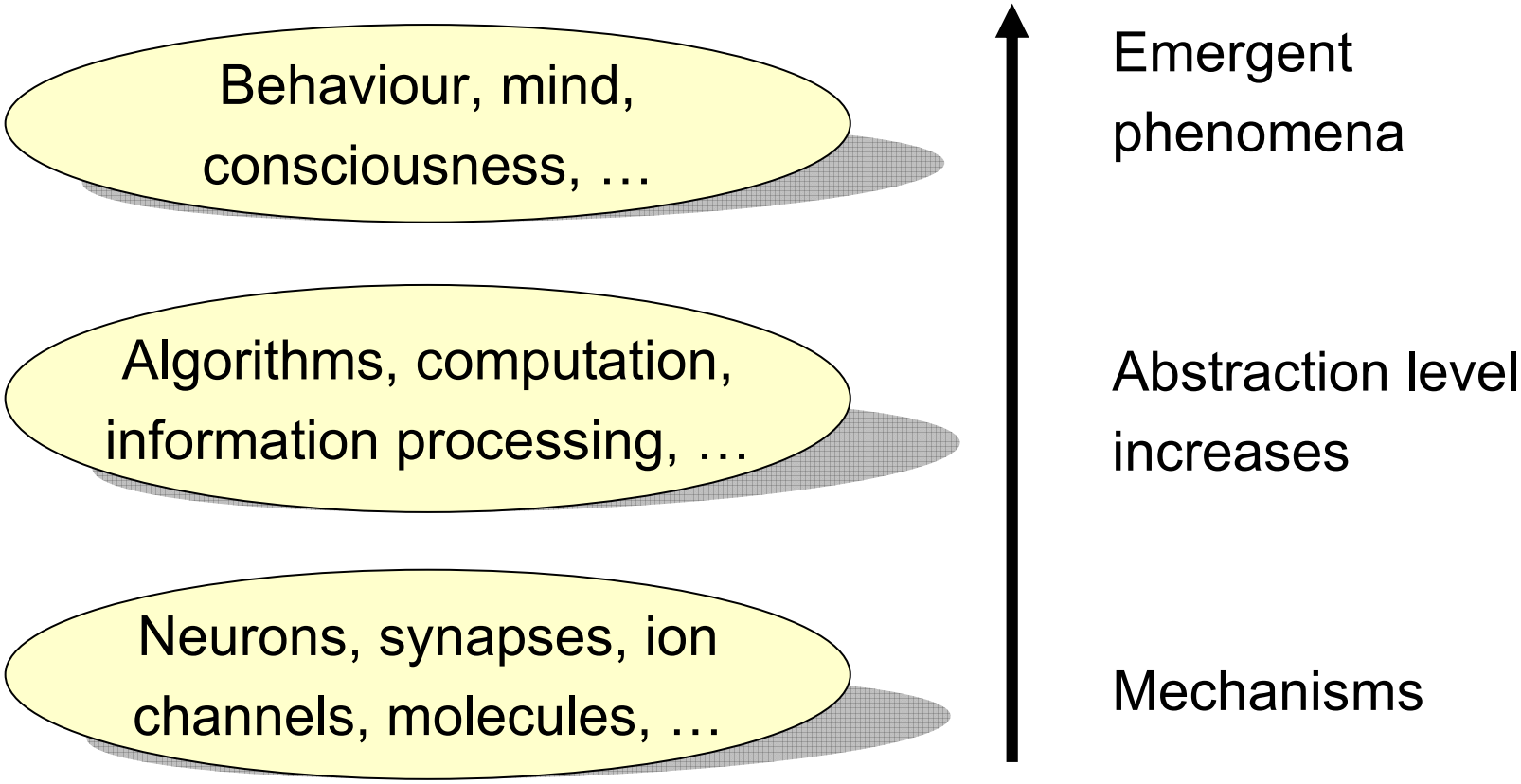
The Engine of Thought—
A Bio-Inspired Mechanism for Distributed Selection of Useful Information

