Building Brains for Robots

Harri Valpola

Computational neuroscience group

Department of biomedical engineering and computational science

Helsinki University of Technology

“The aeroplane will never fly.”

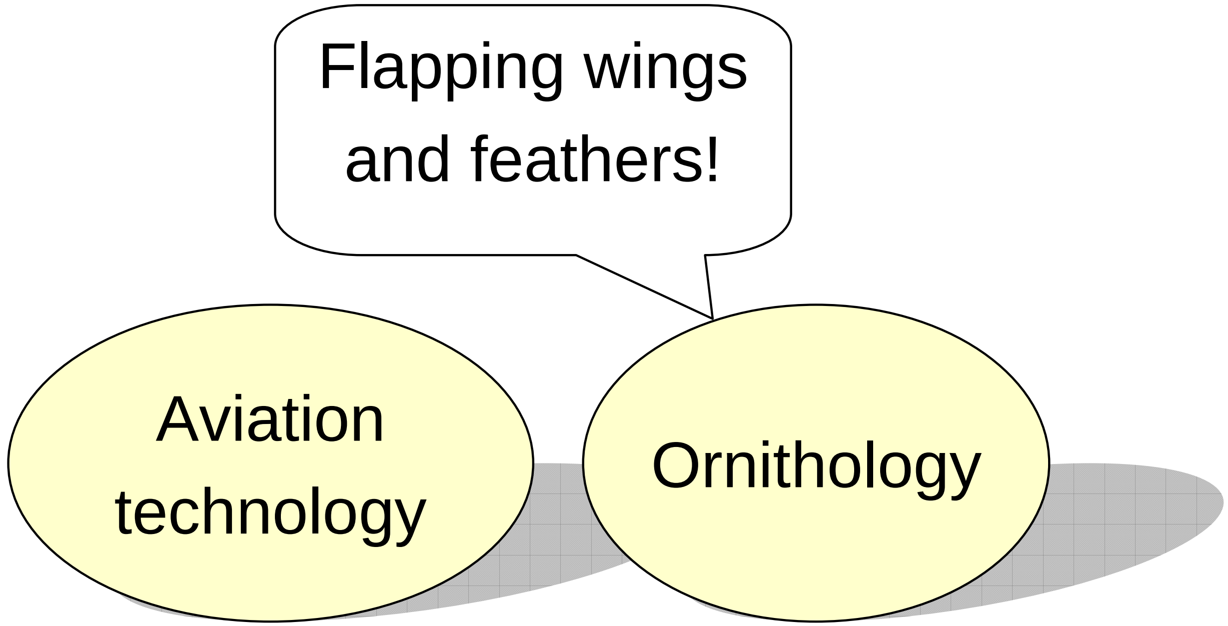
—Lord Kelvin, 1892

Wright brothers: the first flight on Dec. 17, 1903

“Heavier-than-air flying machines are impossible.”

—Lord Haldane, Minister of War, Britain, ***1907***

How to build a flying machine?

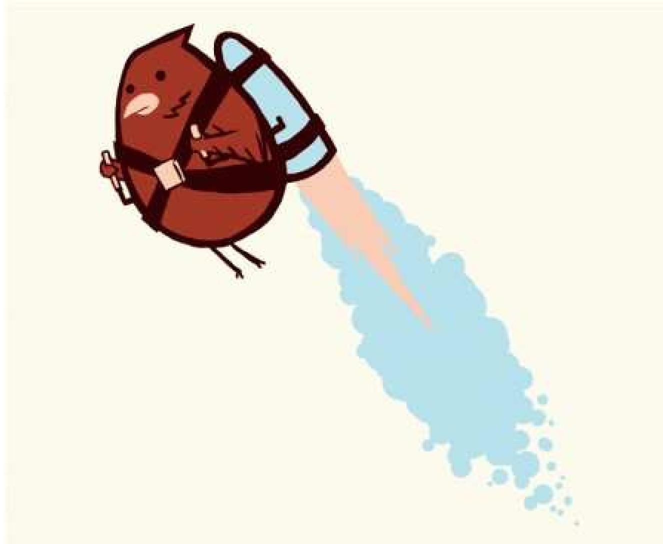


How to build a flying machine?

Aviation technology

Ornithology

How do birds fly?

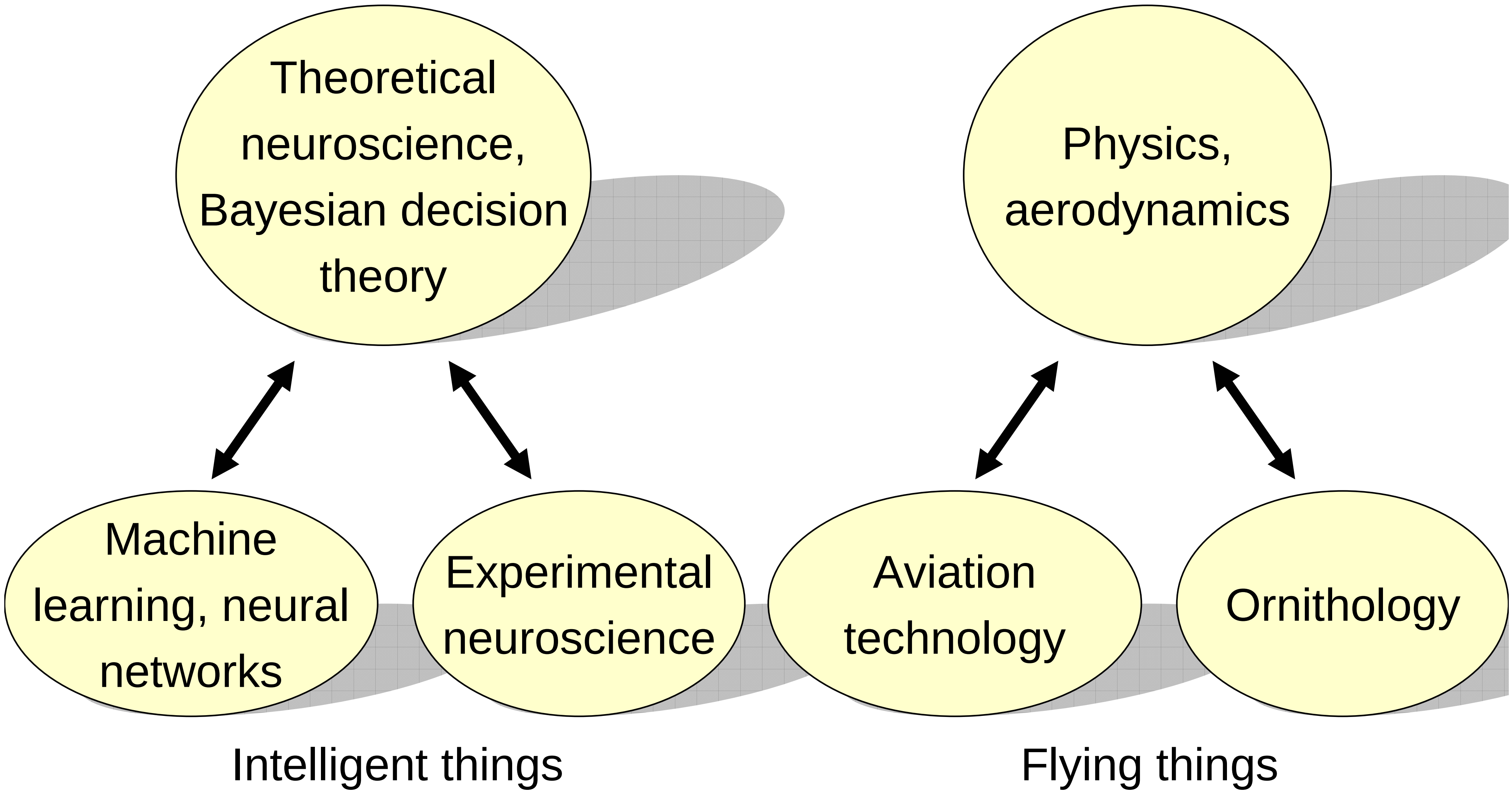


Rockets or balloons!

