Observations and Inspirations

Mutual Inspirations between Cognitive and Statistical Sciences

Shakir Mohamed

Research Scientist, DeepMind

shakir@deepmind.com

@shakir_za

Sheffield Machine Learning Research Retreat 2017

Abstract

Observations & Inspirations: The Mutual Inspirations between Cognitive and Statistical Sciences

Where do we obtain our inspiration in cognitive science? And in Machine Learning? These questions look at the parallels between these two fields. Fortunately, seeking out the parallels between minds and machines is one of our long-established scientific traditions, and this talk will explore the exchange of ideas between the two fields. The parallels between the cognitive and statistical sciences appear in all aspects of our practice, from how we conceptualise our problems, to the ways in which we test them, and the language we use in communication. One of these mutually useful tools are the conceptual frameworks used in the two fields. In cognitive science the most established frameworks are the classical cognitive architecture and Marr's levels of analysis, and similarly in machine learning, that of Box's loop and the model-inference-algorithm paradigm; these will be our starting point. The parallels between our fields appear in other more obvious forms, from cognitive revolutions and dogmas of information processing, to neural networks and embodied robotics. Recurring principles appear: prediction, sparsity, uncertainty, modularity, abduction, complementarity; and we'll explore several examples of these principles. From my own experience, we'll explore the probabilistic tools that connect to one-shot generalisation, grounded cognition, intrinsic motivation, and memory. Ultimately, these connections allow us to go from observation to inspiration: to make observations of cognitive and statistical phenomena, and, inspired by them, to strive towards a deeper understanding of the principles of intelligence and plausible reasoning in brains and machines.

What are the cognitive sciences?

- Neuroscience, physiology
- Psychology
- Sociology and behaviour

Minds

What are the statistical sciences?

- Probability and statistics, machine learning, AI.
- Information theory, signal processing, statistical physics
- Econometrics, game theory, operations research.

Machines

Intersectional Science

Advantages

Strengthens the *motivations* for our research

Refinement and precision in our thinking.

Evidence and *realisation* of learning systems.

Disadvantages

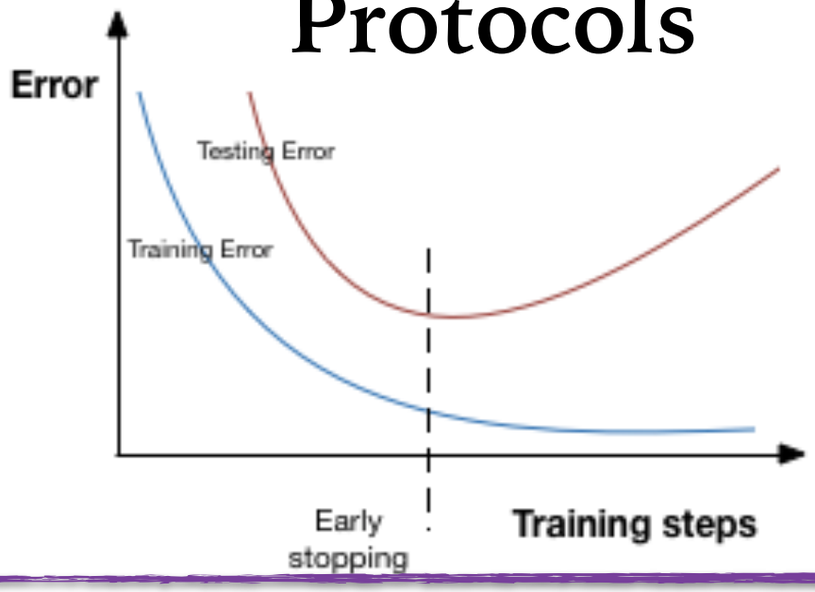
Superficial connections, hype, lack of focus.