Harri Valpola

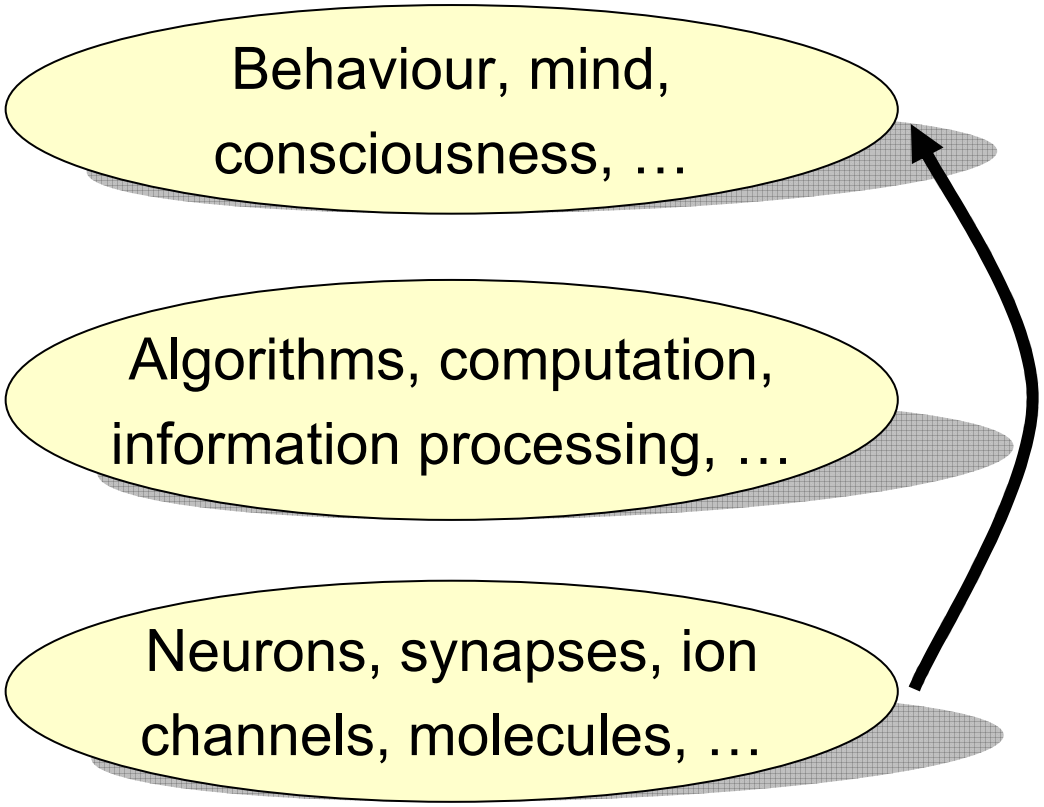
Computational neuroscience group

Helsinki University of Technology

Levels of explanation

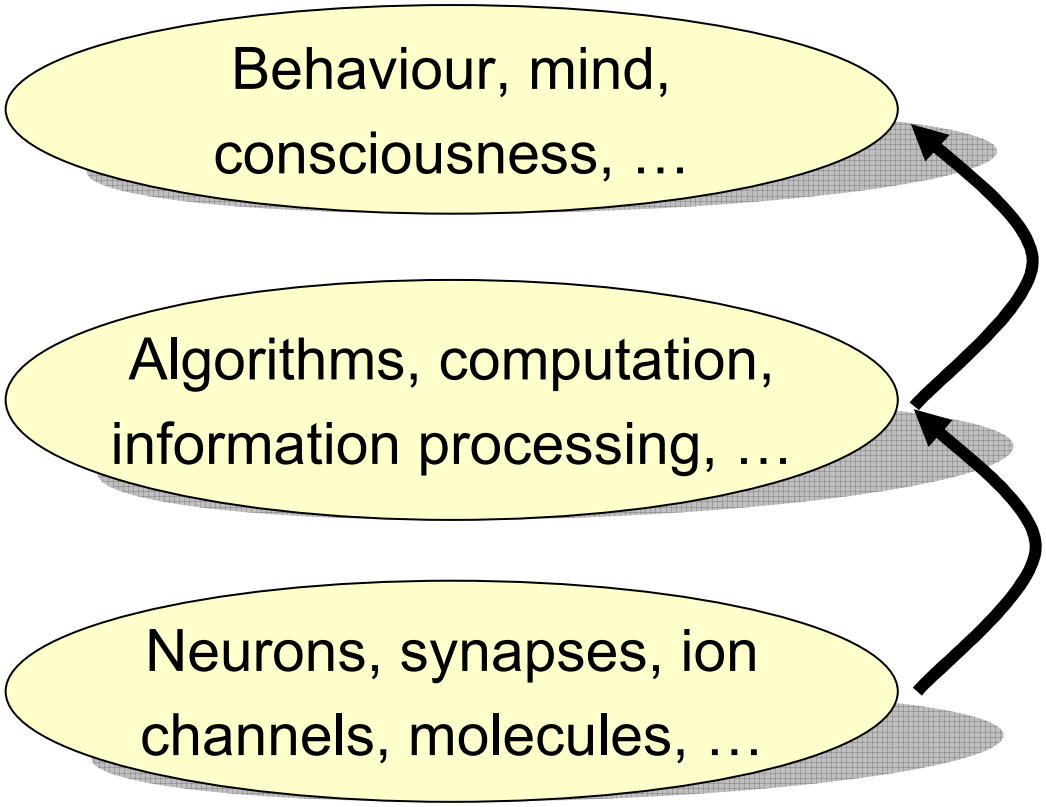


Levels of explanation



NCC: neural correlates of consciousness?

Levels of explanation



Information-processing correlates of consciousness

Neural correlates of information processing

"… an algorithm is likely to be understood more readily by understanding the nature of the problem being solved than by examining the mechanism (and the hardware) in which it is embodied."

—David Marr (1982) *Vision*, Ch. 1.2

Problems to be solved by the brain

- Make decisions
  - The ultimate purpose of the brain: control movement
- Perception
  - Extract information about the world
  - Needed for control, decision making
  - Find regularities, develop representations
- Simulate the future
  - Needed for decision making

Bayesian theory says:

- Decisions are based on
  1. Beliefs (measured by probability)
  2. Utilities
- The recipe:
  1. Evaluate the probabilities of all possible states of the world (probabilistic inference)
  2. Evaluate the probabilities of all outcomes for each and every potential action (probabilistic inference)
  3. Choose the action which maximises the expected utility
- This is optimal if there are no restrictions on the available computational resources

