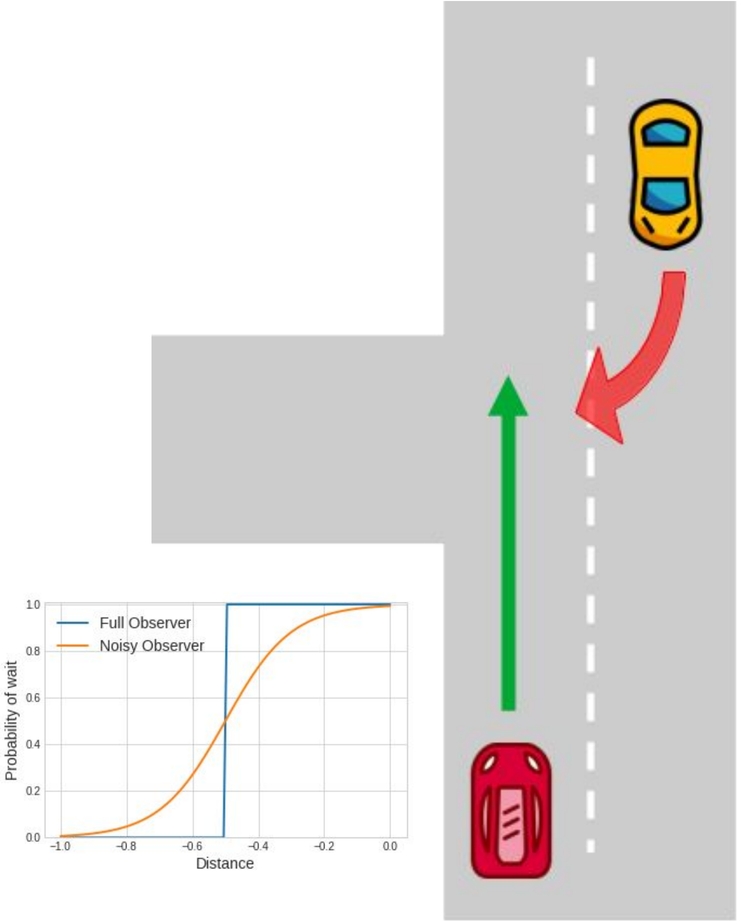
Andrew Howes

Module 4: Building a Model

We build a model of a driver in Colab

1. Define the goal
2. Define the environment
3. Define the cognitive limitations
4. Derive the optimal behavior
5. Inspect model validity

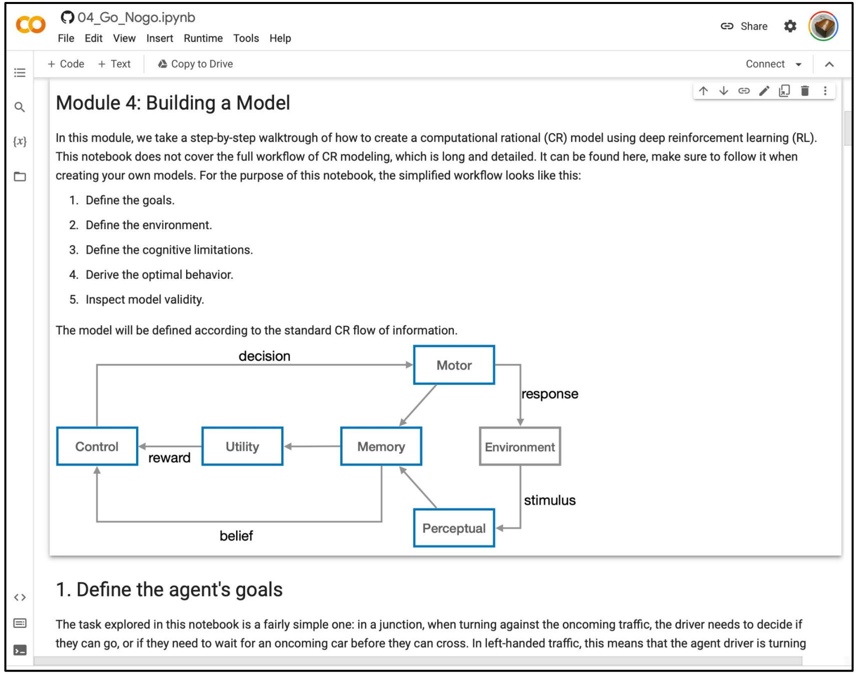
Jussi P.P. Jokinen



icons: flaticon

Let’s set up Colab

Instructions



Module 4: Building a Model

In this module, we take a step-by-step walktrough of how to create a computational rational (CR) model using deep reinforcement learning (RL). This notebook does not cover the full workflow of CR modeling, which is long and detailed. It can be found here, make sure to follow it when creating your own models. For the purpose of this notebook, the simplified workflow looks like this:

1. Define the goals.
2. Define the environment.
3. Define the cognitive limitations.
4. Derive the optimal behavior.
5. Inspect model validity.

The model will be defined according to the standard CR flow of information.

1. Define the agent's goals

The task explored in this notebook is a fairly simple one: in a junction, when turning against the oncoming traffic, the driver needs to decide if they can go, or if they need to wait for an oncoming car before they can cross. In left-handed traffic, this means that the agent driver is turning

A warm up task (5 mins)

1. Find a discussion partner
2. Introduce yourselves
3. Come up with an HCI topic you might (hypothetically) want to build a cognitive model of
4. Write it down

Coming up next in this module

1. Cognition
2. Modelling
3. Cognitive models

Introduction

1. Cognition

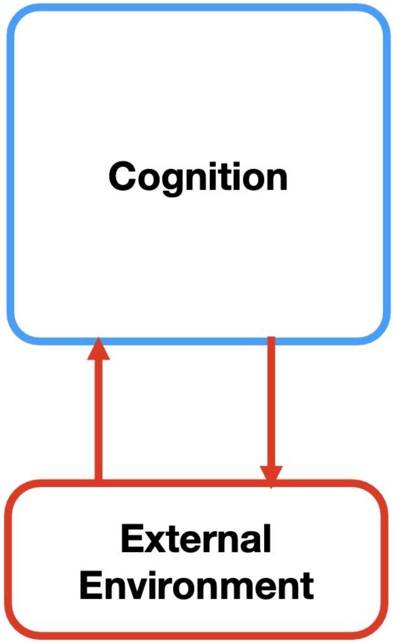
Some slides in this section are based on the book: Hornbaek, K., Kristensson, P.O. and Oulasvirta, A. 2023. *Introduction to Human-Computer Interaction*. Under contract with Oxford University Press.

Warm up: Let’s compare these two scenarios

Q: How are the two different in terms of: goals, motor demands, perceptual demands, cognitive demands?

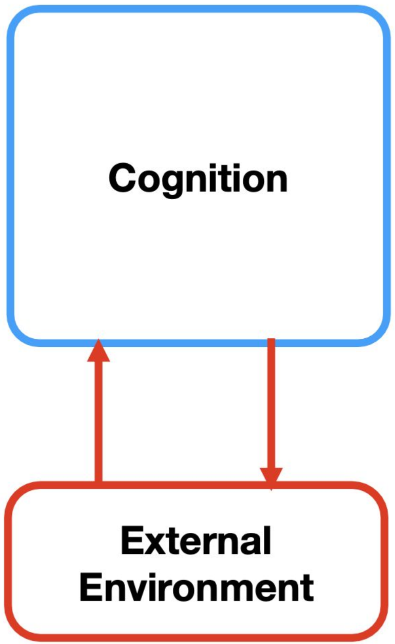
General properties of human cognition

1. It is goal-oriented
2. It adapts
3. It learns
4. It carries out computations on representations
5. It is limited
6. It requires energy and effort



Next up

1. **Perception**
2. Motor control
3. Cognition



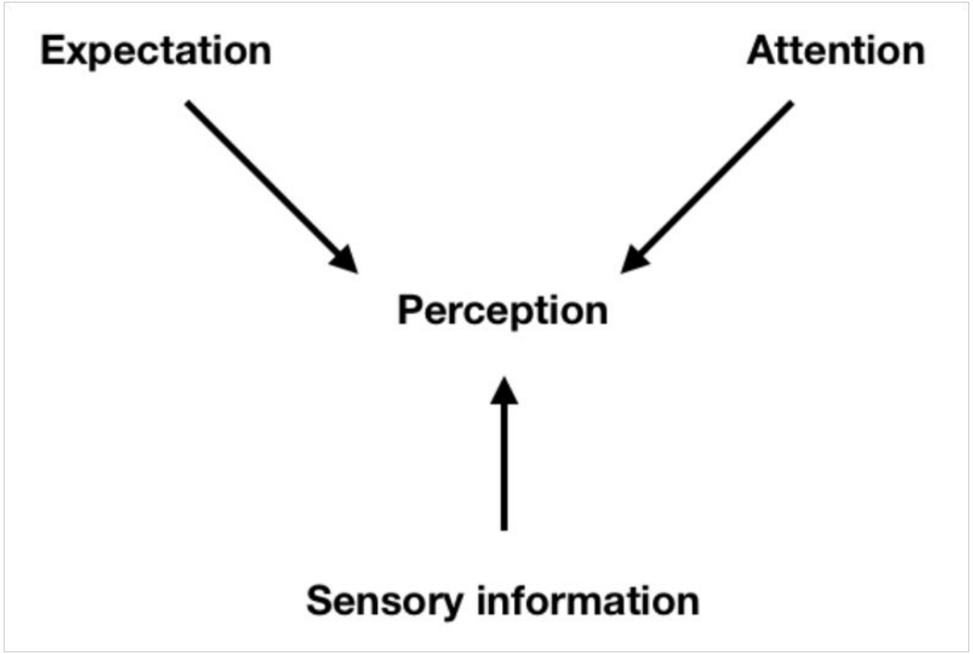
Perception serves multiple roles in HCI

- Perception is needed to regulate actions
- UIs communicate their state via perception
- Visual, auditory, and tactile perception each have their own characteristics and roles in HCI

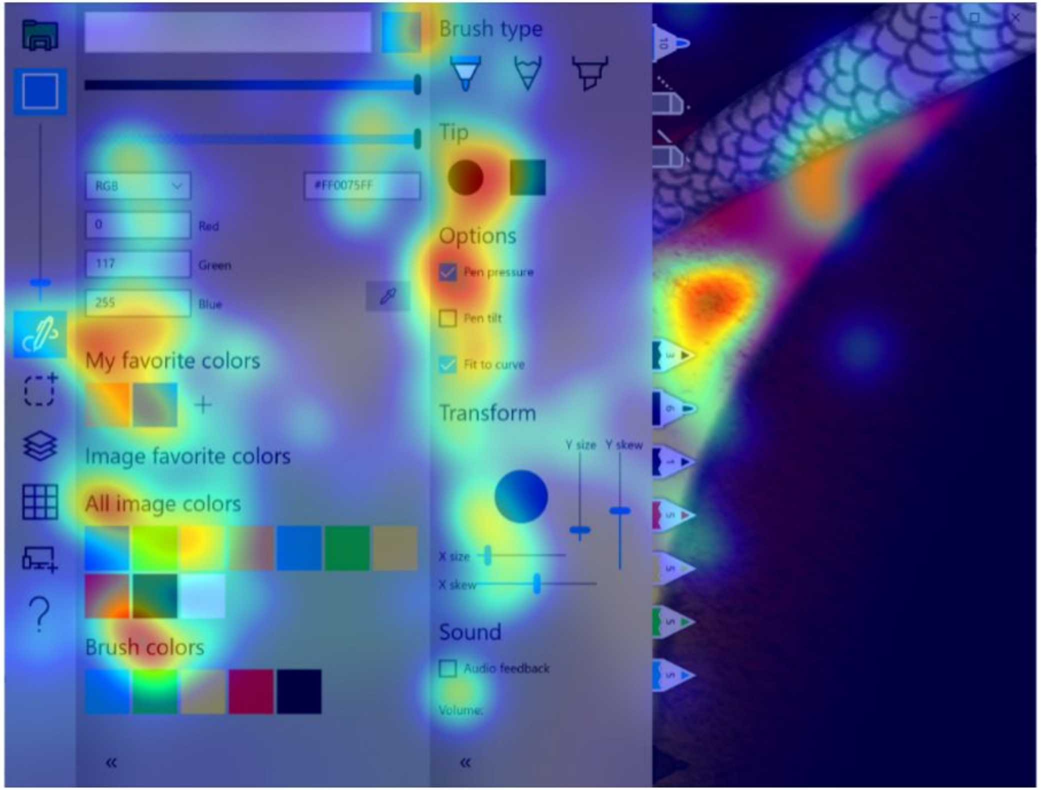
In this example, perception helps us find elements on the display, guide fingers, maintain awareness of the background, and control walking.