Cross-pollination

Motivation and Language

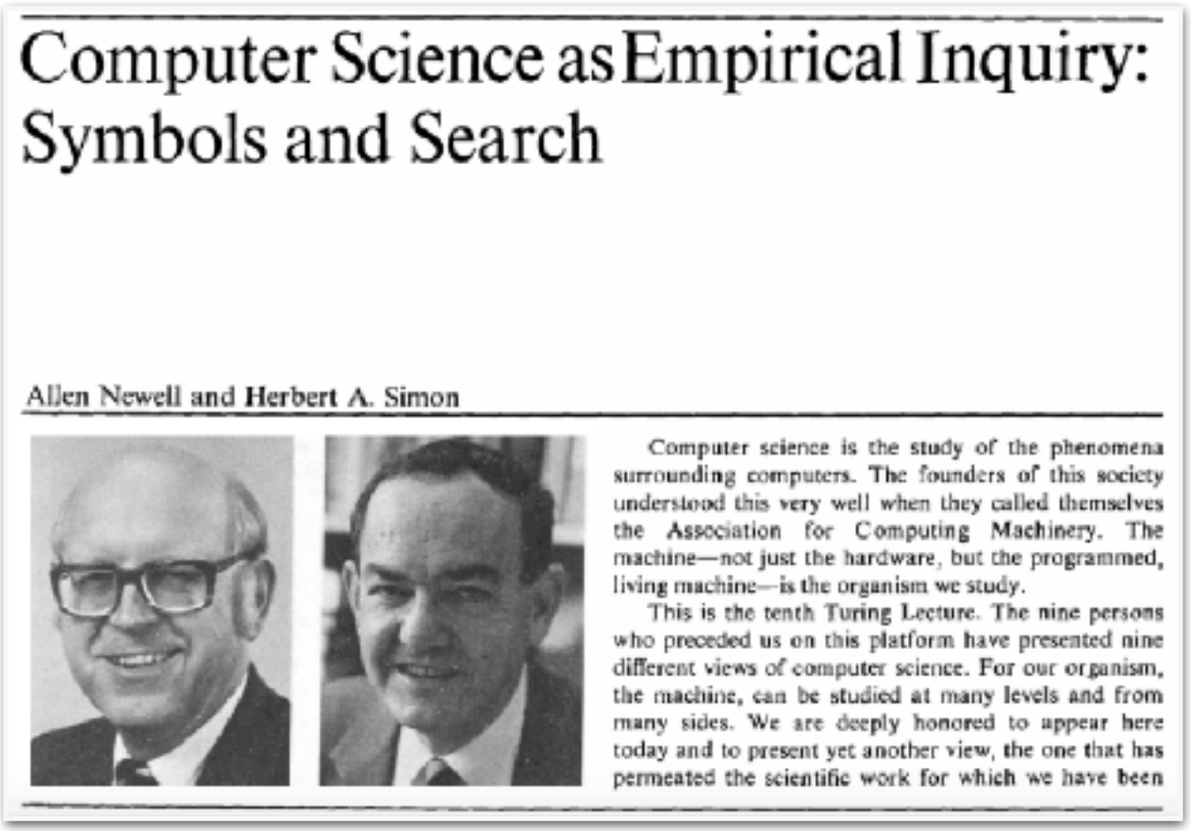
Conceptual and Scientific Frameworks

Testing Cases and Protocols



Classical Architecture

1975: Newell and Simon, Winners of the Turing Award

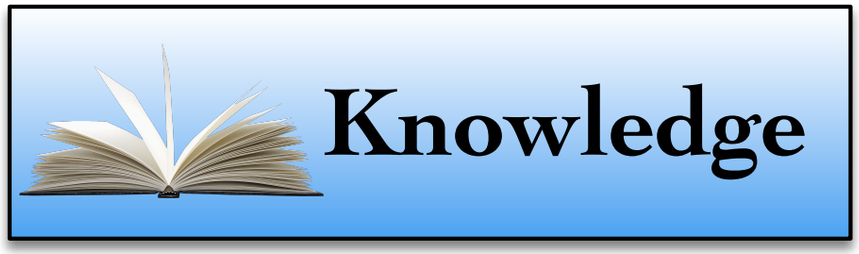


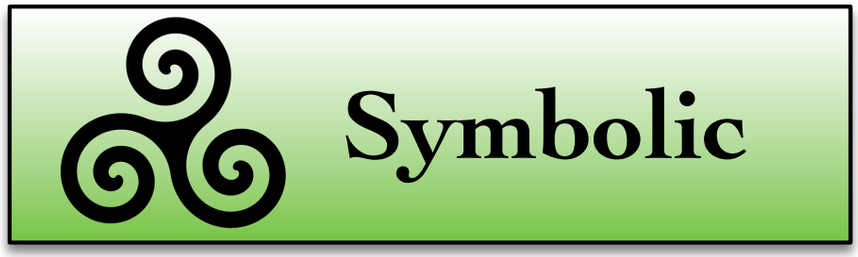
Computer Science as Empirical Inquiry: Symbols and Search

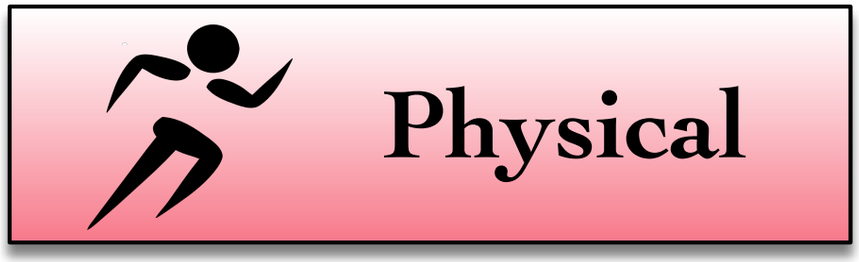
Allen Newell and Herbert A. Simon

Computer science is the study of the phenomena surrounding computers. The founders of this society understood this very well when they called themselves the Association for Computing Machinery. The machine—not just the hardware, but the programmed, living machine—is the organism we study.