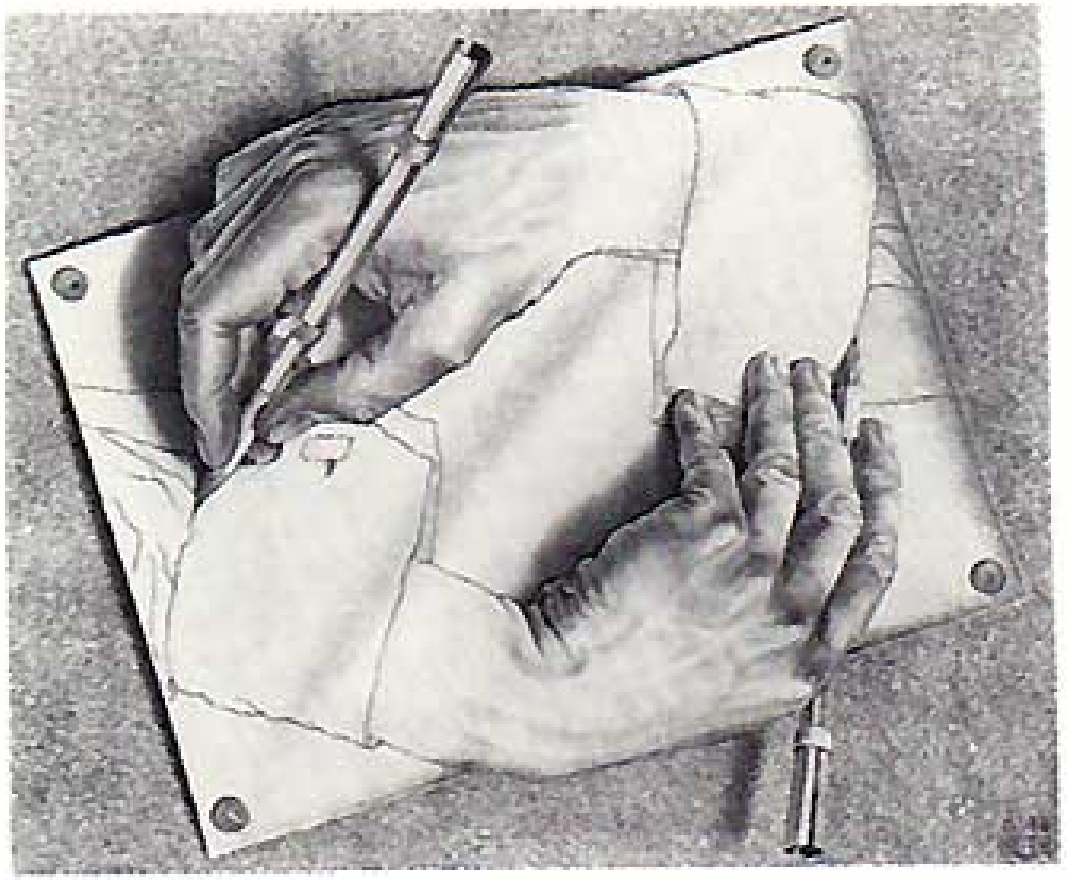
Key problem:
How to select useful information?

- But… computational resources *are* restricted →
- It is impossible to consider all the states and actions →
- It is necessary to select information in order to make decisions

- Selection is a type of decision, in other words:
- In order to decide we need to decide… Infinite regress!