Perception is not determined just by the external reality

- Perceptual experience appears like it reflects sensory data
- But, it is a constructed representation
- Perception is affected by our expectations that draw from prior experiences
- Perception is also shaped by how we deploy attention to sample information



Example: Visual saliency



Next lecture

Characteristics of three sensory modalities

Sensory modality	Key characteristics	Design considerations
Vision	Fast, high bandwidth for parallel processing, field of view about 180 degrees	Visual, spatial, and lexical aspects of graphical displays, like contrast, acuity, use of color, visual primitives, symbols, and text
Hearing	Very fast (about 40ms faster reaction times than vision) but more serial presentation, 360 degrees	Properties of sound and voice, such as pitch, timbre, melody, and phrasing
Tactition	Fast but limited to areas of physical contact	Properties of haptic stimulation like amplitude and frequency of vibration

Elementary perceptual tasks

Discrimination: telling whether a difference occurs in sensory data

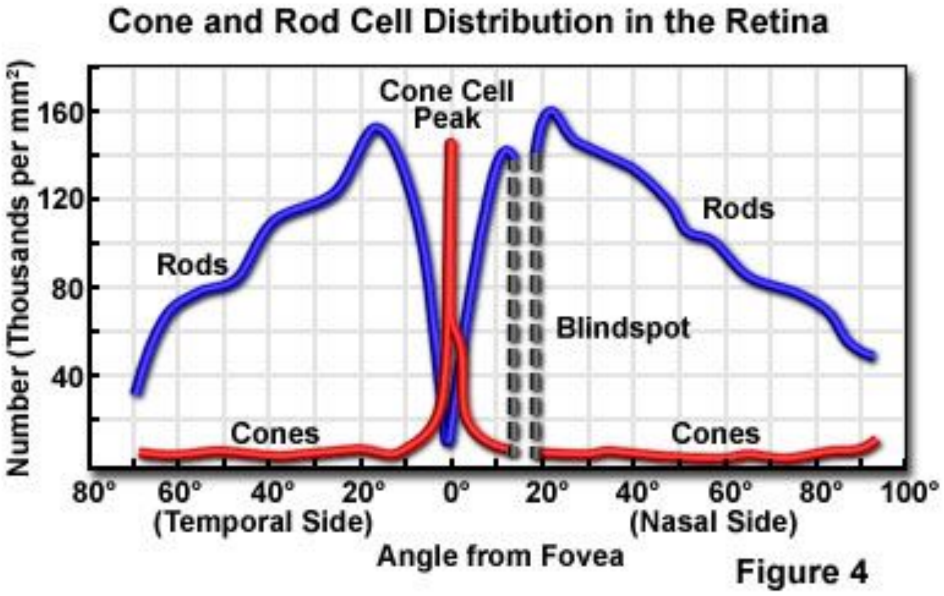
Detection: telling whether an event of interest occurs or not

Recognition: categorizing a stimulus as something

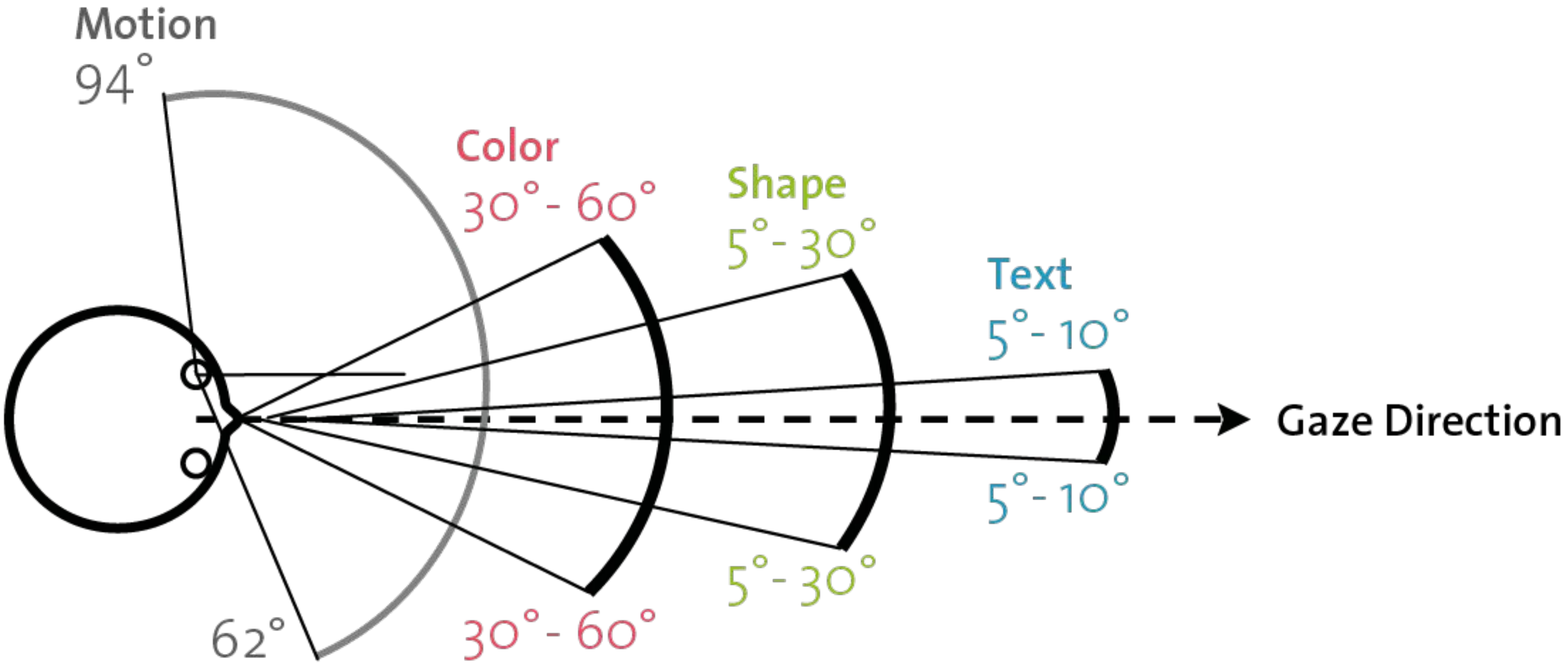
Estimation: estimating a property of an object of event in the environment

Search: localizing an object of interest

Vision is foveated



Some visual primitives are perceived better than others in the periphery.



Eye movements

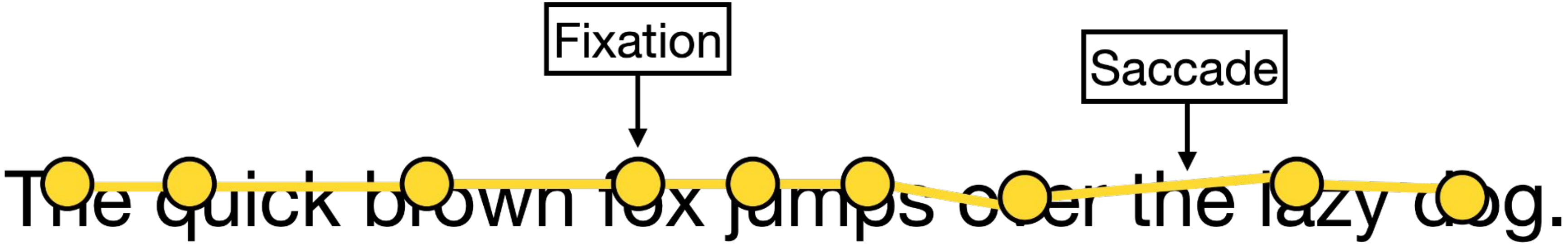
The oculomotor system controls eye movements using three main modes:

Fixations (200-400ms) encode information about the visual scene. They consist of multiple micro-fixations, each a few tens of milliseconds.

Saccades (30-70ms) move the gaze point in ballistic leaps during which the scene is not perceived.

Smooth pursuit tracks moving targets, such as when following an animated character moving on the display. No saccades occur.

Example: Reading



Fix number	Timestamp	Duration	GazepointX	GazepointY
1	48	160	378	761
2	247	199	581	752
3	466	179	751	738
4	666	160	840	740
5	865	239	602	748
6	1483	199	476	767
7	2500	160	664	805
8	2859	140	602	773
9	4733	140	596	795
10	4892	259	713	771

Credit: Duncan Brumby / UCL

Visual attention

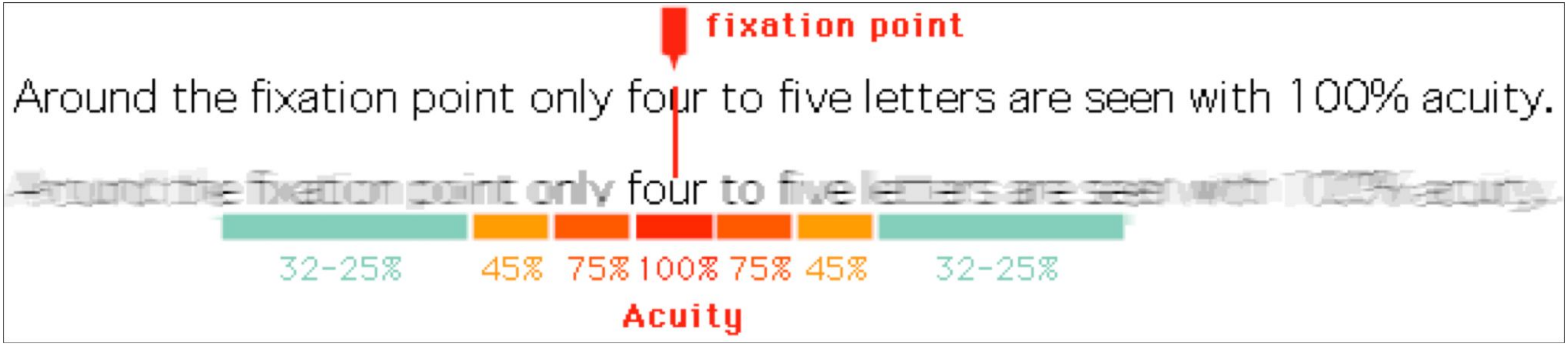
Attention focuses perceptual processing on a region or object.

Visual attention consists of three processes:

1. Selective attention: the ability to shift attention to a desired object or location
2. Divided attention: the ability to share attention between one or more objects, locations, or tasks
3. Vigilance: the ability to sustain attention on something for a longer period

Change blindness is a failure of a user to detect a change within the visual field. In **inattentional blindness** this failure occurs irrespective of any visual disruption such blinking.

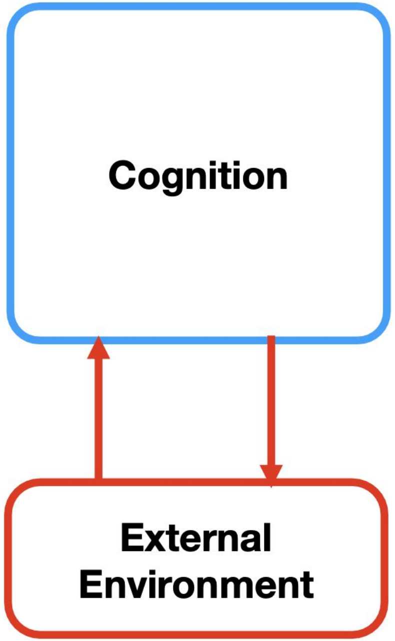
Example: Perceptual span while reading



- How much information do we process from a given point of fixation?
- Rayner (1975): to recognise a word it must be within three to four character spaces of fixation.