This is the tenth Turing Lecture. The nine persons who preceded us on this platform have presented nine different views of computer science. For our organism, the machine, can be studied at many levels and from many sides. We are deeply honored to appear here today and to present yet another view, the one that has permeated the scientific work for which we have been



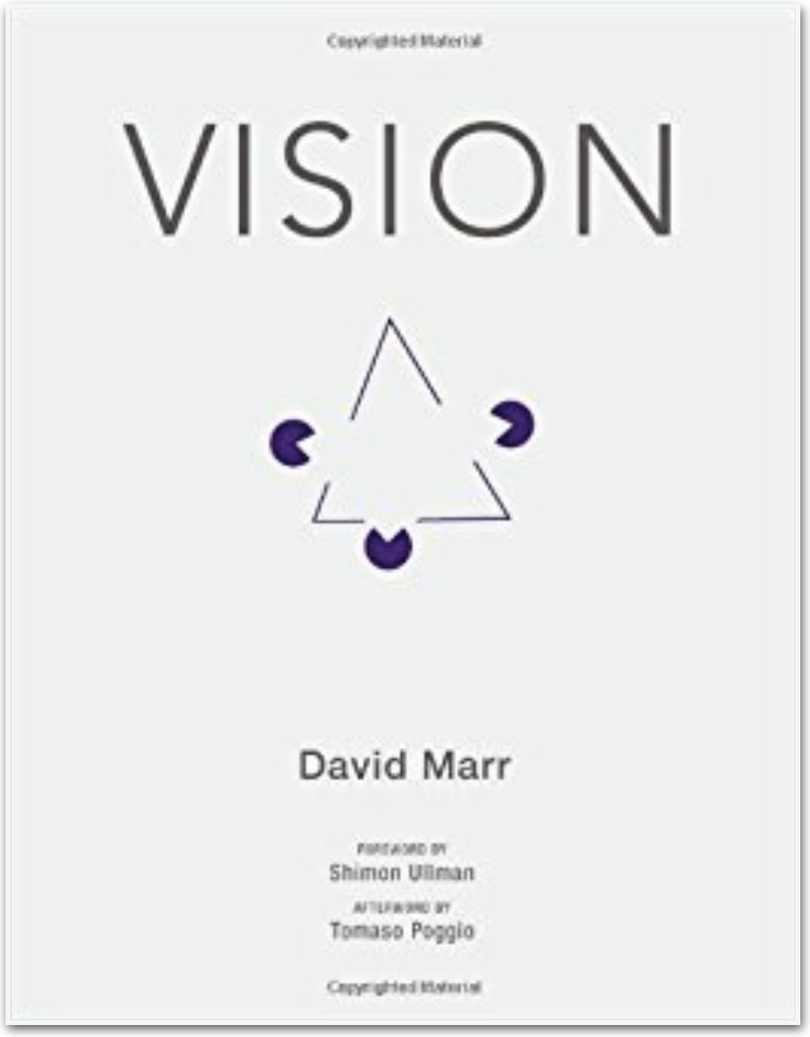


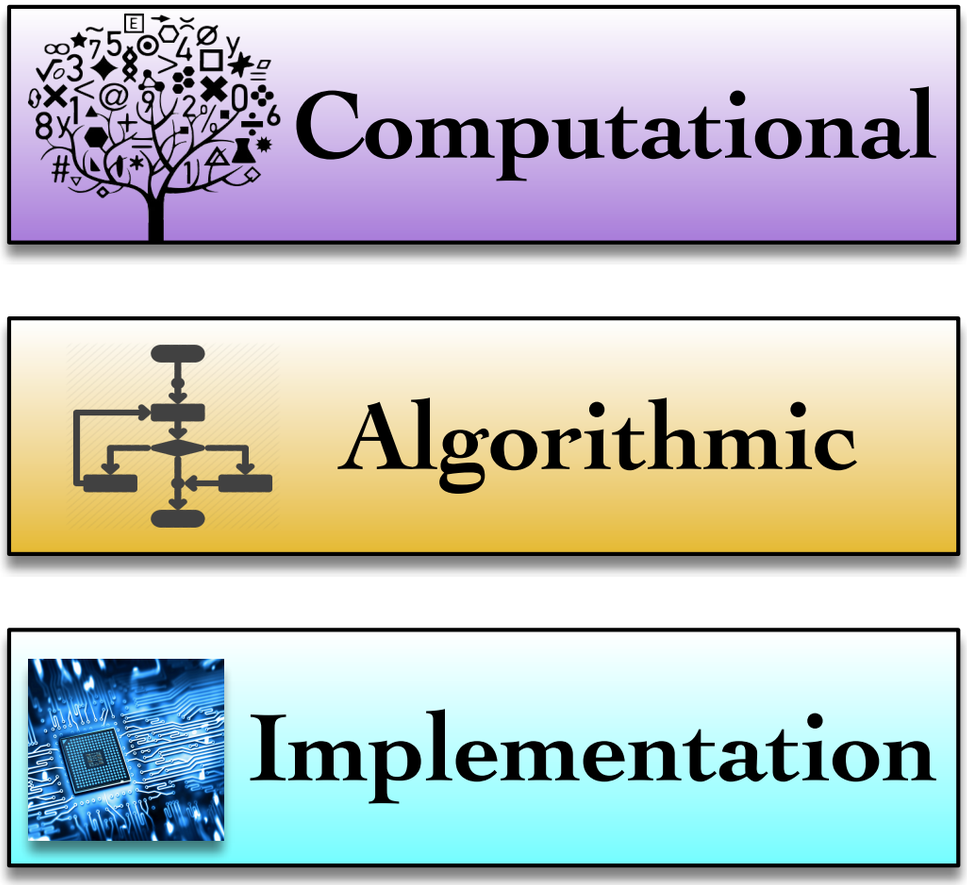


Classical Cognitive Architecture

Levels of Analysis

1982: David Marr's book, Vision.