Let's study the solution adopted by the brain

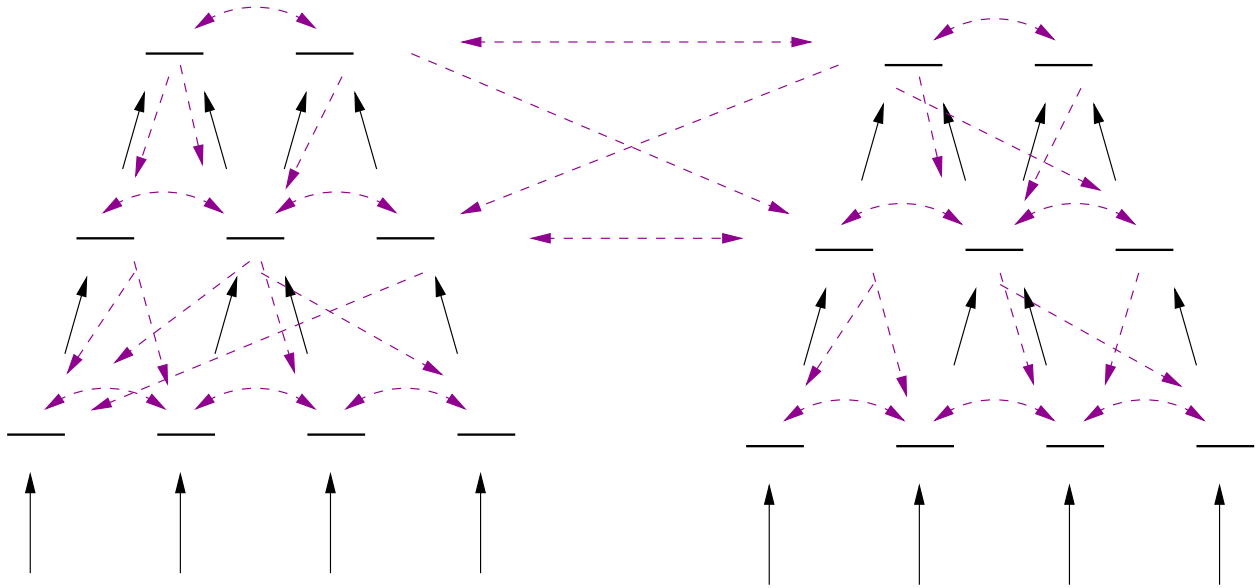
Neuroscience



Technology

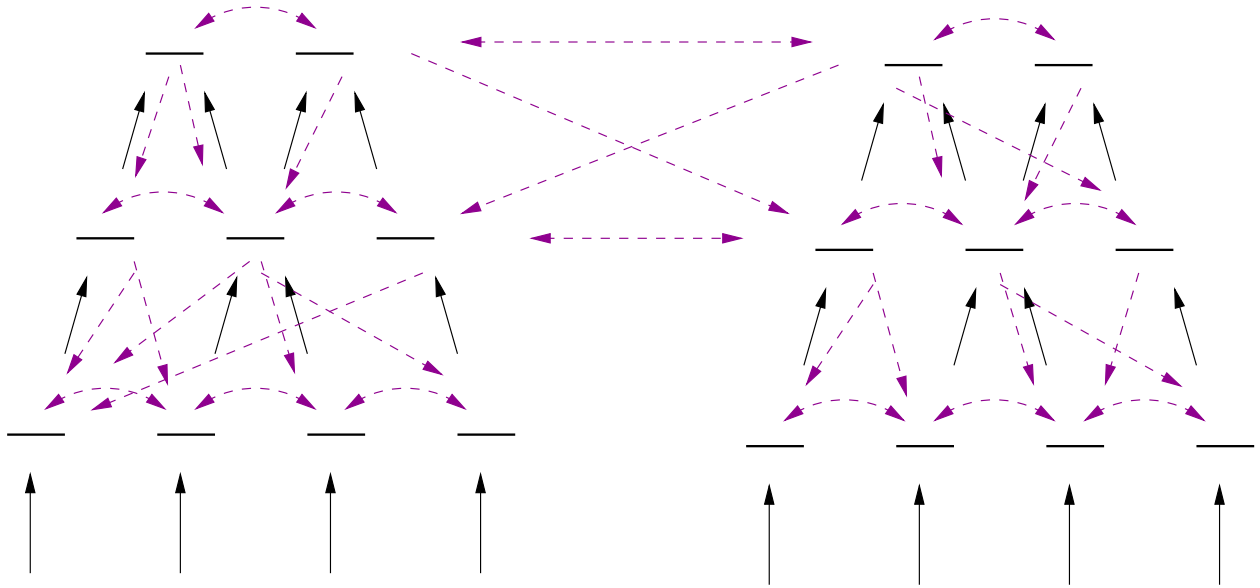
Hierarchy of areas

- The cerebral cortex is connected as a hierarchy of areas
- The representations get more abstract on higher levels



Brain's solution: distributed selection

- Each cortical area selects information to be represented
- Biased-competition model of attention: attention emerges from local selection and global communication