Credit: Duncan Brumby / UCL

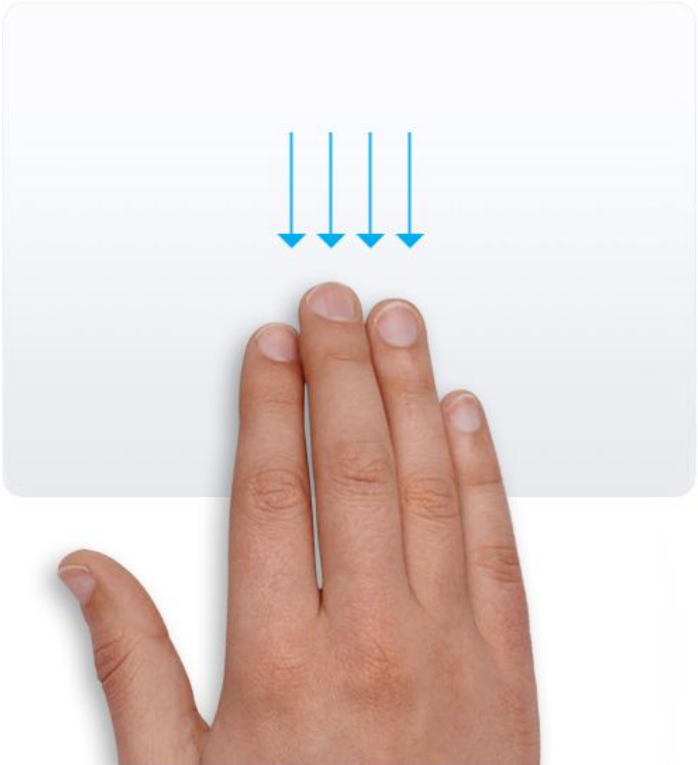
1. Perception
2. **Motor control**
3. Cognition



Motor control plays a decisive role in input to computers

From scrolling with a touchpad to commanding a large display at a distance, users rely on motor control to interact

- Motor control refers to the regulation of movement
- Motor commands are signals sent to muscles to activate
- Motor control includes integration of internal representations with sensory datum
- Motor control is critical for HCI tasks like reaching, pointing, pressing, gesturing, steering, intercepting



Key concepts

1. End-effector
2. Degrees-of-freedom
3. Open-loop control
4. Closed-loop control
5. Aimed movement
6. Interception task
7. Speed-accuracy tradeoff

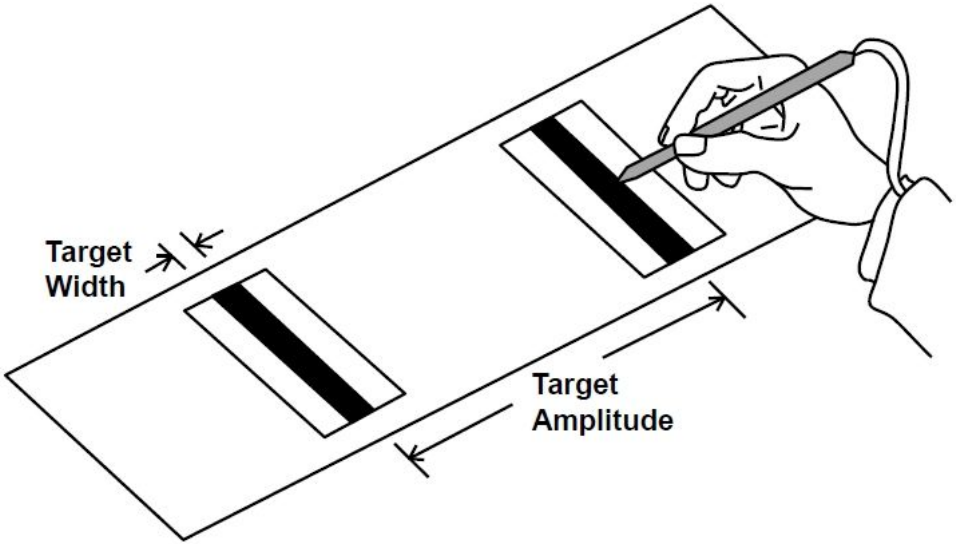
Fitts' law is a motor control model commonly used in HCI

Predicts movement time (MT) given index of difficulty (ID):

$$MT = a + b\ ID$$

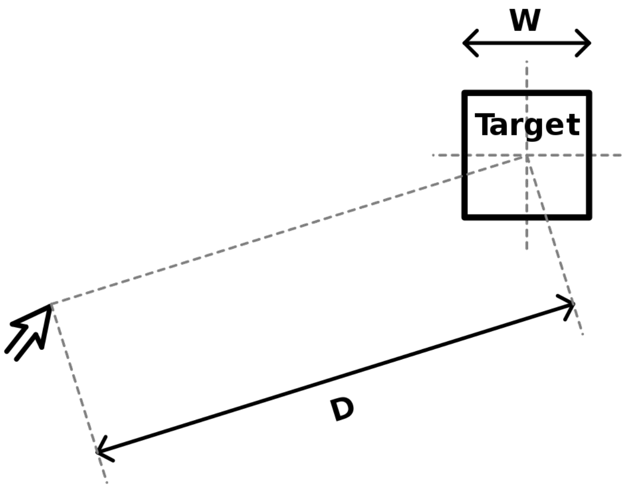
where

- $ID = \log_2 (D/W+1)$
- a and b are task-dependent coefficients
- D and W are the distance and width

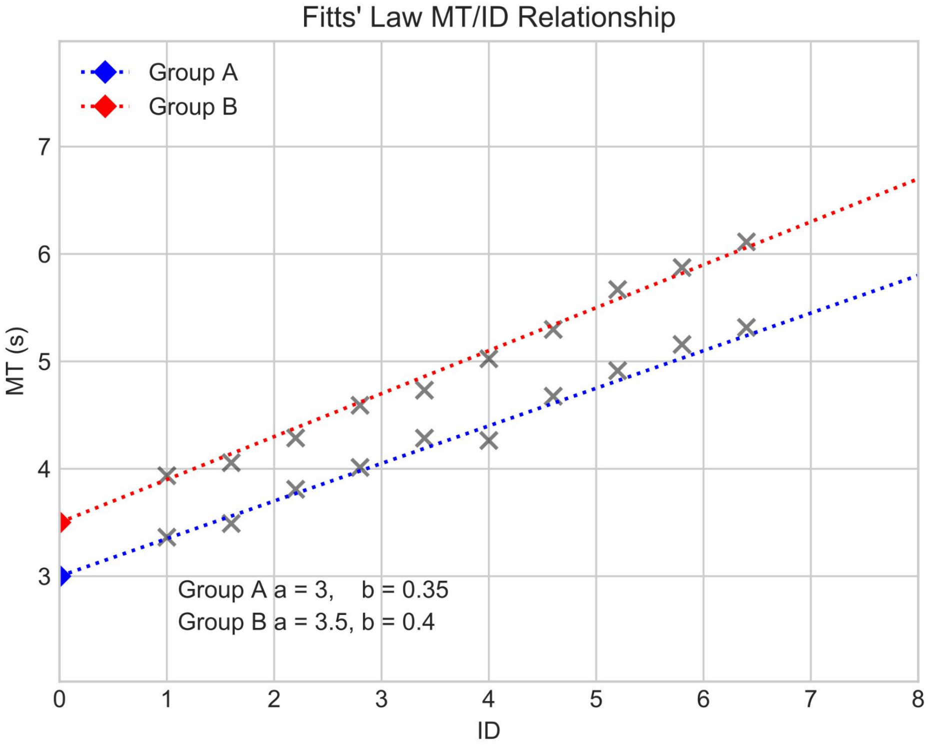


Speed-accuracy trade-off in aimed movements

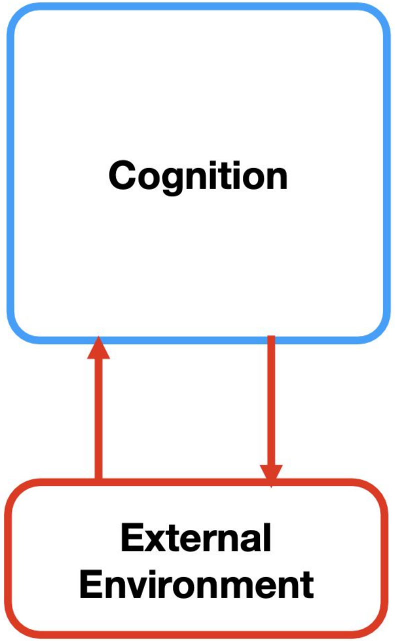
Fast or accurate? Choose one!



$$MT = a + b\ ID = a + b \log_2 (D/W+1)$$



1. Perception
2. Motor control
3. **Cognition**



Key cognitive capacities needed in HCI

Memory

"Powerpoint has image editing capabilities."

Attention

Searching for the icon, guiding finger to press it

Control

"What should I press now to edit a photo?"

Reasoning

"Could I edit the photo in Powerpoint and take a screenshot to store it?"

Decision-making

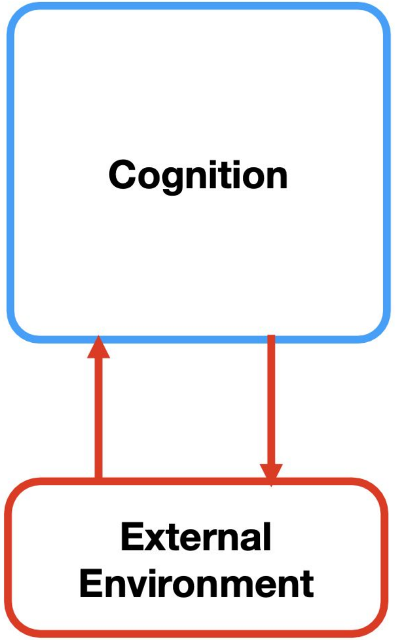
"I don't have time to learn this, can I ask Mary to do it?"

Capacities and their functions: A closer look

Supervisory control: adaptively deciding goals, allocating cognitive resources to tasks, and changing of course of action when required.
Memory: forming, maintaining, and accessing beliefs about objects that are not directly perceivable.
Attention: selectively processing some part of the perceptual field.
Reasoning: applying transformation rules to beliefs to form new beliefs.
Decision-making: gathering information and choosing between options.