Marr's Levels of Analysis

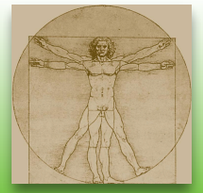
Phenomenological Levels

Sun et al's phenomenological levels.

Sociological

Psychological

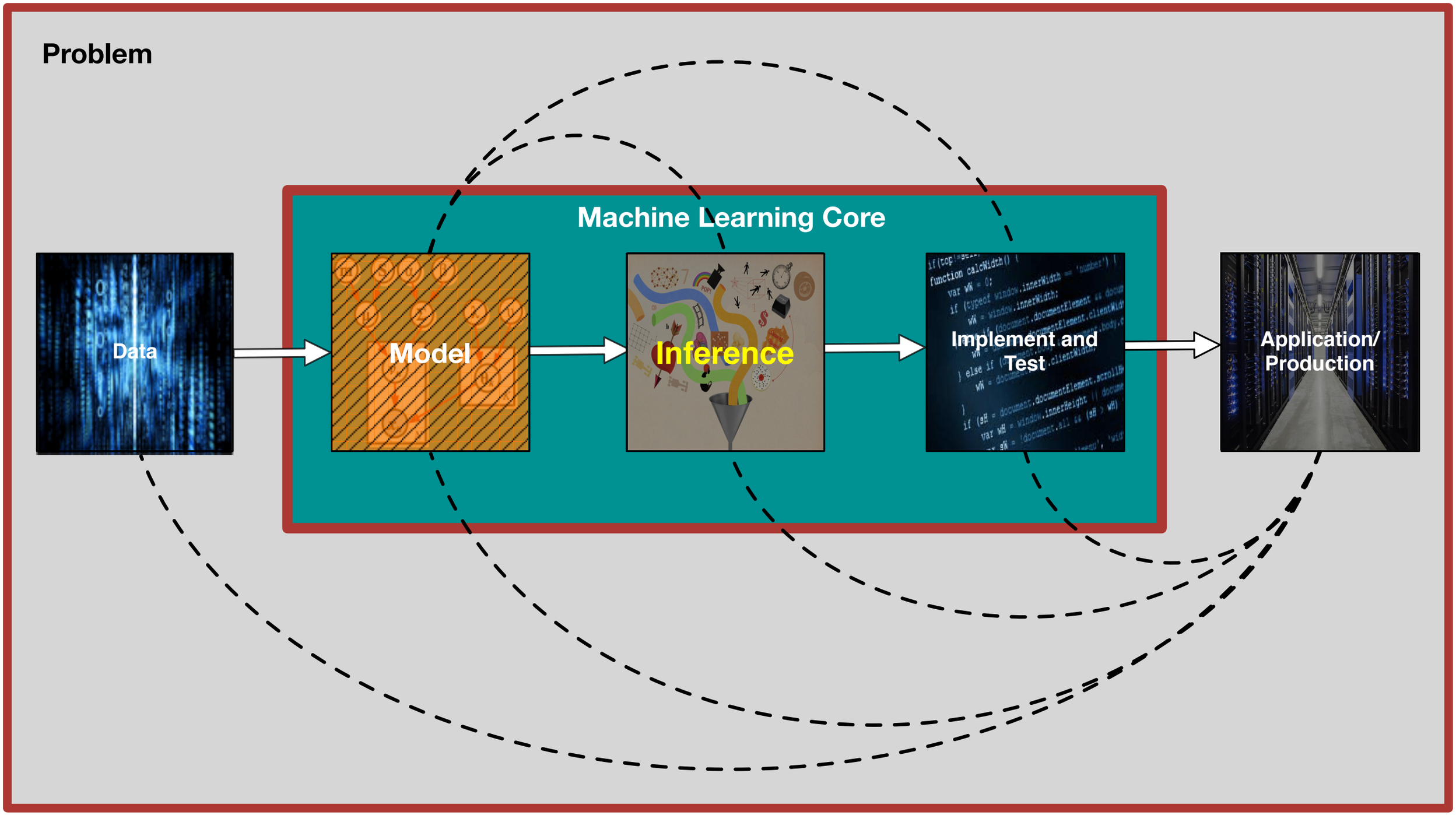
Componential



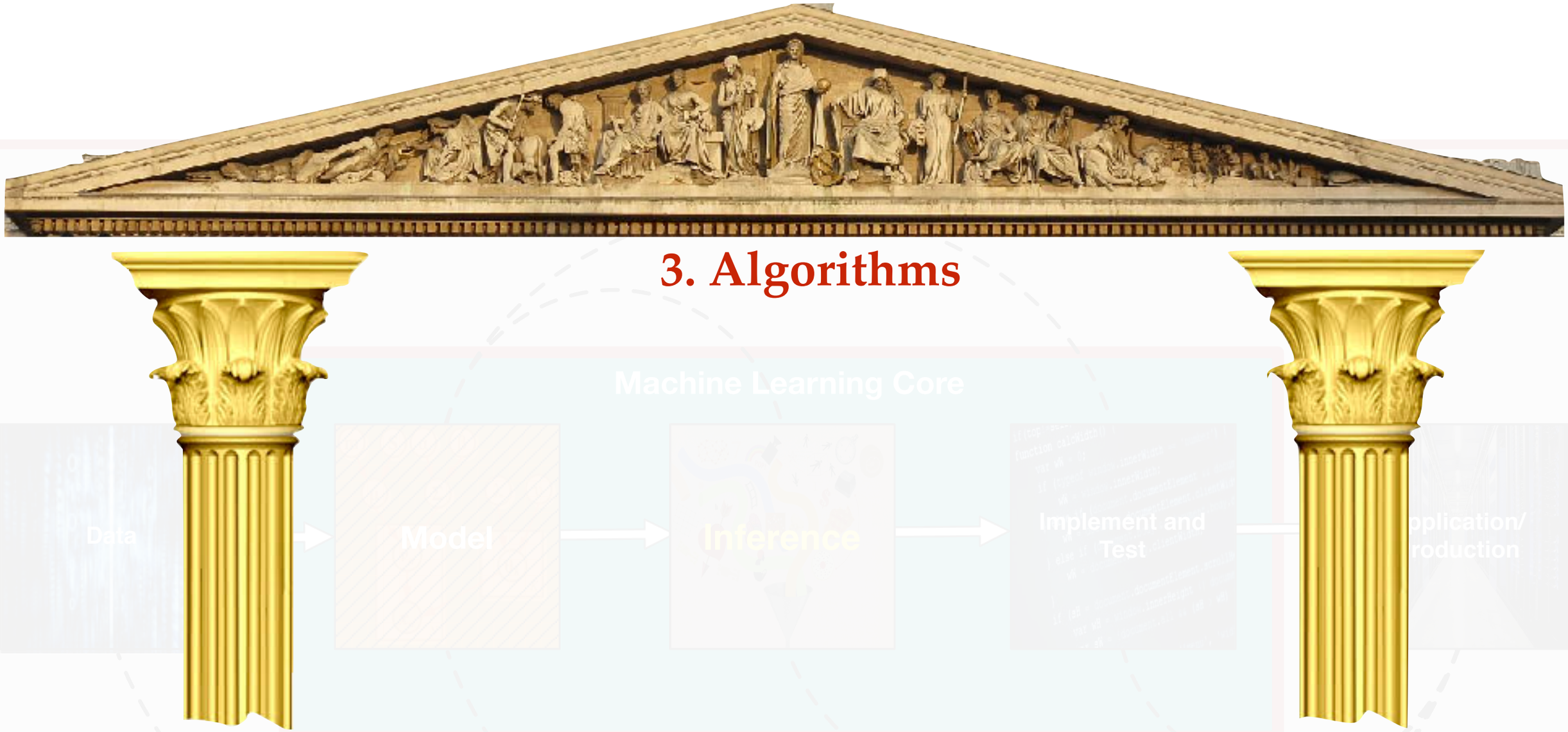
Physiological

Sun's Phenomenological Levels