Visual input

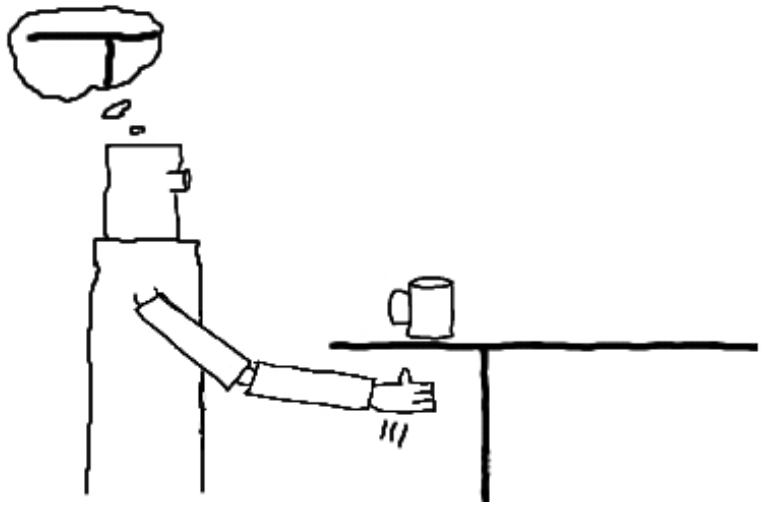
Auditory input

Attention and learning: selection on different timescales

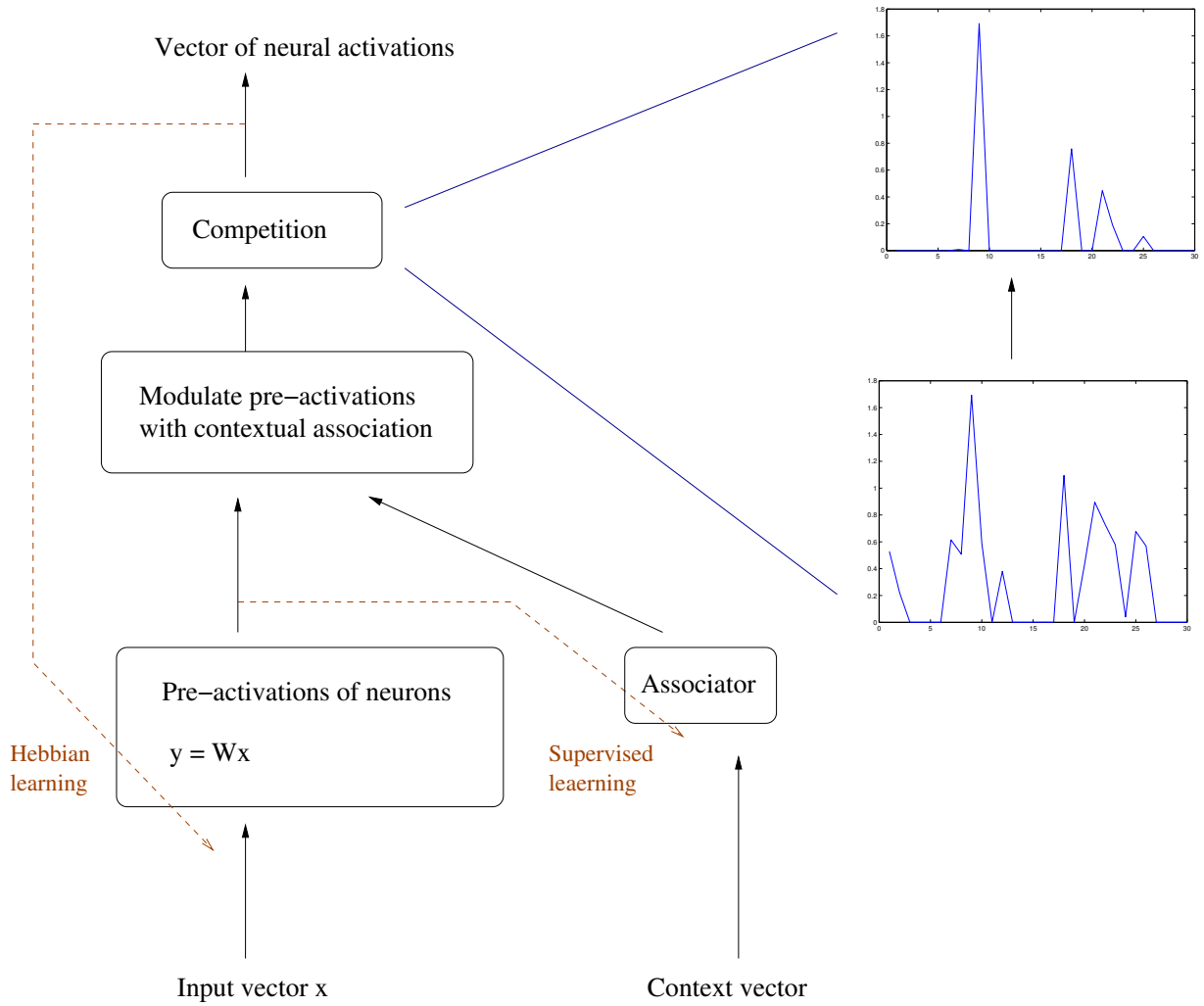
- Our work: biased competition + competitive learning
- Within the Bayesian framework, the only difference between perceptual inference and learning is the timescale
- Attention and learning in the cortex are intimately coupled
- Both are a form of selection, only timescales differ