General properties of human cognition (again)

1. It is goal-oriented
2. It adapts
3. It learns
4. It carries out computations on representations
5. It is limited
6. It requires energy and effort



1. Cognition is goal-oriented

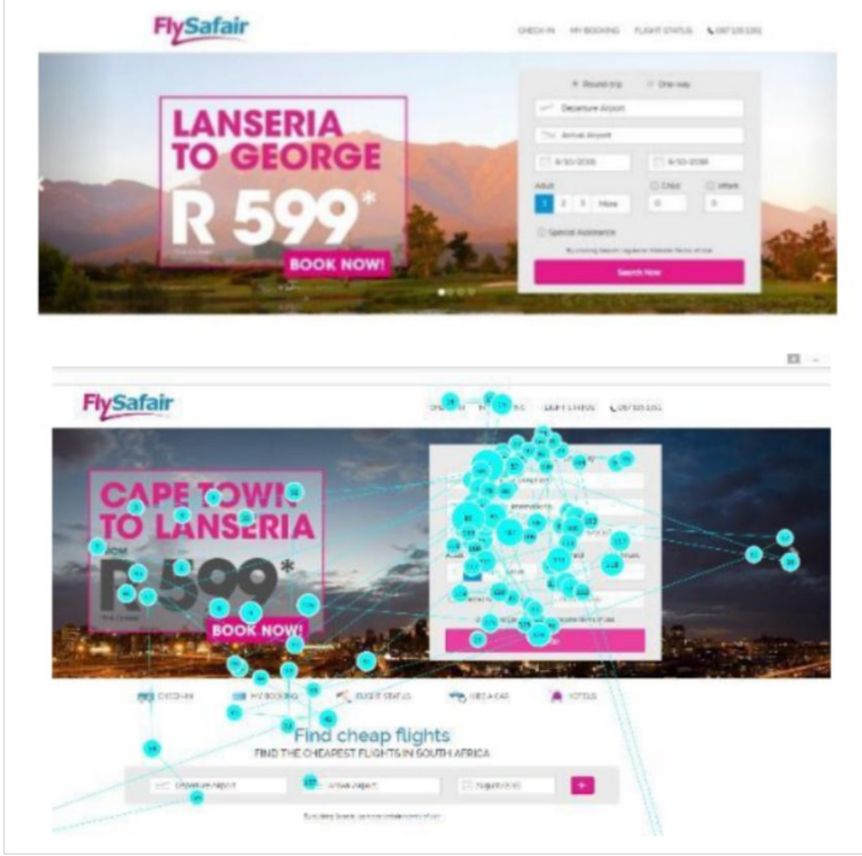
Problem: At any given time, numerous stimuli barrage our senses and we have several options on how to direct cognitive resources to them.

Cognitive control refers to our ability to direct thinking and action toward some goal:

- Setting goals
- Directing resources and attention
- Multitasking
- Task-switching
- Inhibiting distracting ideas

Example: Inattentional blindness

When users are asked process an unrelated task, they remember nothing of the banner, even though it is visually salient



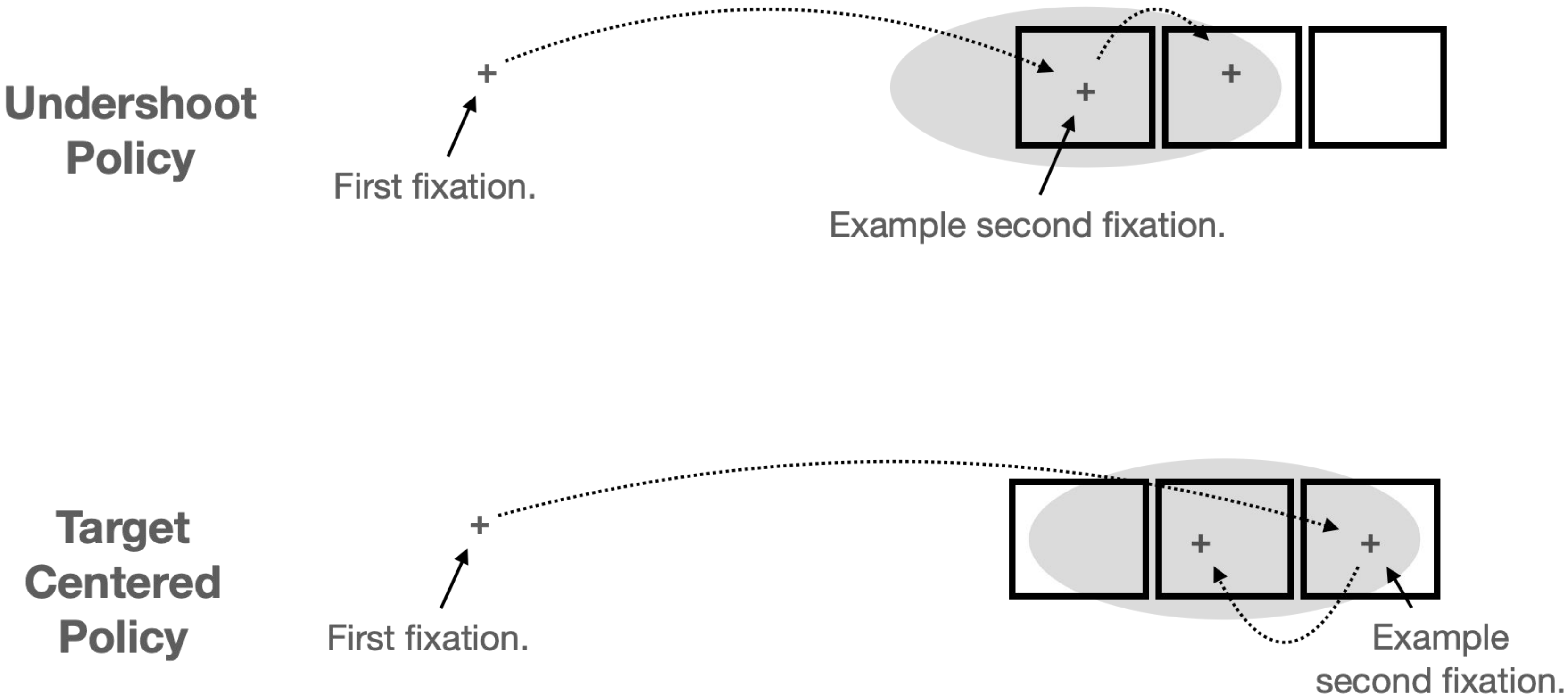
Yet, according to eye-tracking data, they saw it!

H. Gelderblom and L. Menge. The invisible gorilla revisited: using eye tracking to investigate inattentional blindness in interface design. In Proceedings of the Conference on Advanced Visual Interfaces, pages 1–9, 2018.

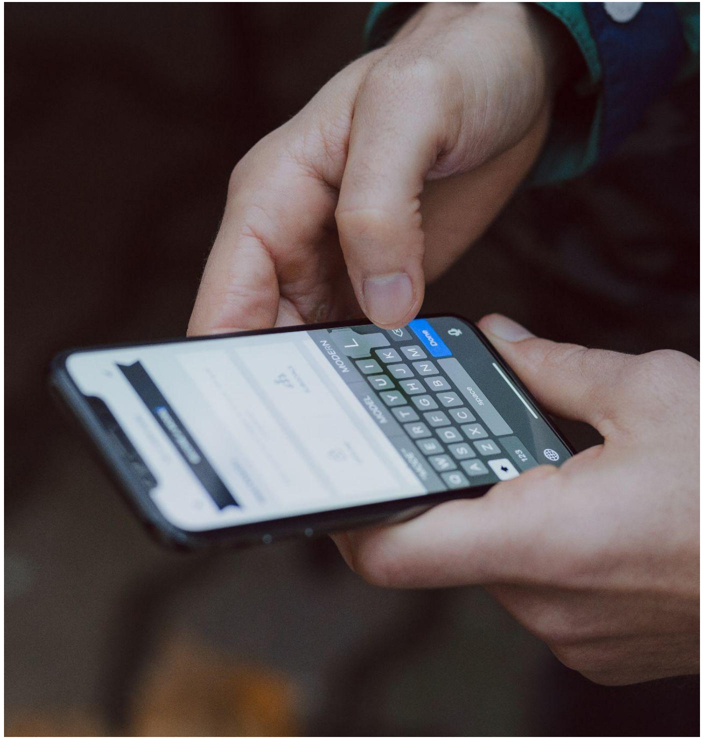
2. Cognition adapts

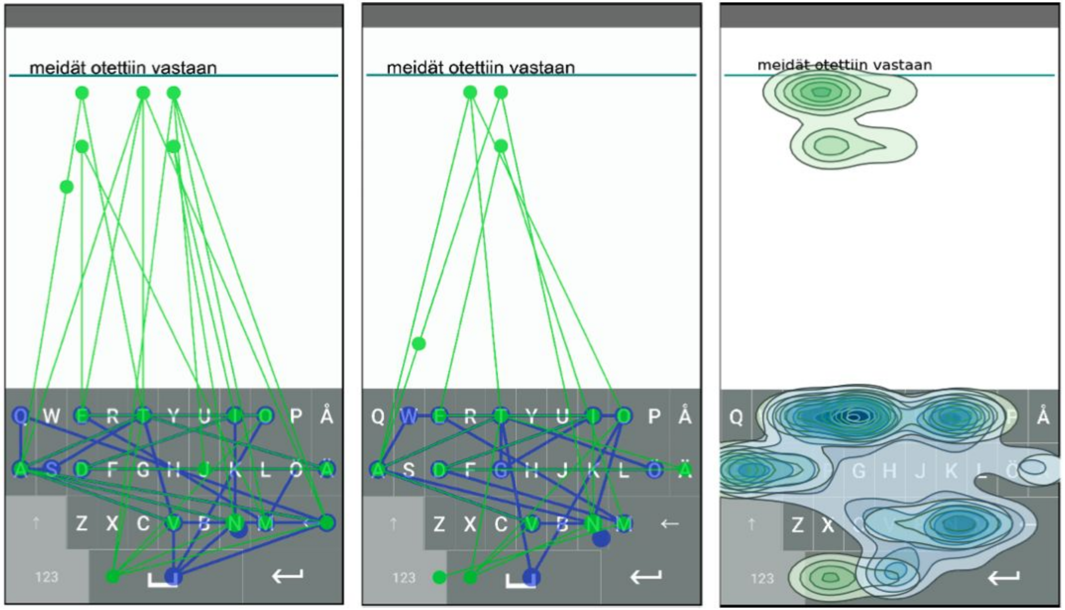
- Change is the norm in HCI: the systems people use and the way people carry out work keeps changing all the time
- Our cognitive, motor, and perceptual processes need to adapt, constantly
- Old beliefs need to be updated and new ones formed
- Old plans need to be updated and new ones formed
- Cognition is not simply passively processing information from external environments and reacting to it but actively finding ways to function better

Example: Adaptive eye movements in HCI



Example: Typing





(a) Model typing, with some errors and no automatic error-correction

(b) Model typing, with no errors and with automatic error-correction

(c) Model typing, no errors and no automatic error-correction

3. Cognition learns

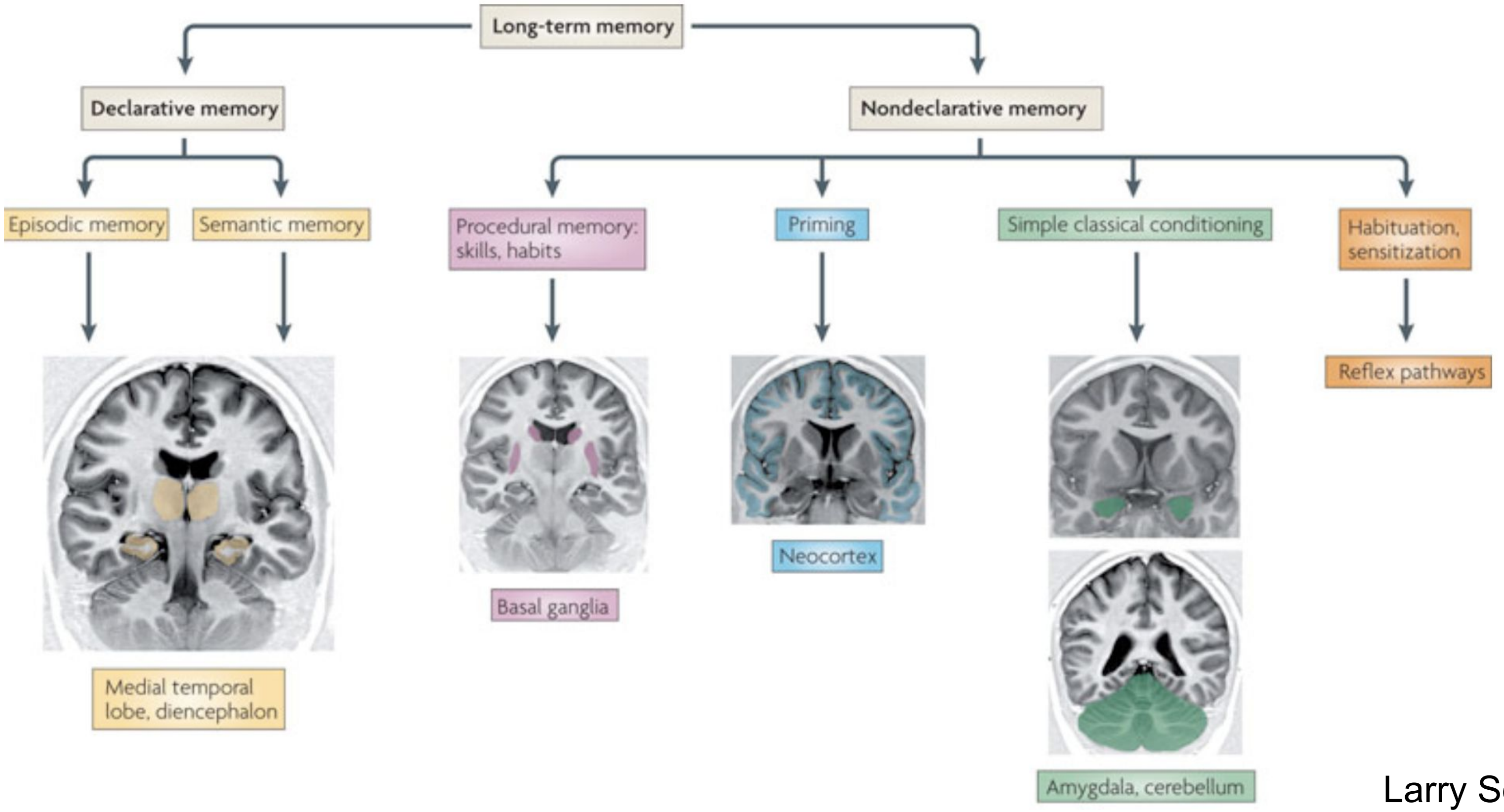
Computers are complex but opaque systems