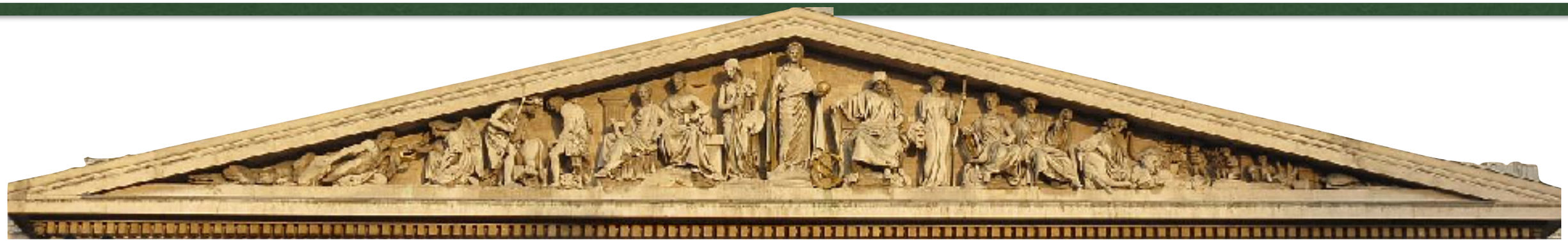
Modelling Lifecycle



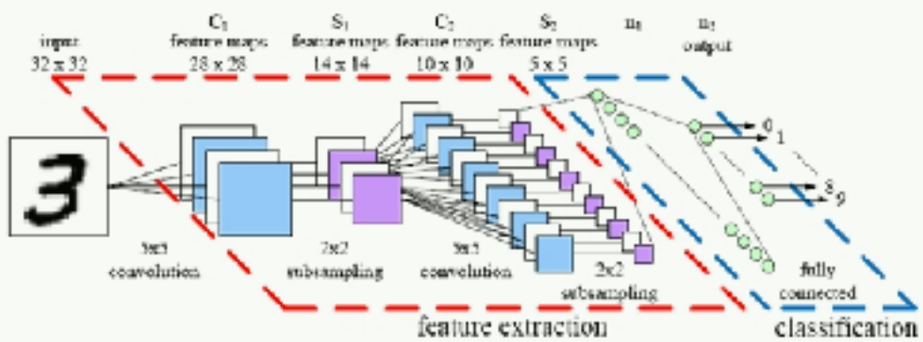
Model - Inference - Algorithm



Model - Inference - Algorithm

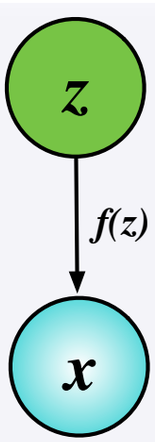


A given model and learning principle can be implemented in many ways.



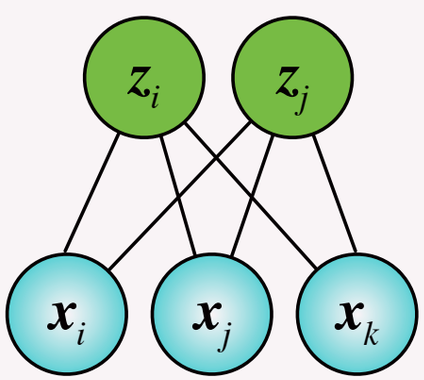
Convolutional neural network