Aviation technology

Ornithology

Bridging the gap



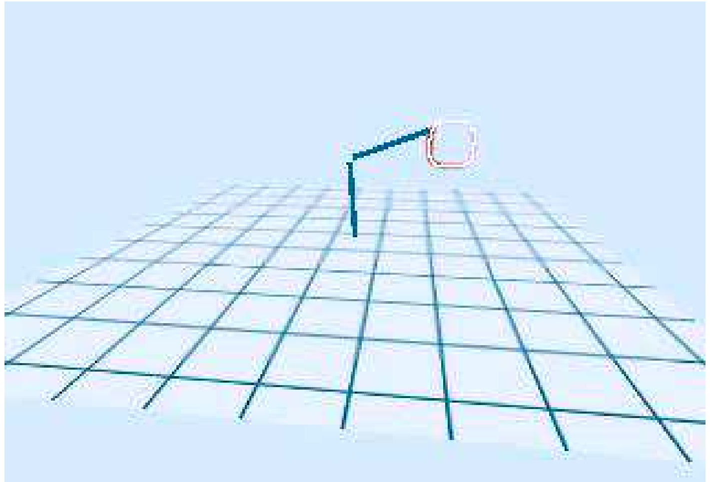
Computational neuroscience group

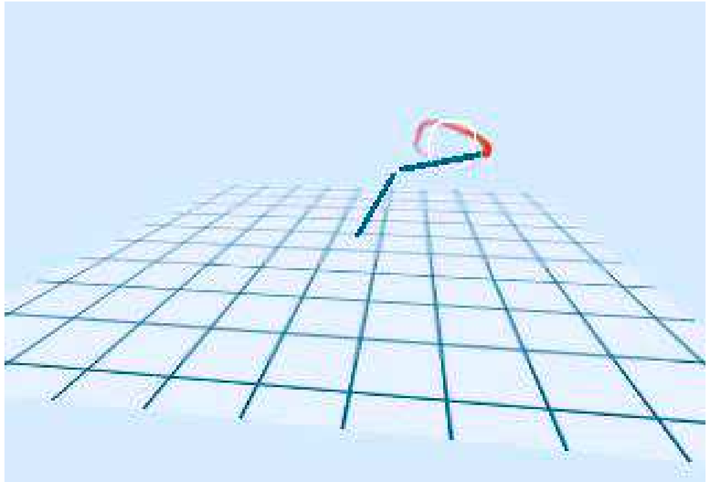
What: Figuring out how the brain works.

How: Building brains for robots = system-level modelling, implementing a whole vertebrate/mammalian brain.

Why: Because we can.