- We need internal representations to control them
- We have multiple memory systems to that end (next slide)

Cognition learns to use external aids

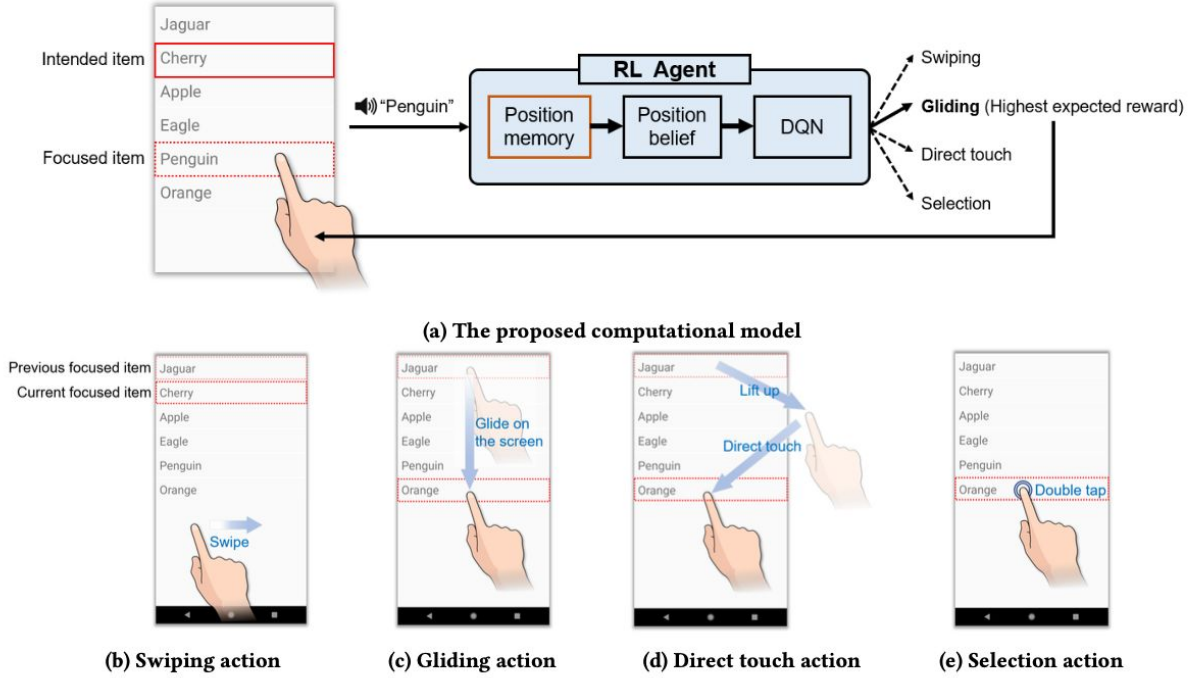
- We use notes, calculators, browsers to augment our abilities
- Over time, such dependencies affect the way we use cognition in interaction
- For example, since the uptake of the graphical user interface, which relies more heavily on visual recognition, we have less need to use our long-term memory to store computer commands that would be typed in a command prompt

Multiple memory systems: Multiple ways to learn



Larry Squire

Example: Menu interaction of non-sighted users



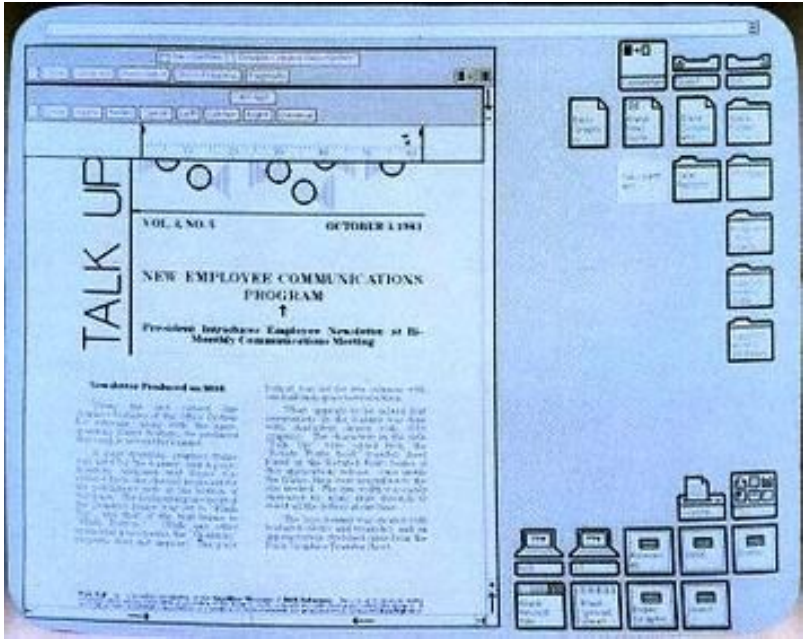
Memory for positions of menu items is created via interactions. The reliability of this memory affects the menu selection strategy users prefer.

Li et al. CHI2023

4. Cognition computes based on internal representations

Cognition can reason about things that are not directly perceivable. Internal models of reality that can be used to rason, formulate goals and plans

- E.g., metaphors help users understand what to expect from a UI
- The desktop metaphor uses spatial concepts that are rooted in our everyday physical experiences



Xerox Star used the desktop metaphor

5. Cognition is limited

- Visual attention is spatially limited
  - People can extract more information from the foveal region and less from the periphery
- Working memory is capacity-limited
  - We can only keep a few mental representations active in our mind
  - The typical working memory capacity that can be simultaneously maintained active in our mind is thought to be about 2–4 items
- Forgetting occurs in long-term memory
  - We cannot remember everything we have experienced and as a result we forget details of things we have attended
- Our capacity for abstract reasoning and planning is limited
  - We can only look a few steps ahead.

In-class task: Measuring cognitive capacity

https://www.psytoolkit.org/experiment-library/

List of experiments (alphabetical)

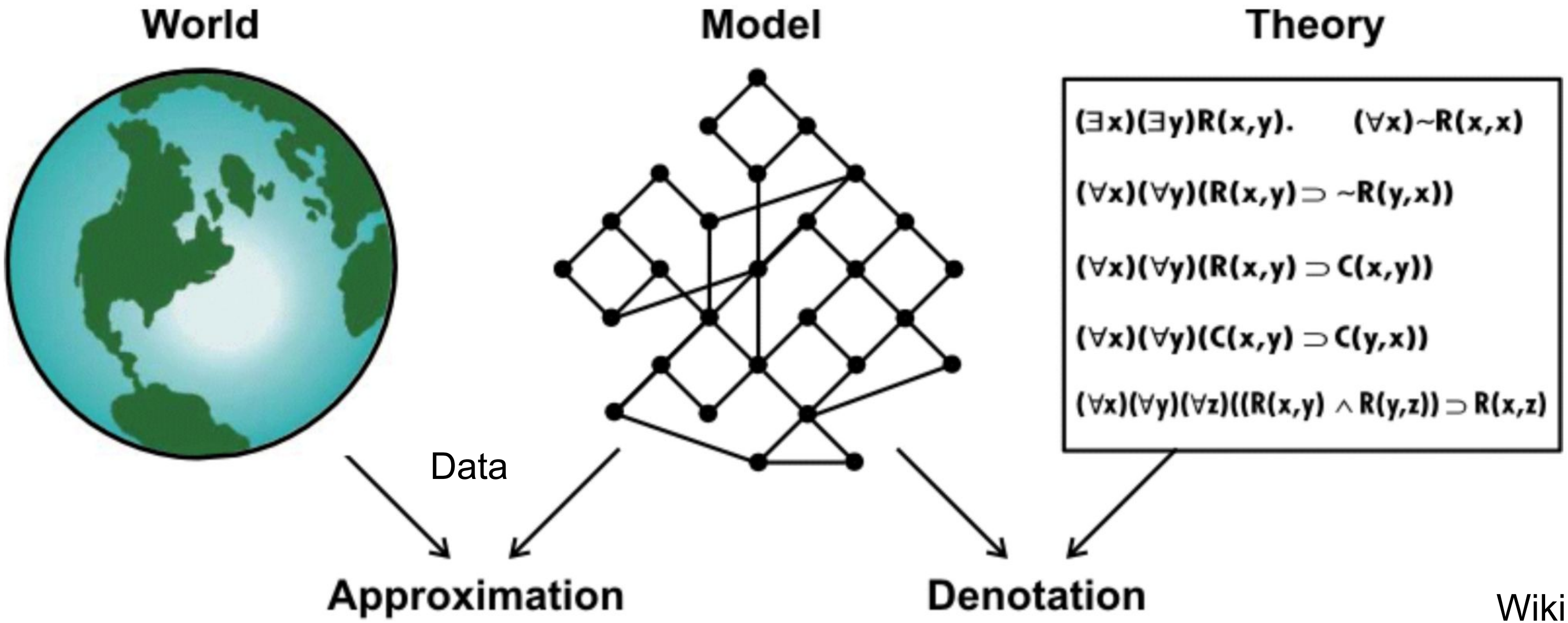
- ABBA task (reversed compatibility)
- Attentional blink paradigm
- Corsi **backward** block test (short term memory measure)
- Cueing, Posner Task
- Deary-Liewald Task
- Corsi and Digit span memory tasks
  - Corsi block test (short term memory measure)
  - A slightly different implementation of the same Corsi Test
  - Digit span test
- Endogenous vs exogenous cueing
- Fitts's Law
- Eriksen Flanker Task
- Go/No-go task
- Implicit Association Task (IAT)
- Inhibition Of Return (IOR)
- Iowa Gambling Task
- Lexical Decision Task
- Mackworth Clock Task to measure sustained vigilance
  - Mackworth Clock Task (basic)
  - Mackworth Clock Task (fancier)

Introduction

2. Modelling

Models connect theories and data via some mechanism

“A scientific model seeks to represent empirical objects, phenomena, and physical processes in a logical and objective way. All models are in simulacra, that is, simplified reflections of reality that, despite being approximations, can be extremely useful.”