+ penalised maximum likelihood

- Optimisation methods (SGD, Adagrad)
- Regularisation (L1, L2, batchnorm, dropout)



Latent variable model
+ variational inference

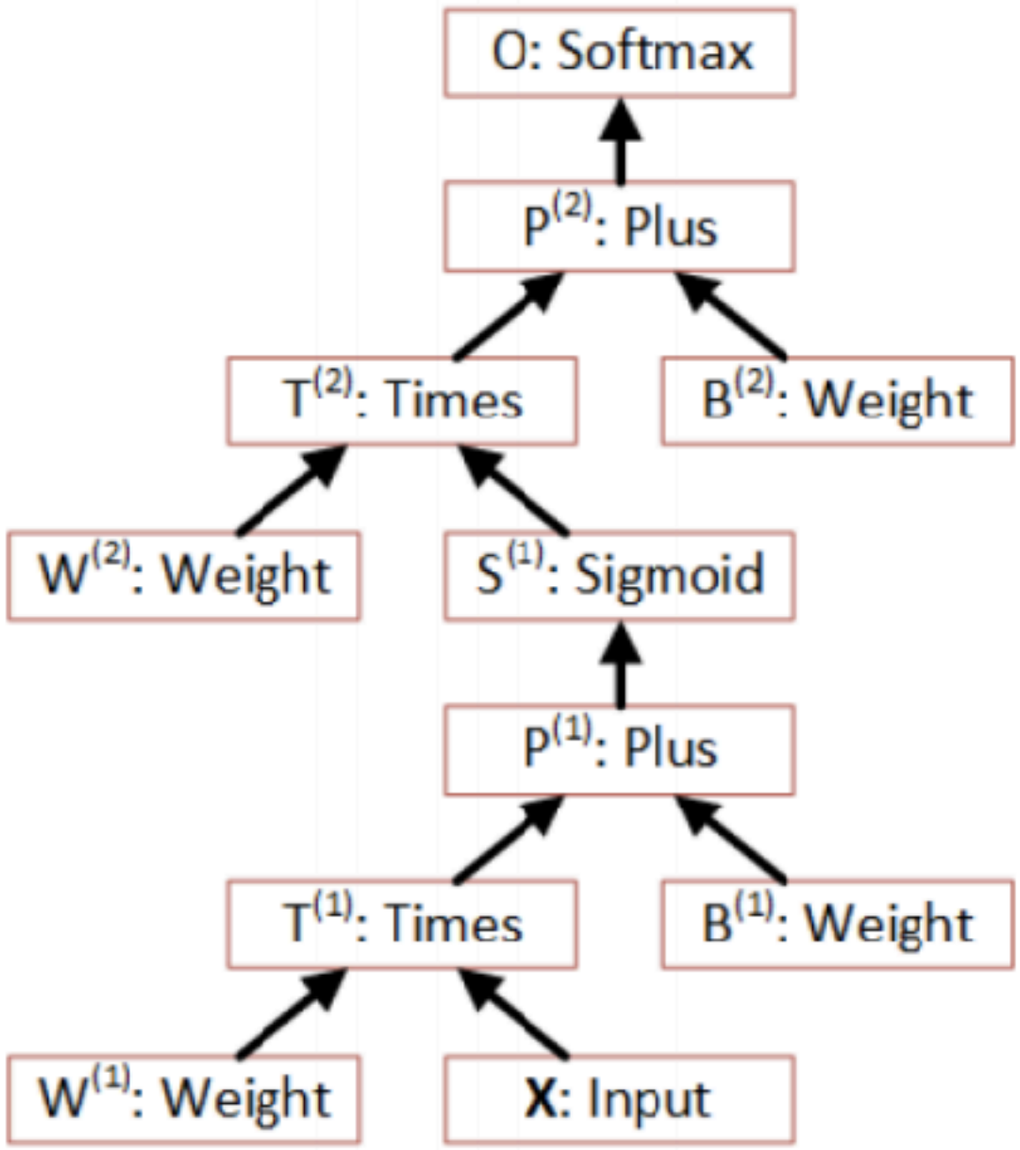
- VEM algorithm
- Expectation propagation
- Approximate message passing
- *Variational auto-encoders*