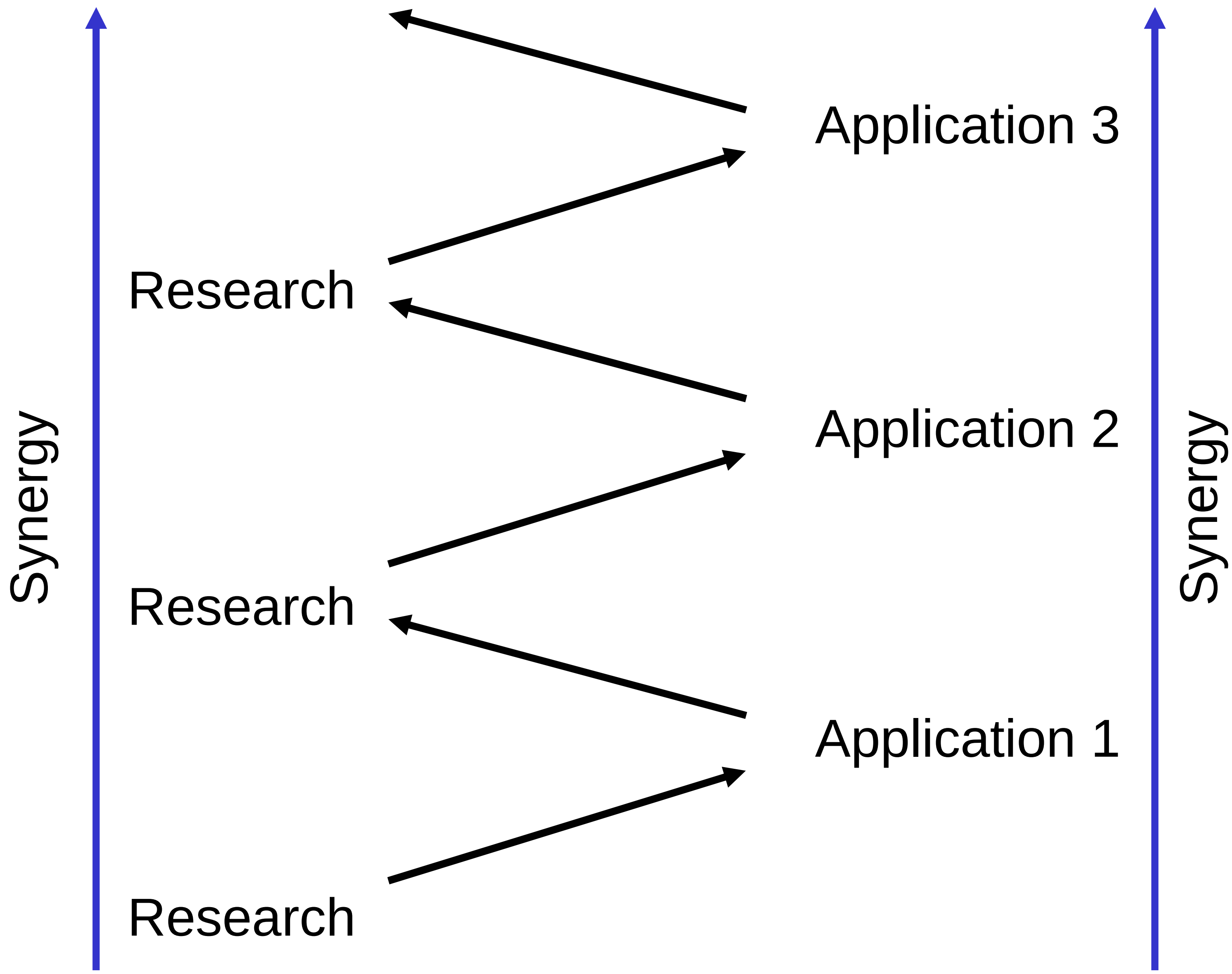
Intact cerebellum and sober





Cerebellar lesion or drunk

Singularity



Pilot Project: Robots for Waste Recycling

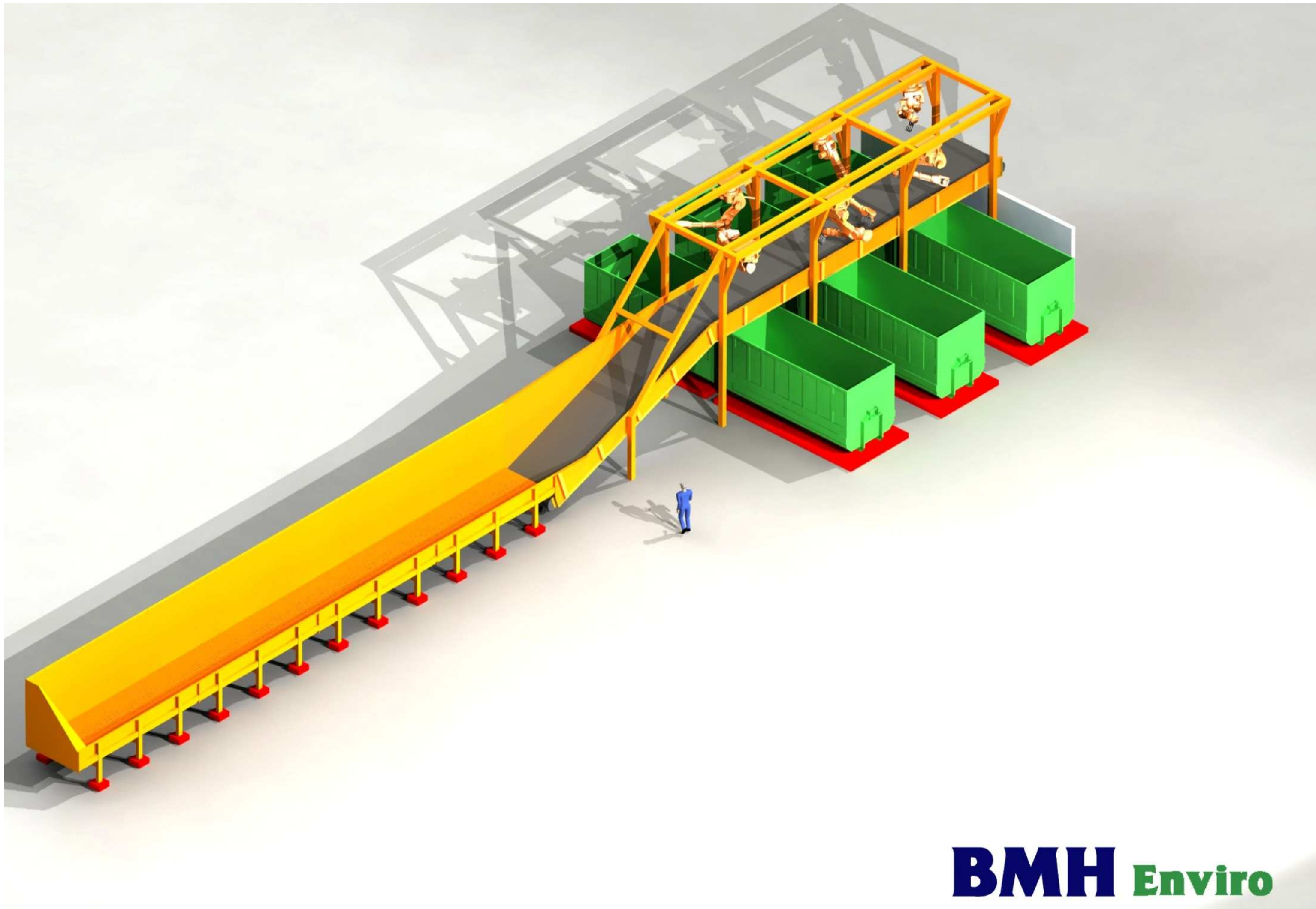
Harri Valpola

ZenRobotics Ltd.