Wikipedia

Modeling is one type of knowledge needed in HCI

Guidelines are practical rules of thumb that summarize best practices or recommendations

- E.g., Shneiderman's Eight Golden Rules

Concepts characterize phenomena and their underpinning mechanisms

- E.g., turn-taking as a concept for understanding human conversations

Taxonomies systematize several concepts

- E.g., morphological analysis of design spaces

Theories consist of constructs and relations among those constructs

- E.g., self-determination theory outlines the motives of people

Models are simplifications of reality. They link concepts using some systematic formalism like code

- E.g., motor control models that predict user performance in pointing

Types of models

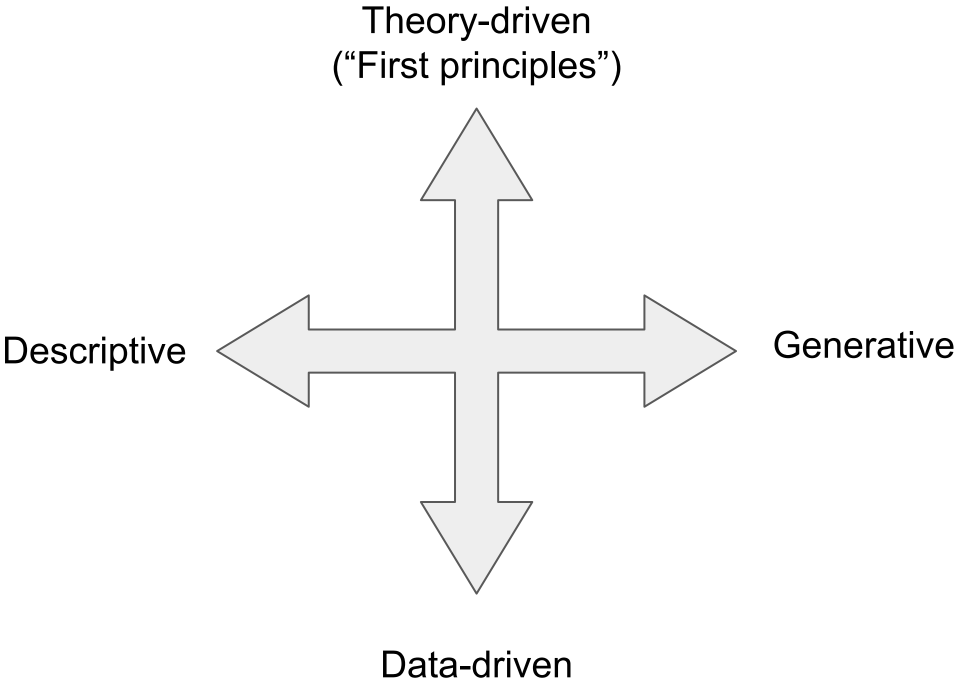
Verbal models

Logic-based models

Mathematical models

Simulation models

Machine learning models

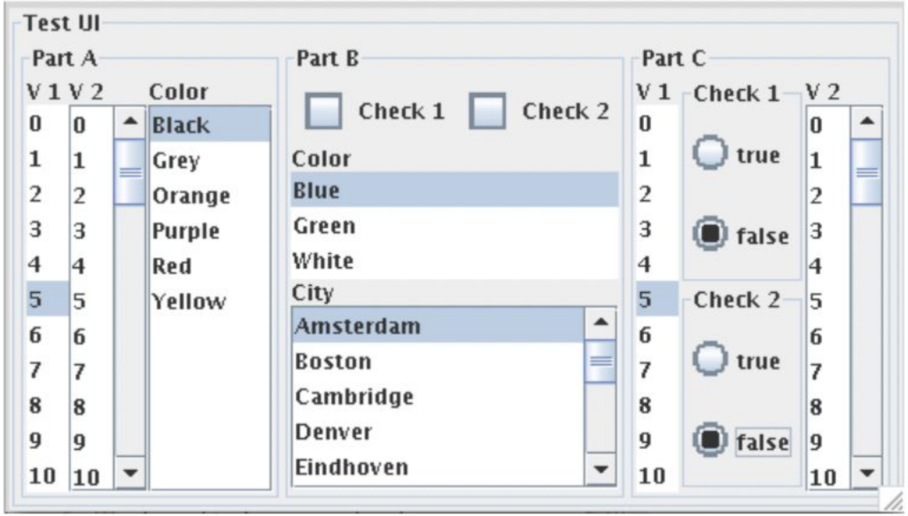


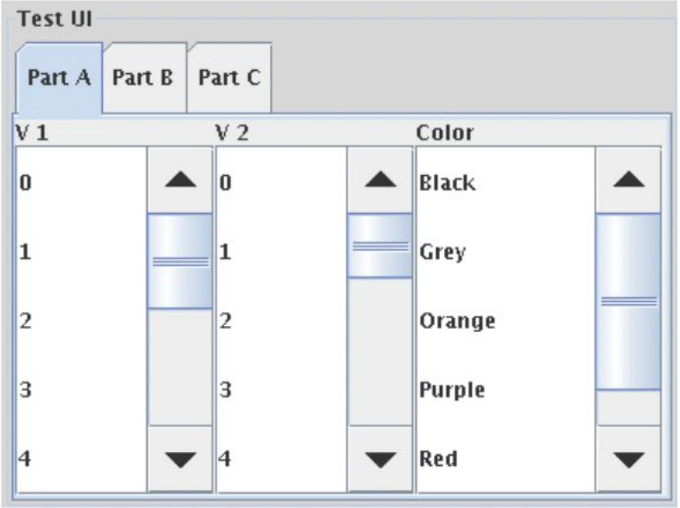
What models do in HCI

1. Direct what to pay attention to
2. Explain empirical findings
3. Predict behavior
4. Test theories
5. Inform design decisions
6. Help designers explore design spaces
7. Drive an intelligent system

Modeling is hard but fruitful

Example: SUPPLE uses motor control models to optimize GUIs for people with motor and vision problems





Work by Krzysztof Gajos

What makes a good model

General qualities

1. Plausibility: Consistent with contemporary evidence
2. Interpretability: Modeling assumptions are transparent and meaningful
3. Verifiability: Assumptions can be empirically tested
4. Flexibility: Can be adapted to new study new behaviors
5. Parameter identifiability: Parameters are can be fit to data meaningfully
6. Parsimoneity: Only necessary assumptions are made

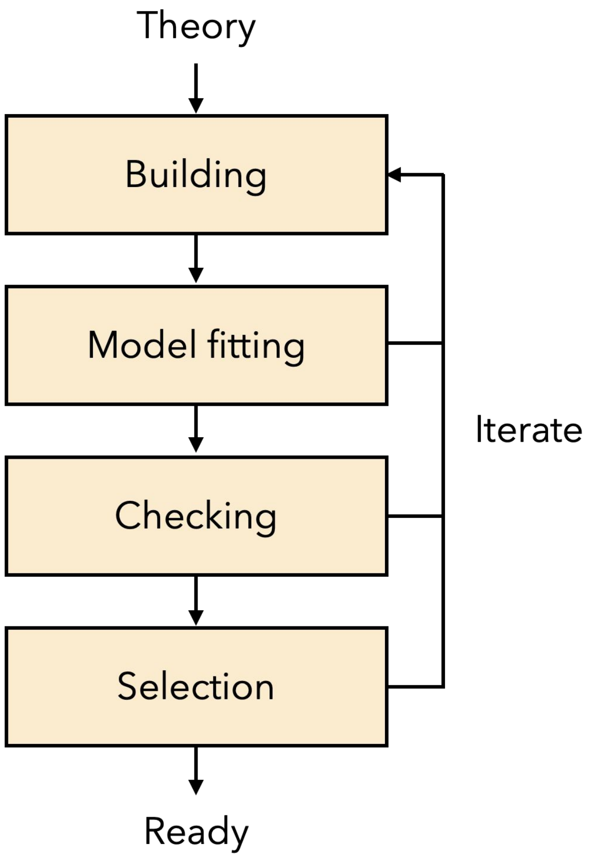
Four powers we need in HCI

7. Descriptive power: Provides a good account of observed data
8. Explanatory power: Explains observed behavior
9. Predictive power: Predicts unseen behavior
10. Design power: Informs decisions and choices in design

How to build a model: A workflow

Building a good model takes more than programming!

1. Derive ideas from theory
2. Build a model in code
3. Fit or train the model
4. Check and verify the model
5. Compare the model against other models
6. Iterate!



Introduction

3. Cognitive models

What is a cognitive model?

Cognitive models are computer programs that:

Causality: Include theory-informed mechanisms for producing behavior and thought

Simulation: Produce behavior and thought moment-by-moment as it unfolds in some situation.

Adaptation: Are able to adapt behavior to the benefit of the agent.

Verifiability: Make testable predictions.

David Marr’s levels of cognition

Computational theory	Representation and algorithm	Hardware implementation
What is the goal of the computation, why is it appropriate, and what is the logic of the strategy by which it can be carried out?	How can this computational theory be implemented? In particular, what is the representation for the input and output, and what is the algorithm for the transformation?	How can the representation and algorithm be realized physically?

Figure 1–4. The three levels at which any machine carrying out an information-processing task must be understood.

Next up

Mathematical vs. computational models

Cognitive architecture models

Deep learning

Large language models

Reinforcement learning

Mathematical vs. simulation models

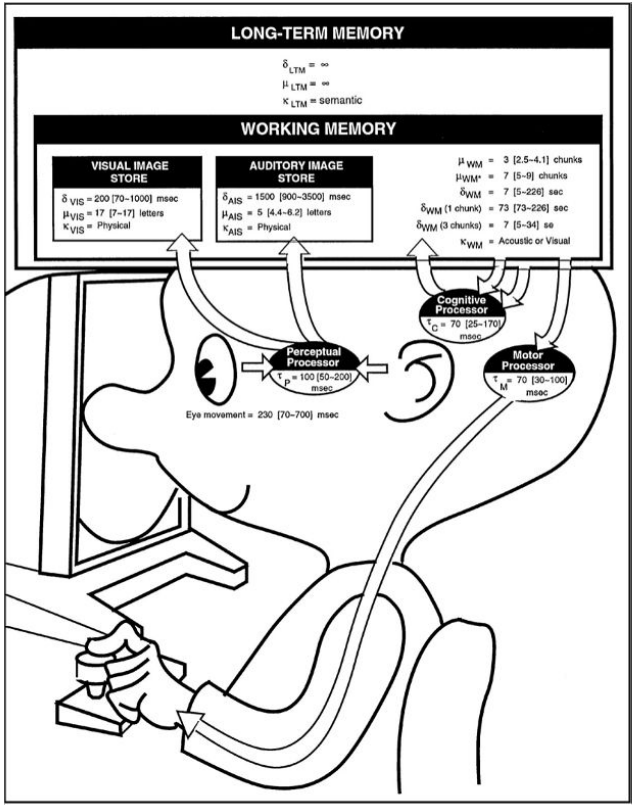
Statistical models (e.g., Fitts’)	Simulation models
Statistical relationships	Statistical and causal
Equations	Computer programs
Predict outcomes of interaction	Predict outcomes and process
Few parameters	Many or lots of parameters
Parameters fit to data	Both fitting and learning
Fast	Slow

Cognitive architecture models

1. Cognition consists of the transmission and processing of information through a series of stages carried out in order to achieve a goal
2. Higher mental processes are understood as the collective action of elementary process. Processes occur independently and they can be isolated
3. Human cognition has a limited capacity for storing and transmitting information