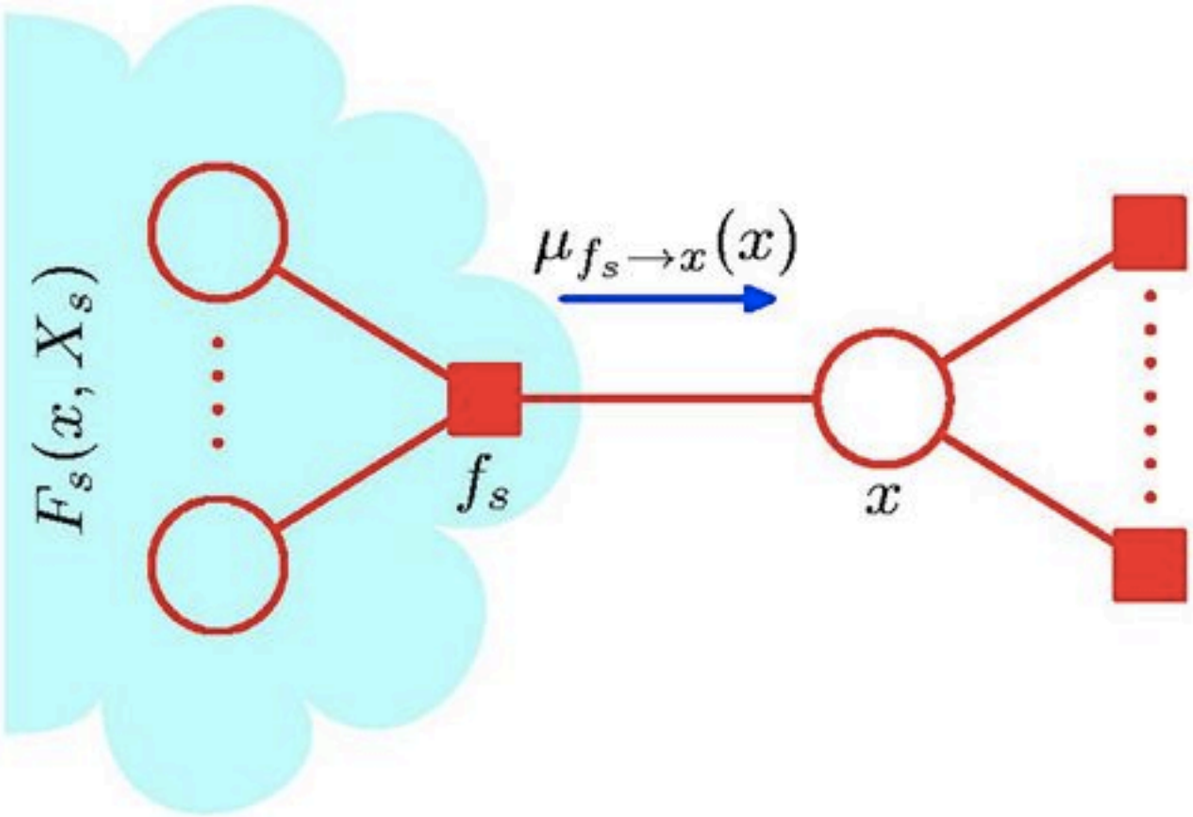


Restricted Boltzmann Machine
+ maximum likelihood

- Contrastive Divergence
- Persistent Contrastive Divergence
- Parallel Tempering
- Natural gradients