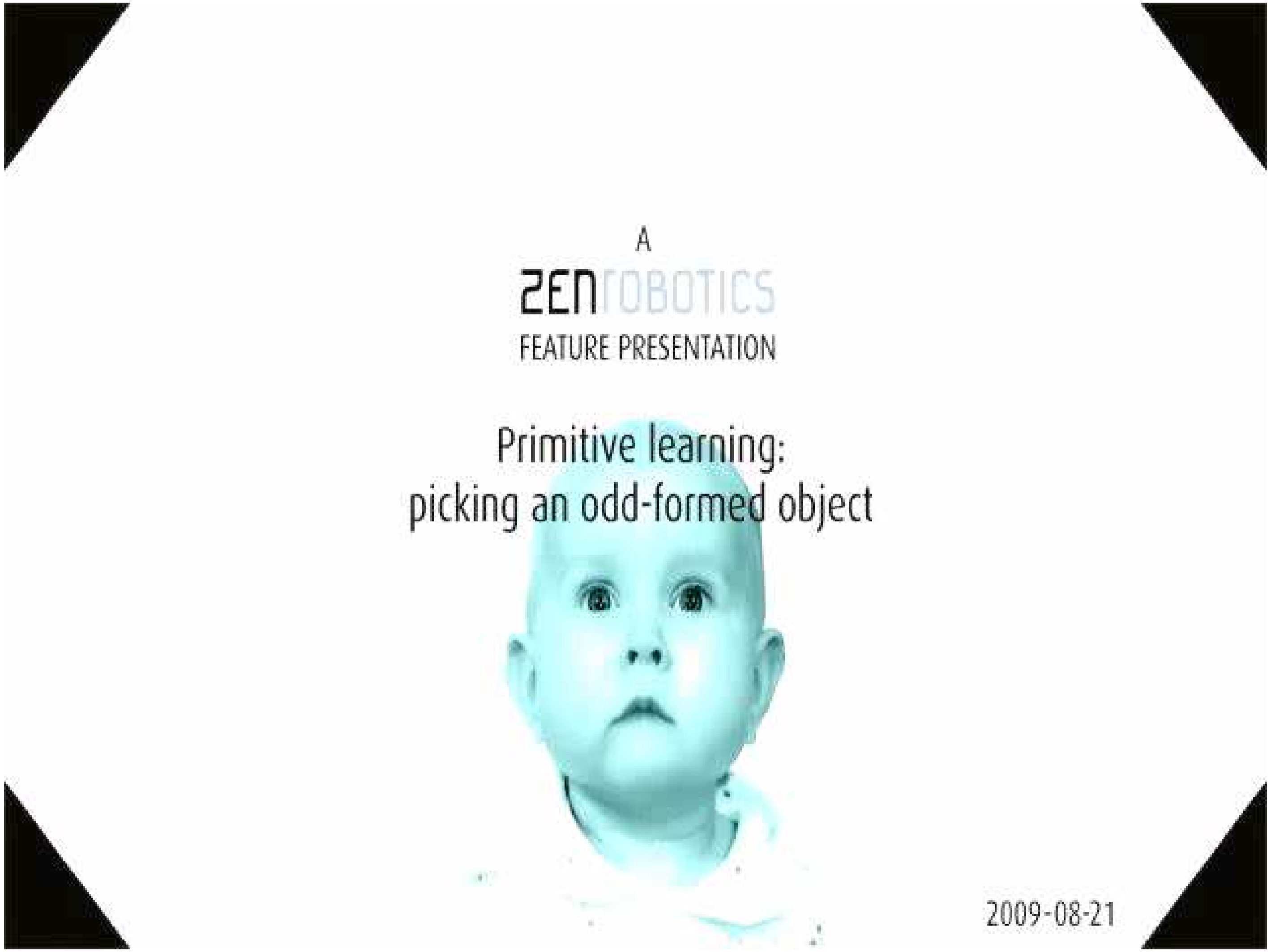
www.zenrobotics.com

Target application: hand-eye coordination for robots

Pilot plant
Scheduled to be in production in 2011



Jack the Gripper



Our main strengths in research

Neuroscience → machine learning / AI / neural nets

→ robotics in unstructured environments

Cognitive architecture: the organisation of the whole brain

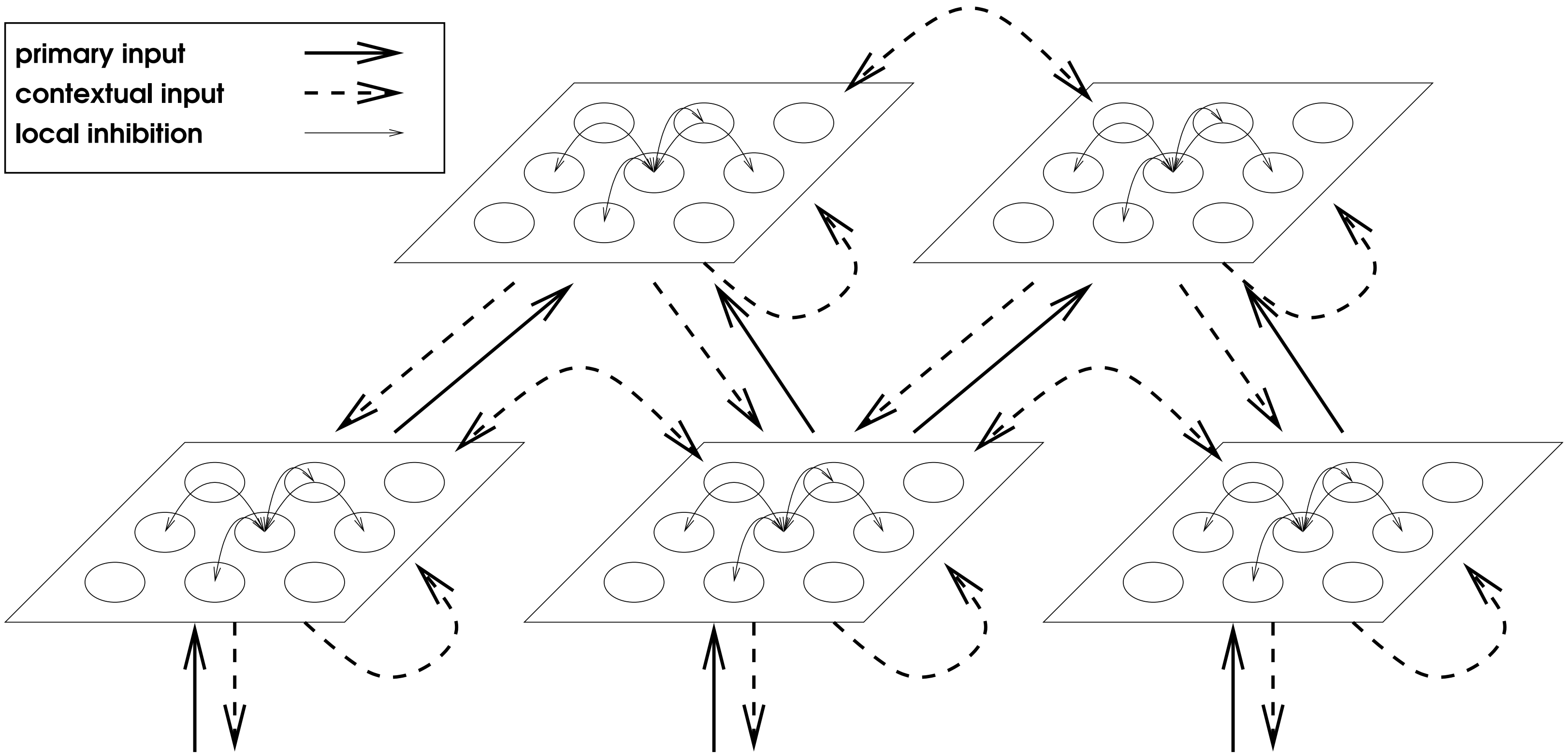
- **Cerebral cortex**
- Basal ganglia
- Cerebellum
- Hippocampus

Cerebral cortex

- Selection of useful information (attention and learning)
  - Unsupervised learning: from raw data to abstract concepts (sensory and motor)
  - Segmentation of objects
- Planning and simulation

Cerebral cortex: a network of interacting modules

- Small areas of cortex recognise their inputs
- The areas share this information and bias their decisions
- Selective attention emerges from the dynamics

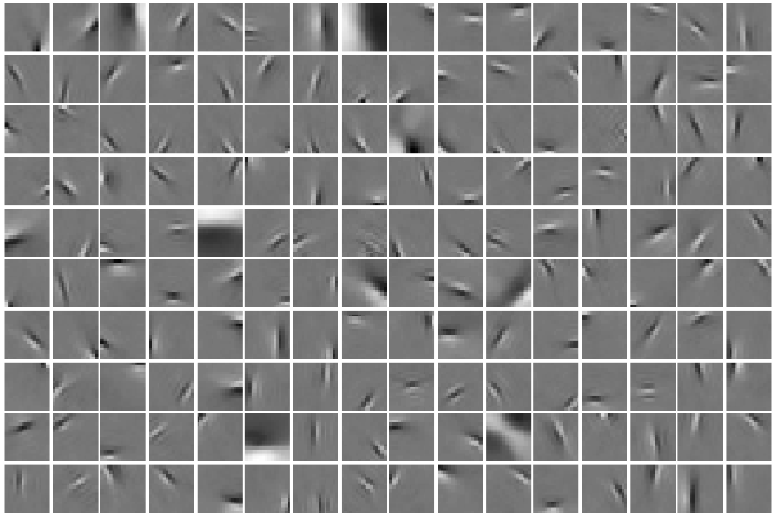


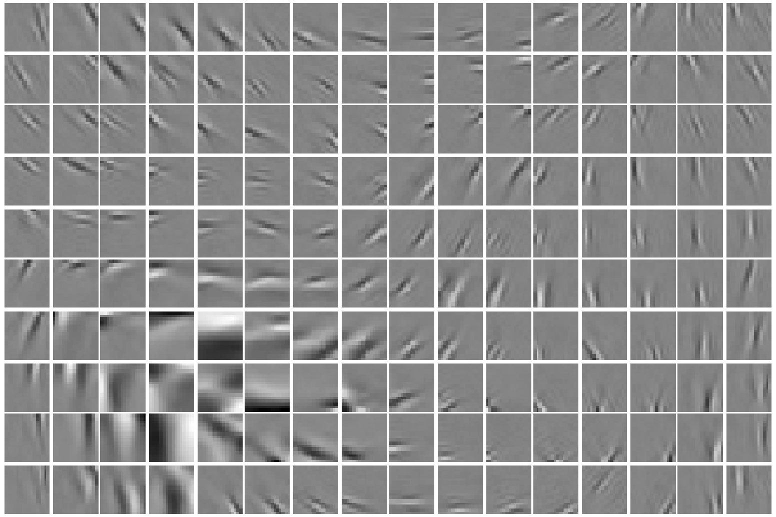
Now we have methods for:

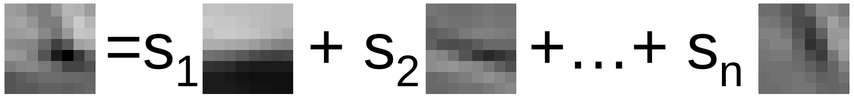
- Learning features (DSS, etc.)
- Learning correlation structures
- Integration, segmentation and selection of information
- Abstraction: low-level sensory and motor → more abstract sensory and motor

- [work in progress]: learn and perceive relations between objects
- [future work]: simulation and planning