The value of information

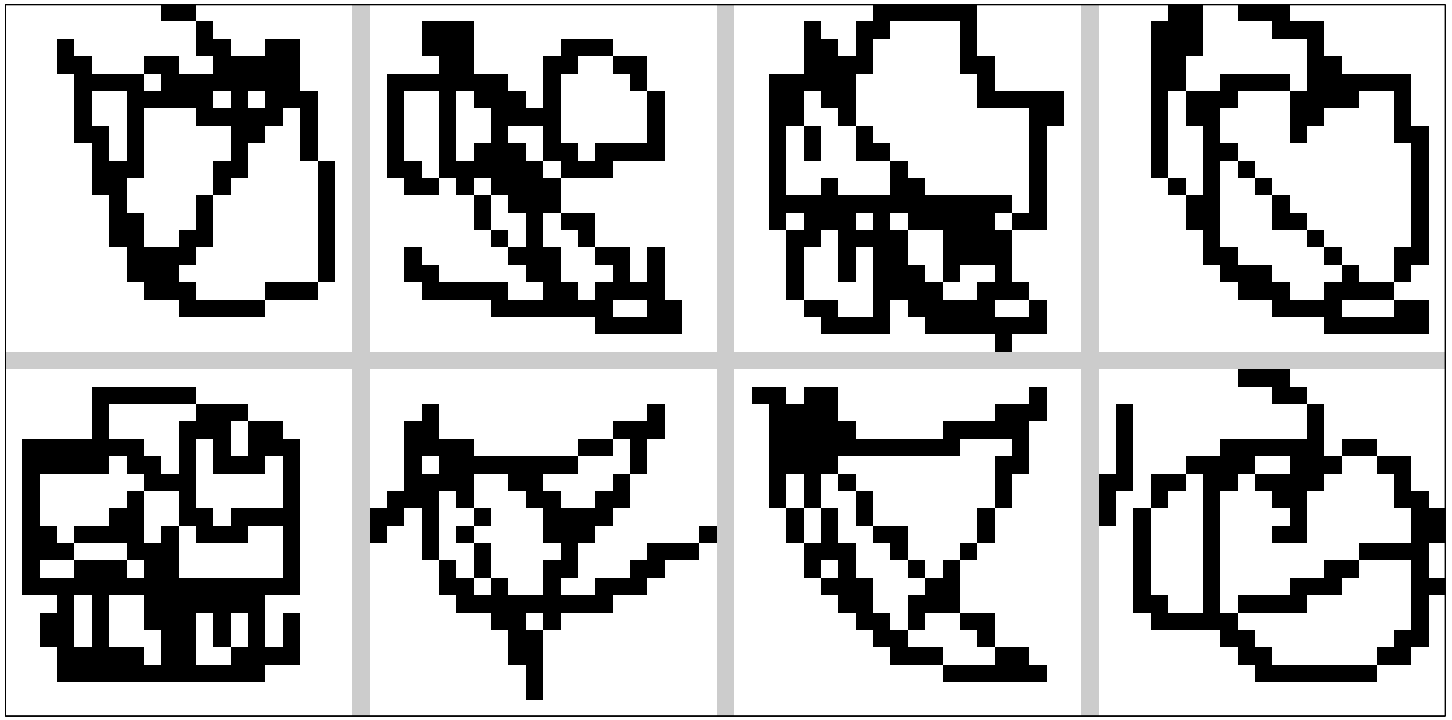
- Motor areas may be able to rely (at least partly) on global reward signals (reinforcement learning)
- Sensory areas or a large brain: credit assignment problem
- More specific but locally available information: predictive power or “are the others listening?”