Examples: GOMS, EPIC, ACT-R, SOAR

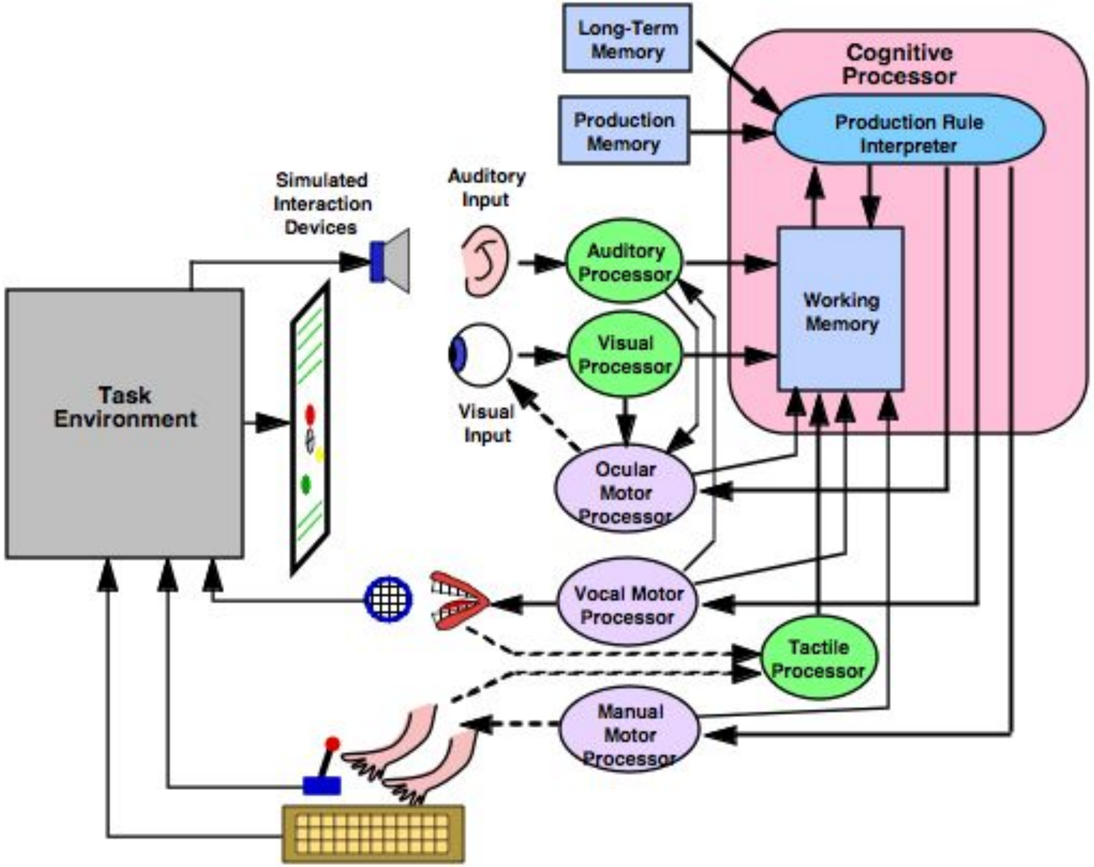
Example: GOMS



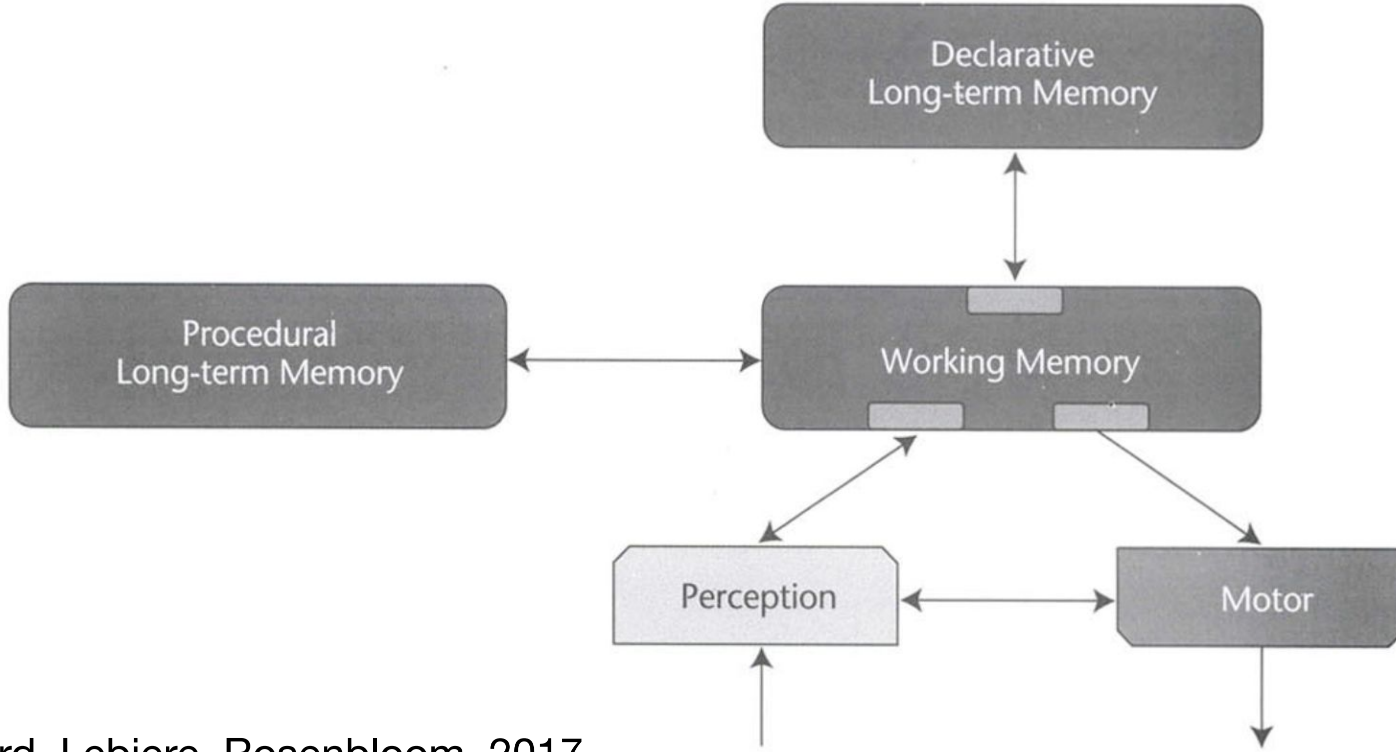

```
GOAL: DELETE-FILE
.    GOAL: SELECT-FILE
.    .    [select:  GOAL: KEYBOARD-TAB-METHOD
.    .              GOAL: MOUSE-METHOD]
.    .    VERIFY-SELECTION
.    GOAL: ISSUE-DELETE-COMMAND
.    .    [select*: GOAL: KEYBOARD-DELETE-METHOD
.    .              .    PRESS-DELETE
.    .              .    GOAL: CONFIRM-DELETE
.    .              GOAL: DROP-DOWN-MENU-METHOD
.    .              .    MOVE-MOUSE-OVER-FILE-ICON
.    .              .    CLICK-RIGHT-MOUSE-BUTTON
.    .              .    LOCATE-DELETE-COMMAND
.    .              .    MOVE-MOUSE-TO-DELETE-COMMAND
.    .              .    CLICK-LEFT-MOUSE-BUTTON
.    .              .    GOAL: CONFIRM-DELETE
.    .              GOAL: DRAG-AND-DROP-METHOD
.    .              .    MOVE-MOUSE-OVER-FILE-ICON
.    .              .    PRESS-LEFT-MOUSE-BUTTON
.    .              .    LOCATE-RECYCLING-BIN
.    .              .    MOVE-MOUSE-TO-RECYCLING-BIN
.    .              .    RELEASE-LEFT-MOUSE-BUTTON]

*Selection rule for GOAL: ISSUE-DELETE-COMMAND
  If hands are on keyboard, use KEYBOARD-DELETE-METHOD,
    else if Recycle bin is visible, use DRAG-AND-DROP-METHOD,
    else use DROP-DOWN-MENU-METHOD
```