Independent component analysis (ICA) for natural images





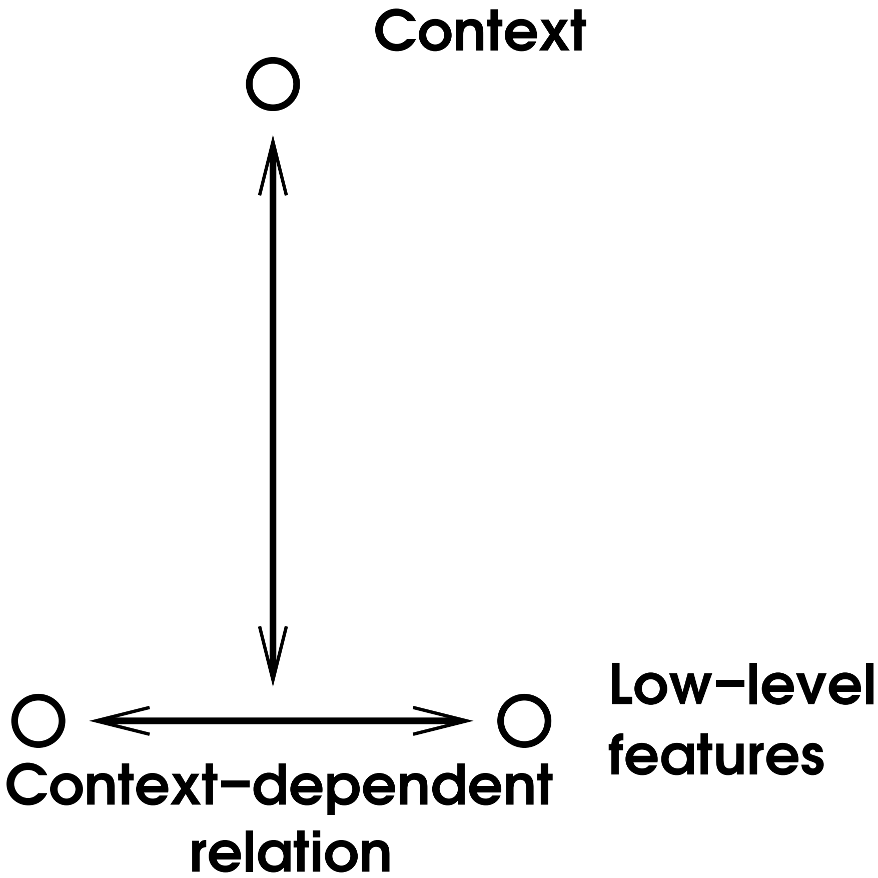


- ICA is an example of unsupervised learning.
- Can learn something like V1 simple cells.

http://www.cis.hut.fi/projects/ica/imageica/

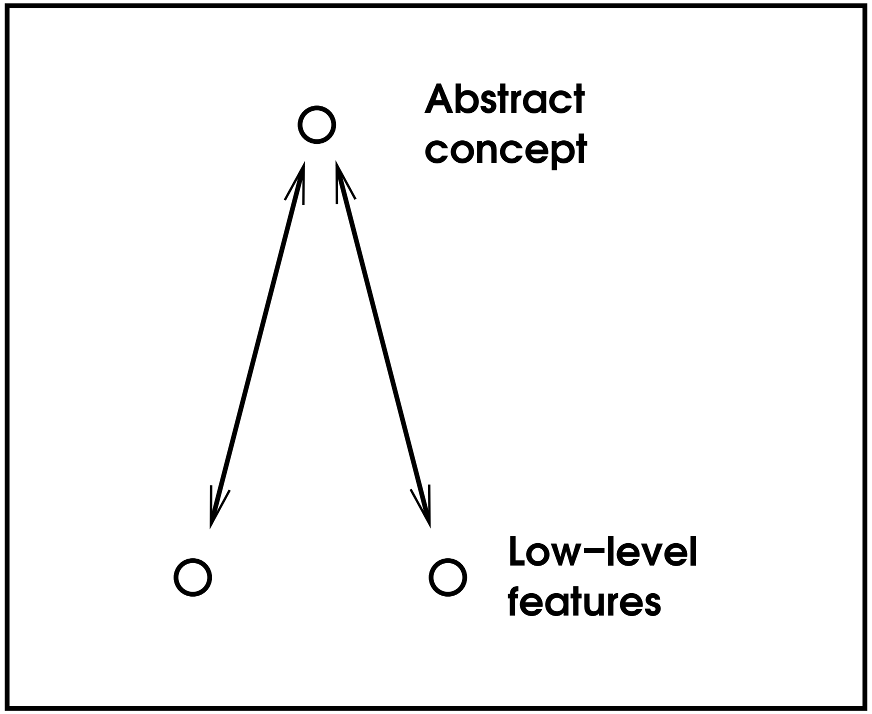
Model for correlation structure

- Recurring template:

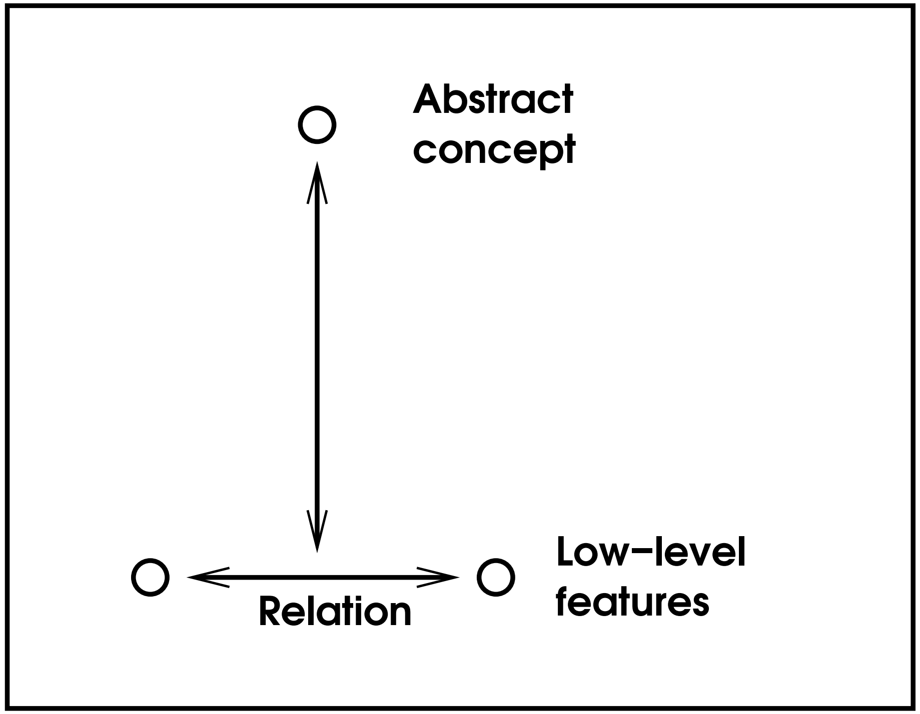


Model for correlation structure

- Key problem: how to learn this efficiently?