A model of a cortical area

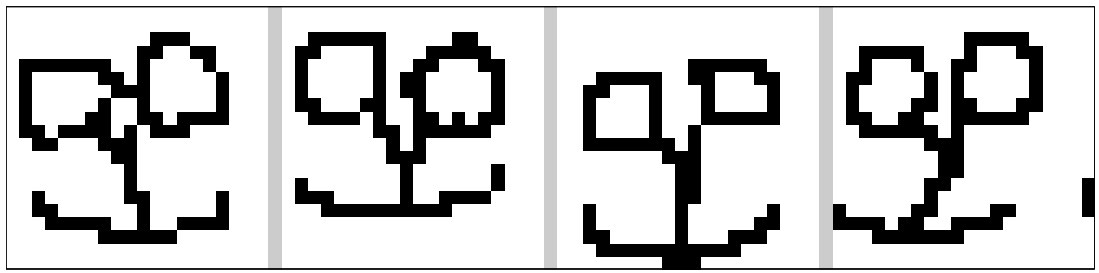


Results 1: data

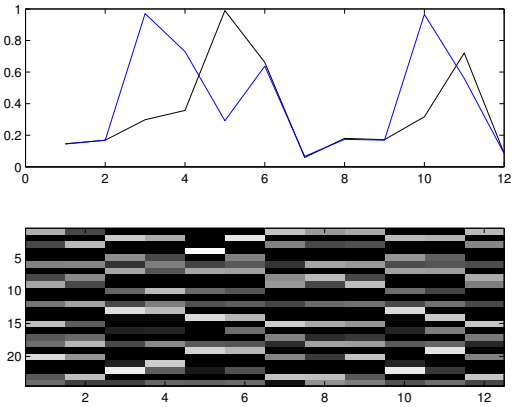


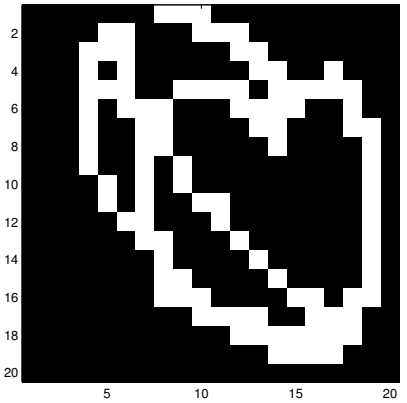
Results 2: abstract categories

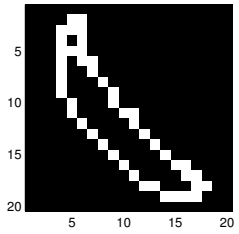
- Four samples of test data, each of which have activated the same coalition of neurons at the highest level
- This invariant recognition of abstract categories was achieved without any supervision even if the objects never appeared in isolation

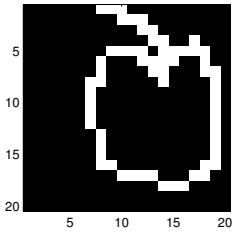


Results 3: switching attention









What about consciousness…?

- We are beginning to understand the information processing on the cortex (perception, attention, learning, imagination, decision making, …)
- Now we can ask how it relates to consciousness.

- My own answer: information processing (the list above) relates to consciousness as metabolism and replication relate to life
- But: most people need to meet a sufficiently intelligent machine to get convinced (a nice robot would do fine)