Example: EPIC

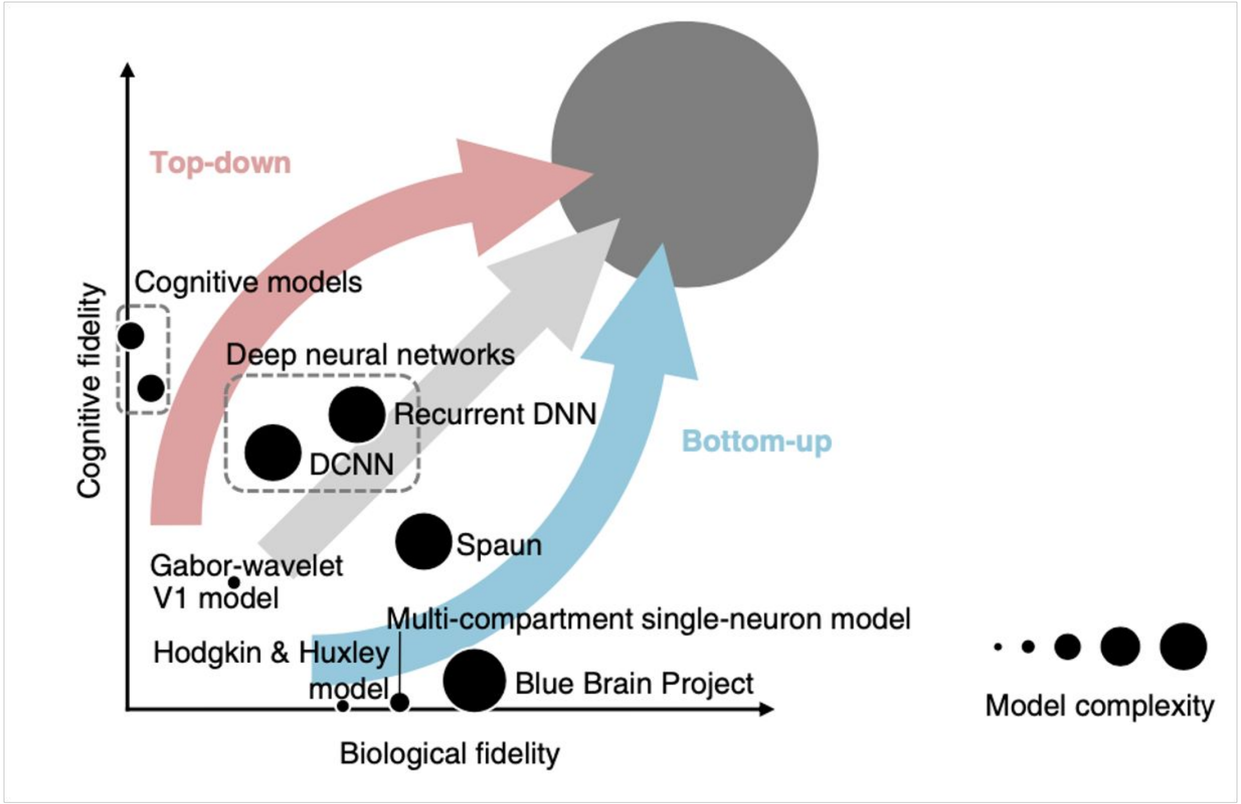


A standard model of cognition



Laird, Lebiere, Rosenbloom, 2017

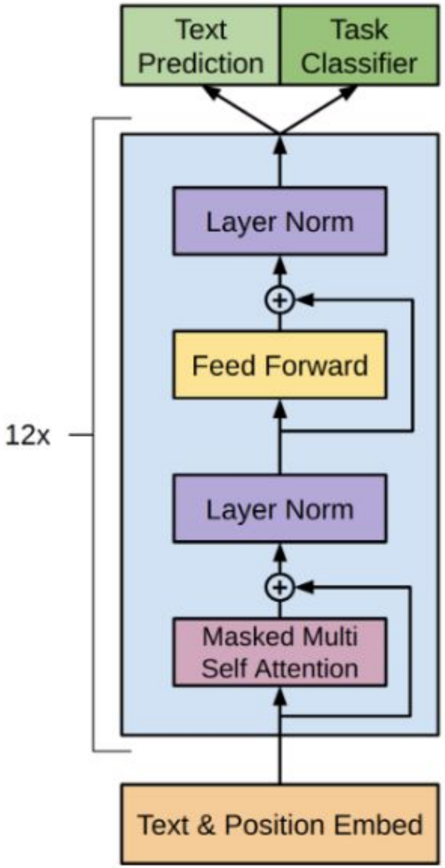
Deep learning as a model of cognition

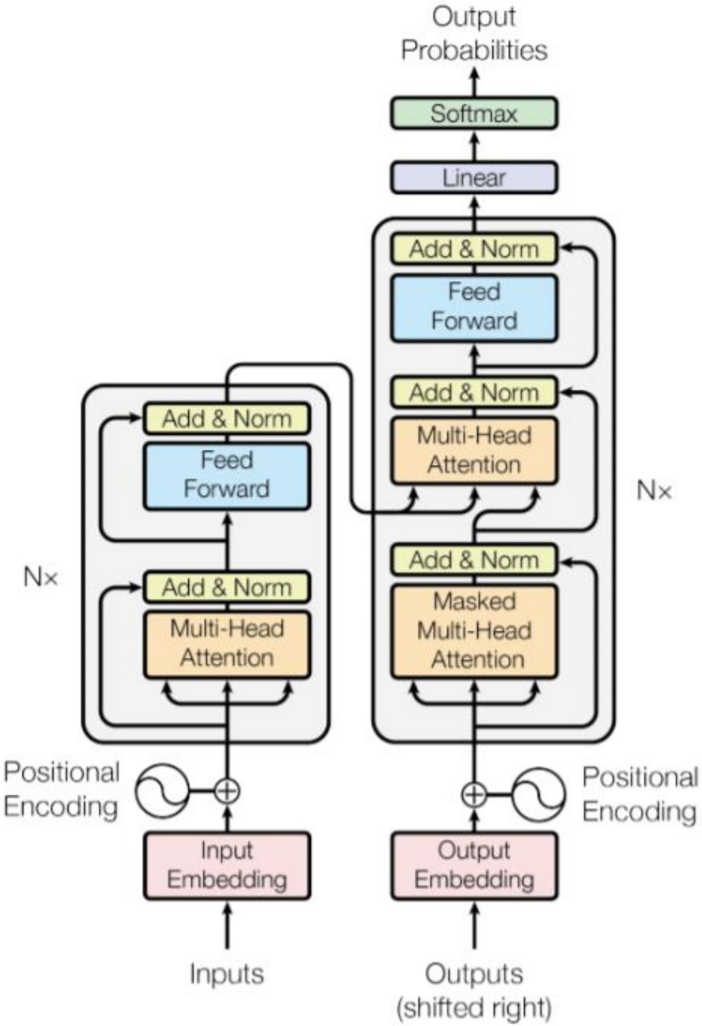


Douglas and Kriegeskorte 2019

Transformers

Supervised learning
Self-supervised learning





Deep learning and the predictive coding hypothesis

The predictive coding hypothesis

- The brain constantly attempts to predict future events before they occur
- These predictions are revised and updated via error signals generated upon comparison of predictions with observations

A conceptually simple and plausible hypothesis of how the brain works

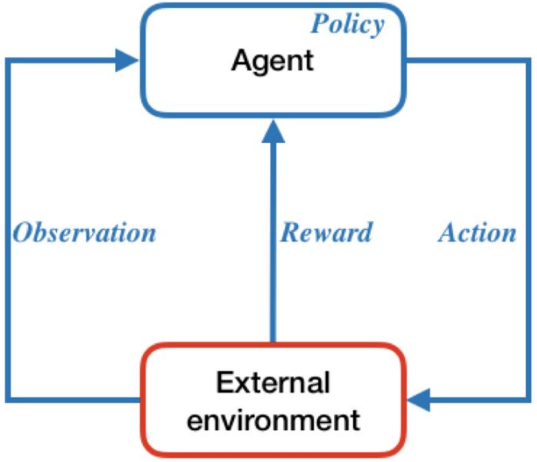
- Broader than language
- Early evidence from electrophysiological studies show that surprisal predicts brain activation

Prediction error is also integral to how LMs work

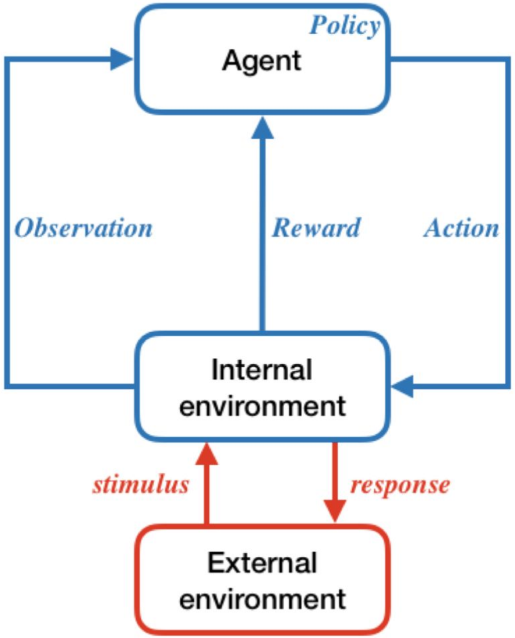
- On-going research: Is next word prediction also what the brain does?

Reinforcement learning models

Cognition is about choosing sequences of organismically beneficial actions.

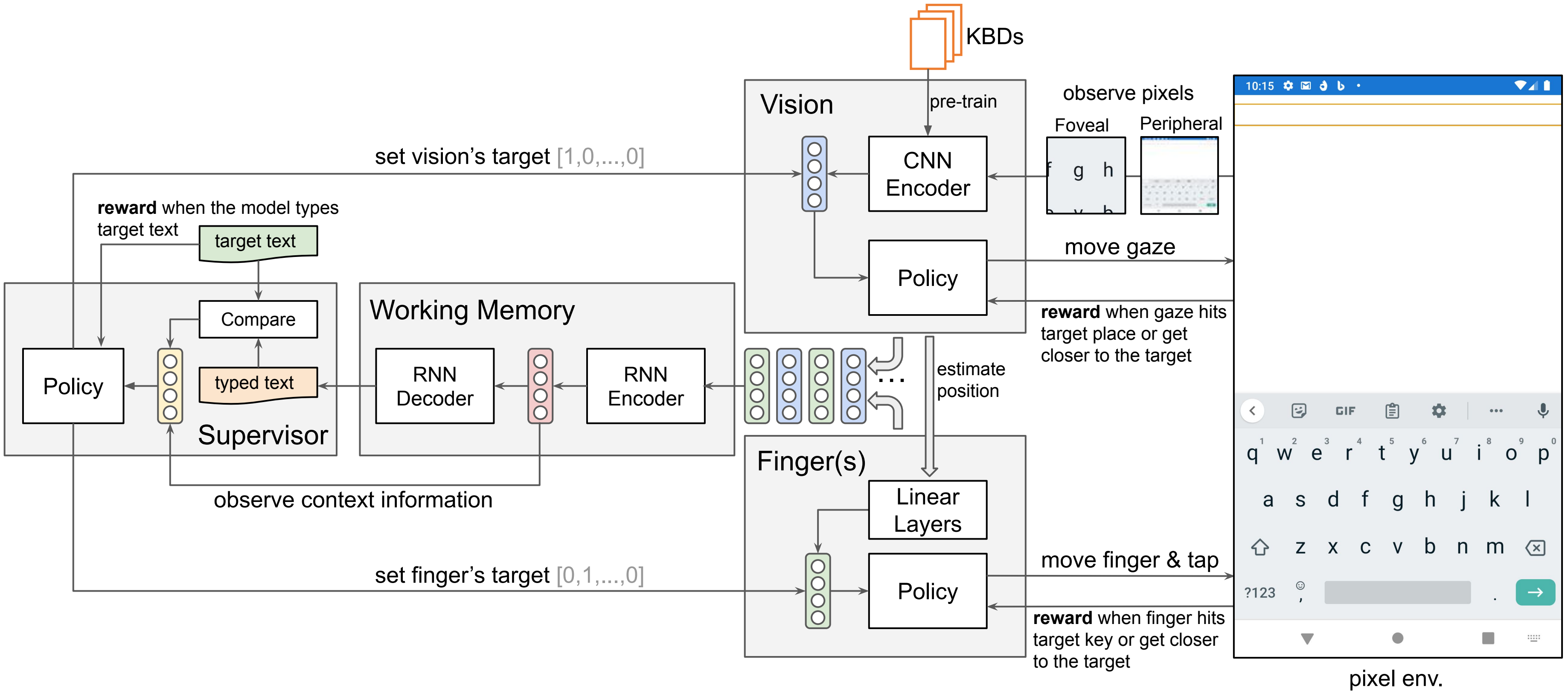


Standard reinforcement learning



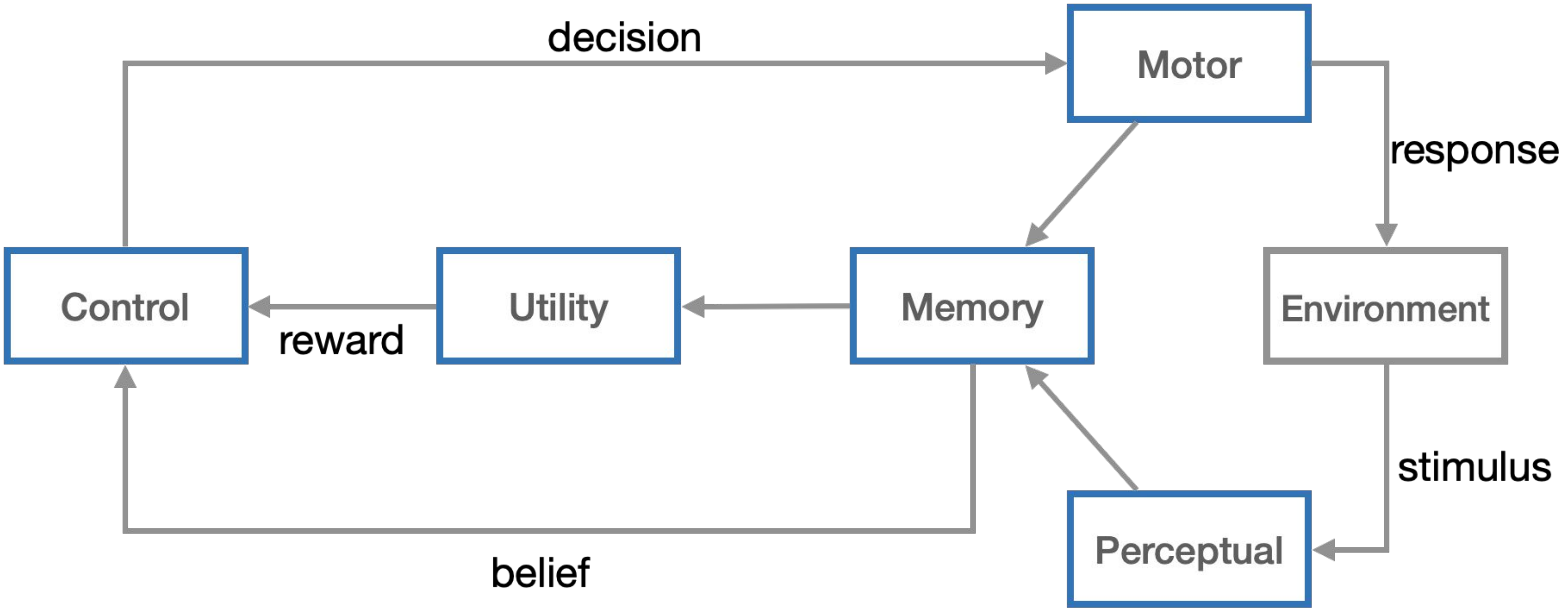
RL as a model of human cognition

Deep RL = Deep learning and RL



On-going work on "Deep typist"

Cognitive architecture used in Modules 3 and 4



Introduction

Discussion

"It doesn't matter how beautiful your theory is, it doesn't matter how smart you are. If it doesn't agree with the experiment, it's wrong."

--Richard P. Feynman

http://mattbors.tumblr.com/post/100671142990/the-ultimatum-game-more-comics-at-the-nib

Sources

Some slides based on manuscript for the textbook: Hornbaek, K., Kristensson, P.O. and Oulasvirta, A. 2023. *Introduction to Human-Computer Interaction*. Under contract with Oxford University Press.