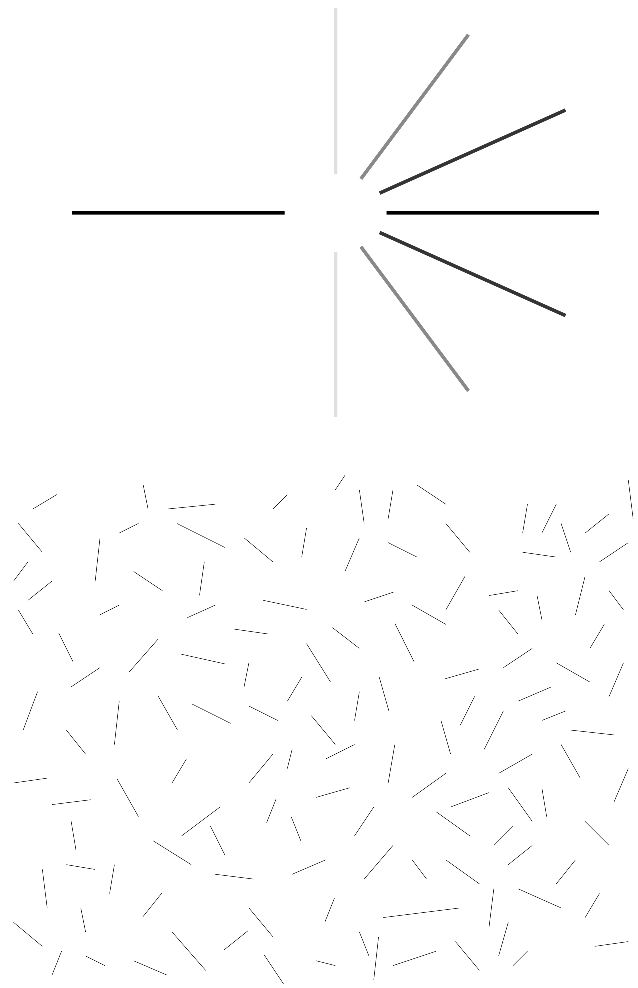
Regular neural nets



Model for correlation structure

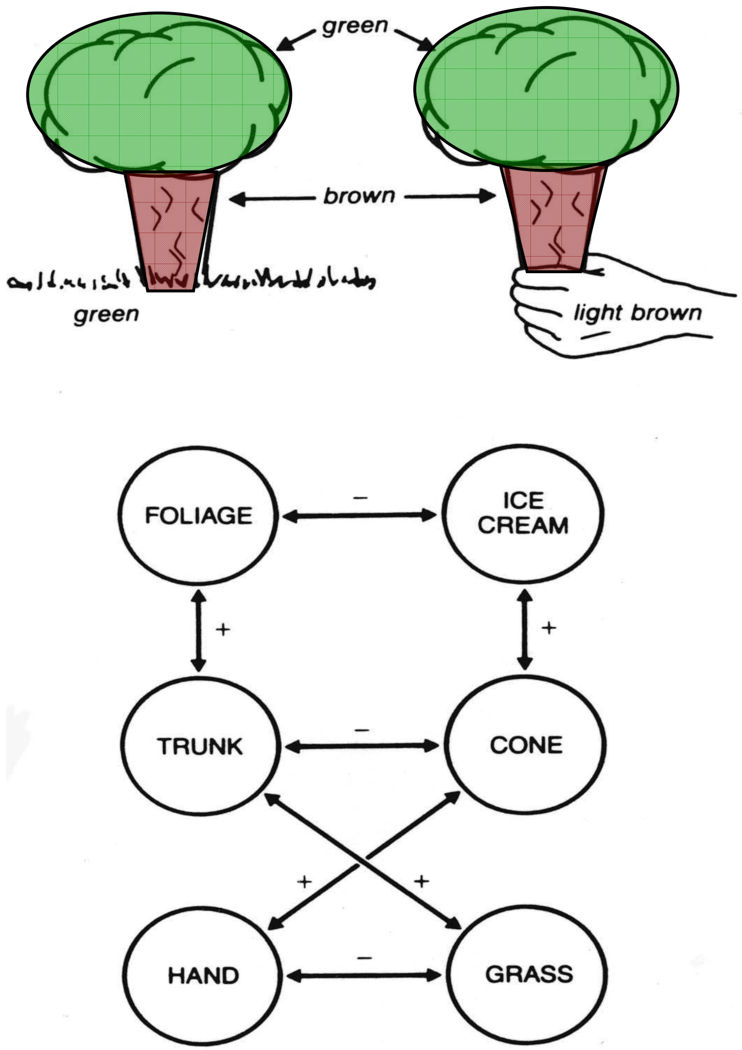
Segmentation: grouping features to objects

- Long-range associations are stronger between correlated features
- Anatomical basis for segmentation, Gestalt principles
- Dynamical model: features of the same object become synchronised



Attention vs. interpretation

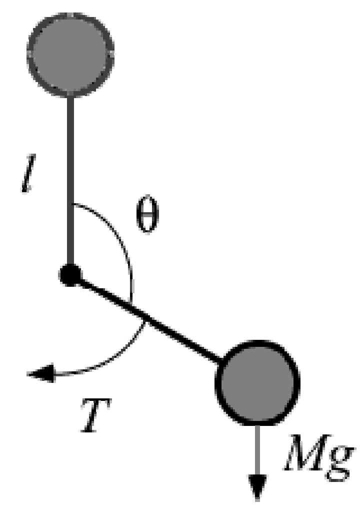
- The same mechanism can select information and choose between consistent explanations



http://ilab.usc.edu/classes/2002cs564/lecture_notes/06-Schemas.ppt

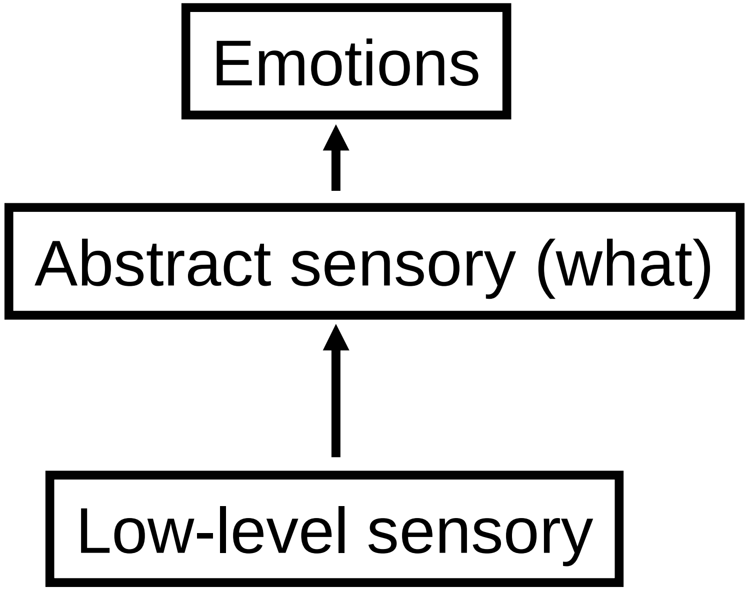
Example of abstraction: case pendulum

- Task: swinging up a pendulum with a weak motor → needs multiple back-and-forth swings (torque in phase with angular velocity)
- Low-level sensory input: sin θ, cos θ, dθ/dt
- Low-level motor output: torque force acting on the joint

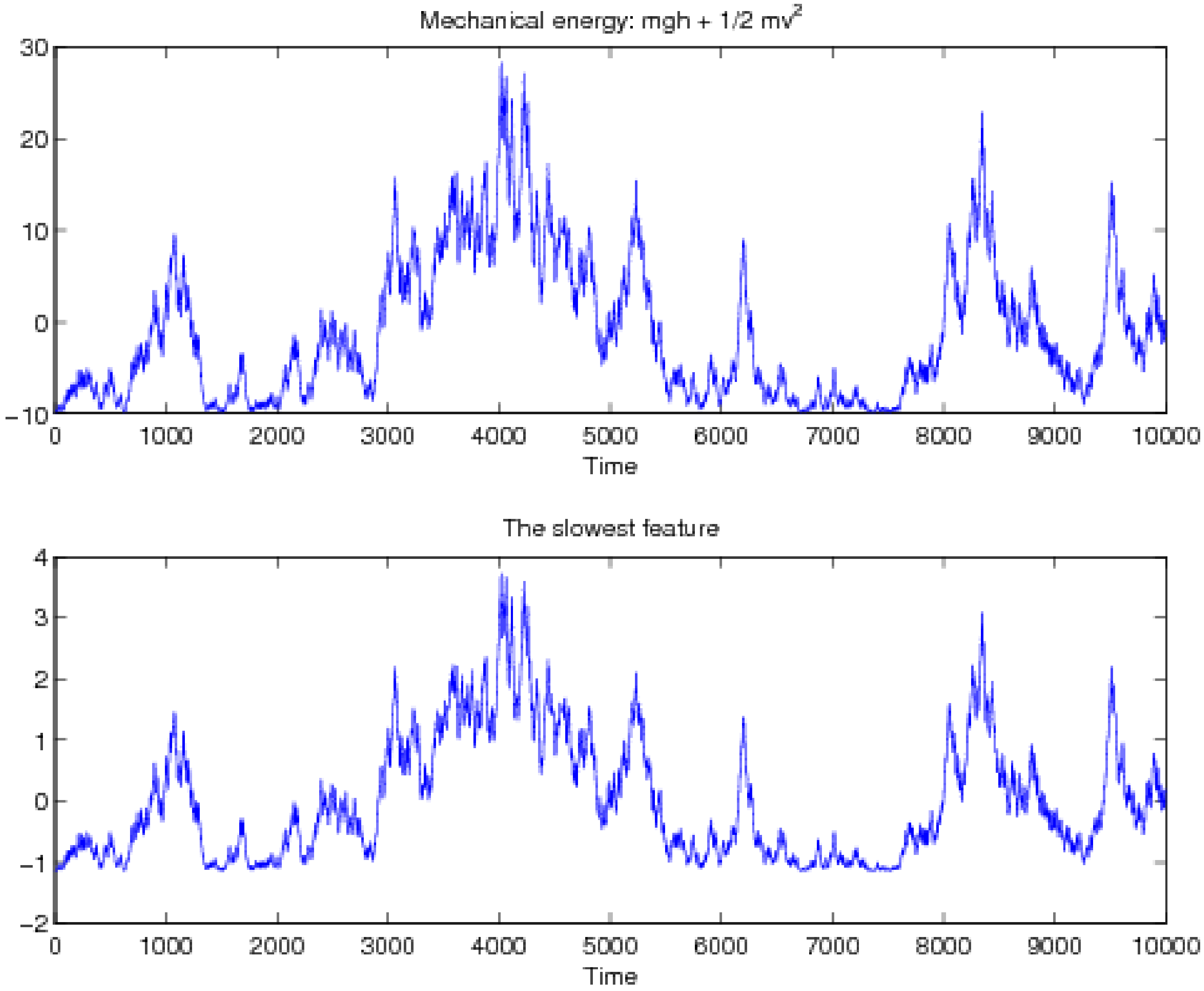


Learning sensory abstraction by slow feature analysis

- Exploration (generating data): keep flipping the torque direction randomly
- Find a feature which changes the slowest → mechanical energy
- This is the feature which best predicts the success in the task (emotions)

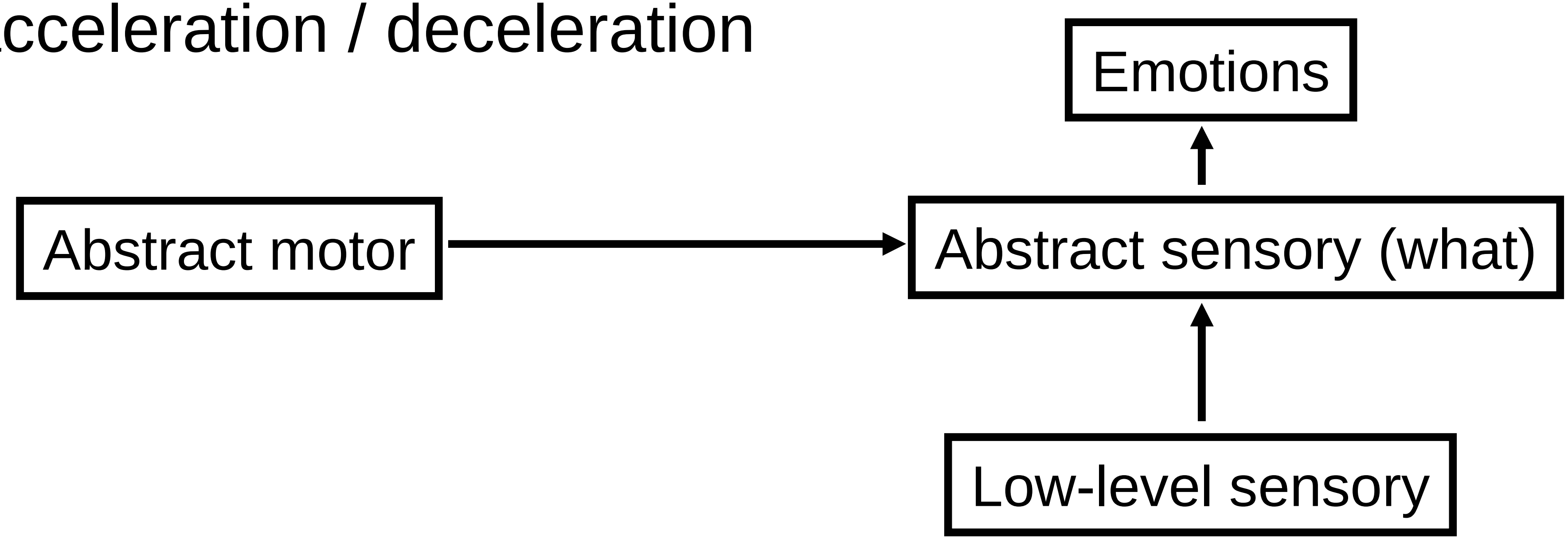


The resulting “what” feature



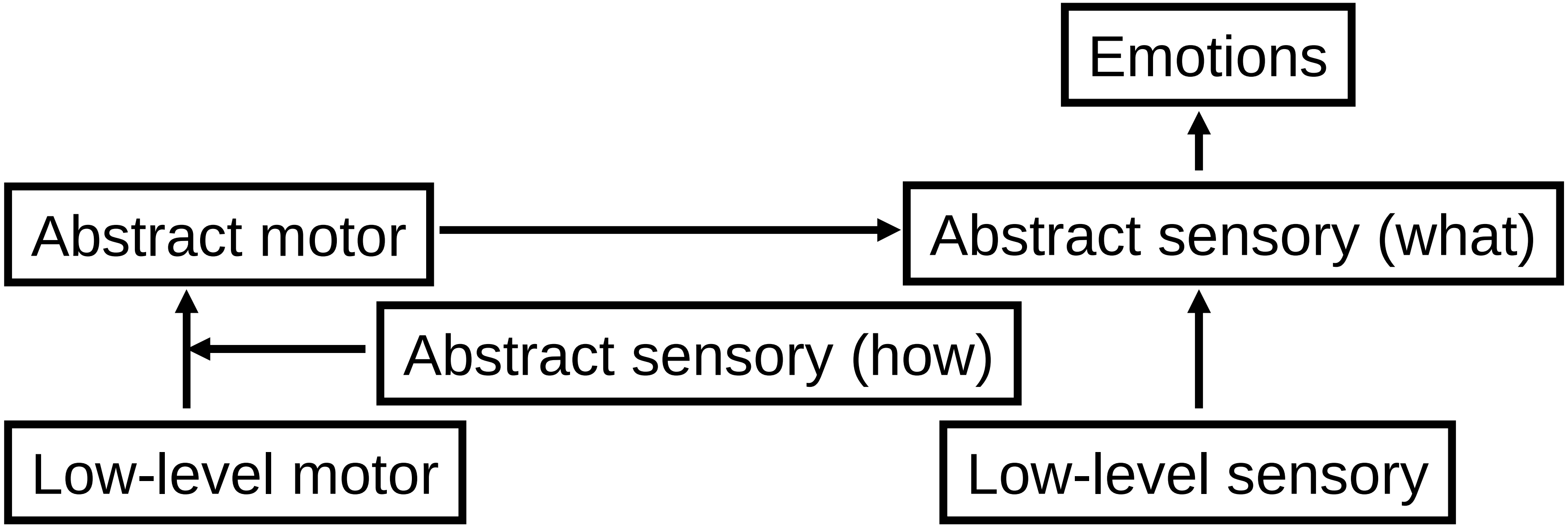
Abstract motor features

- Time derivative of the slow feature is a good candidate for a relevant action: maybe something important was done?
- Change in mechanical energy = acceleration / deceleration



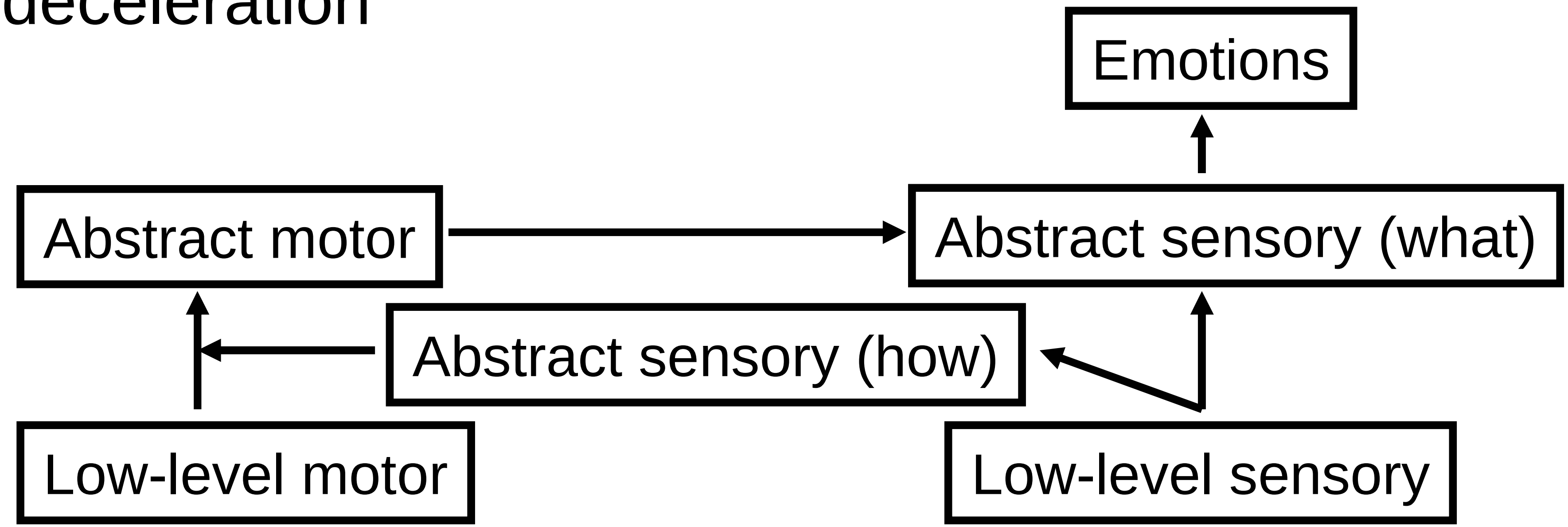
Correlation structure between low-level and abstract motor features

- Search for latent variables which represents the mapping between low-level and abstract motor features

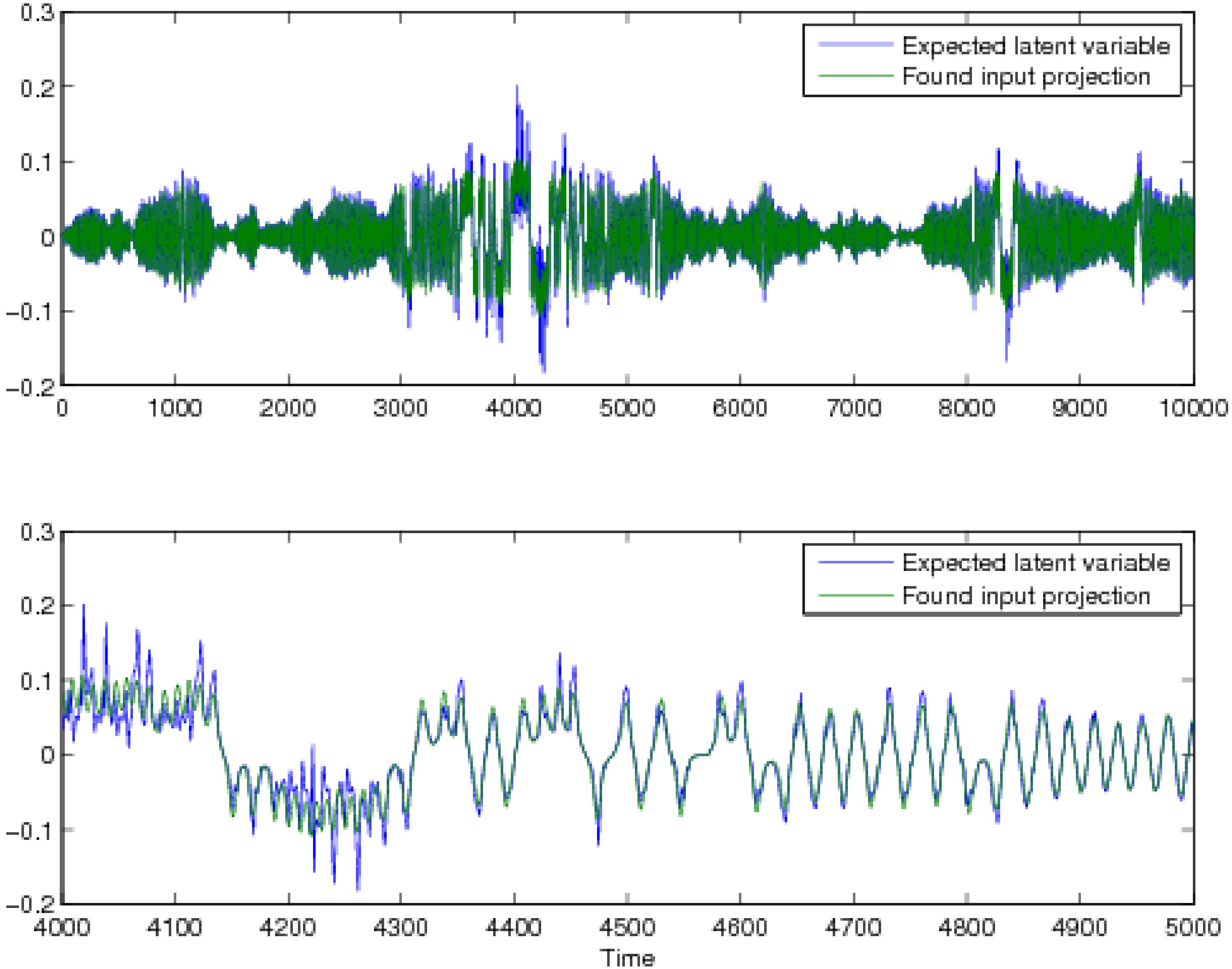


Learn to predict the “how” feature

- The relevant missing latent variable turns out to be essentially $d\theta/dt$
- With the model, control can be lifted from controlling the torque to controlling acceleration / deceleration



The resulting “how” feature



Thank you!

www.zenrobotics.com