Architecture - Loss





1. Computational Graphs

2. Error propagation

Widespread Parallels

Information Theory and Statistical Learning

Cognitive revolution

Barlow's dogma of neural information processing

Normative models of cognition

Machine Learning

Analogical reasoning

Neural networks

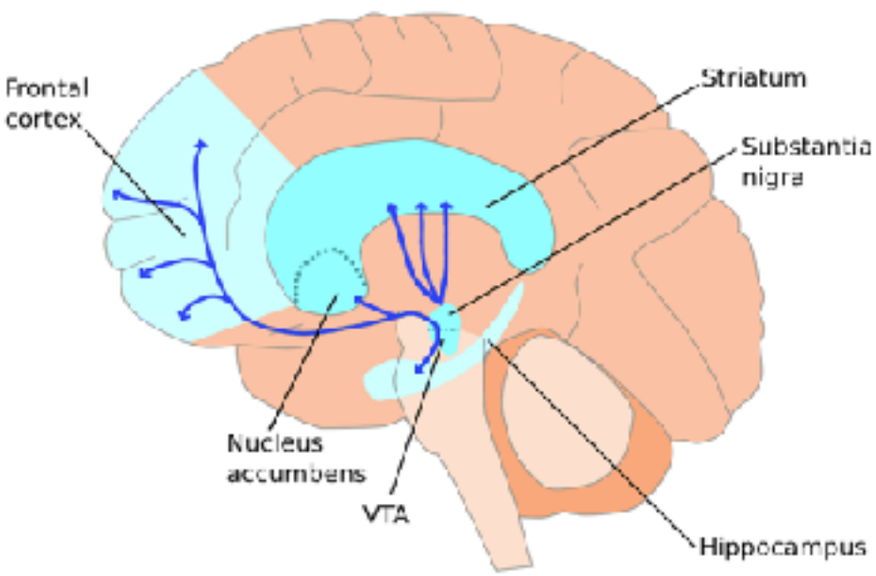
Embodied cognition

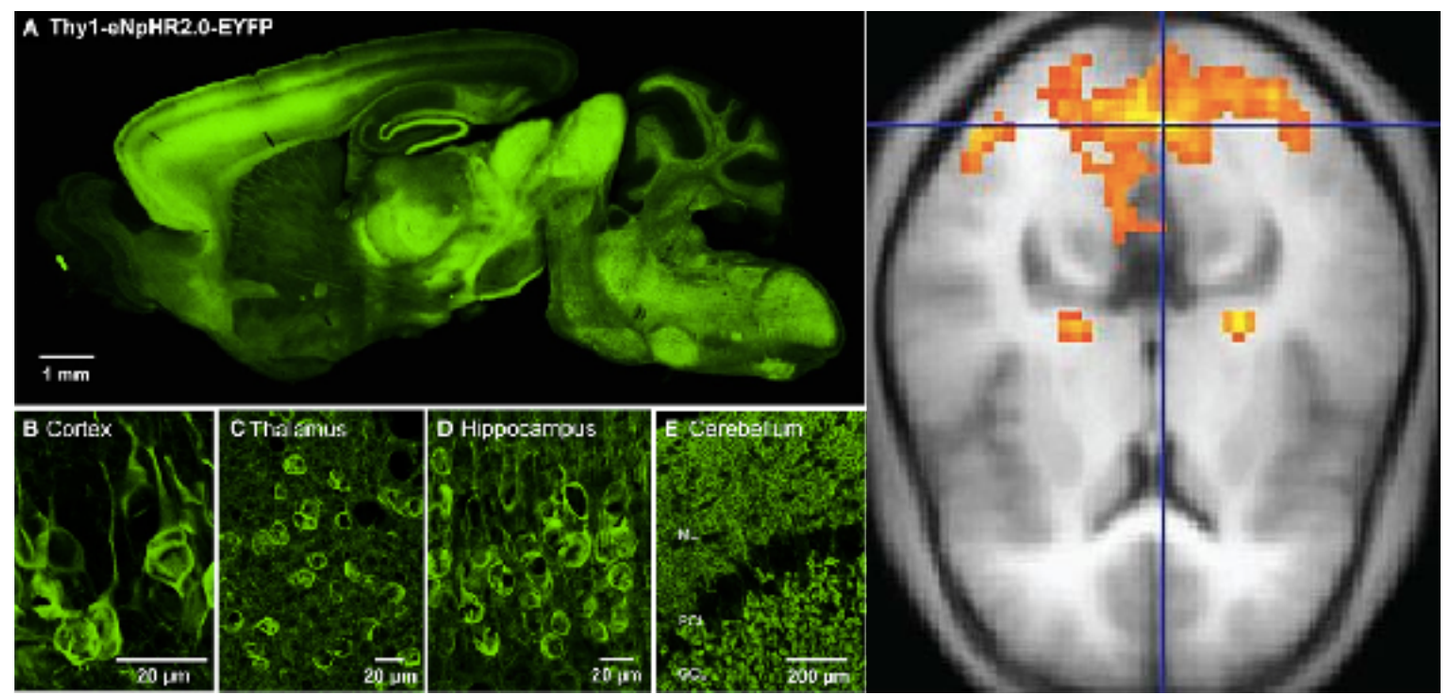
Episodic memory

One-shot generalisation

Shared Principle: Prediction

- Classical and instrumental conditioning tasks -role of striatum.
- FMRI and single-cell recordings of dopaminergic neurons.
- Optogenetic activation to show casual link between prediction error, dopamine and learning.

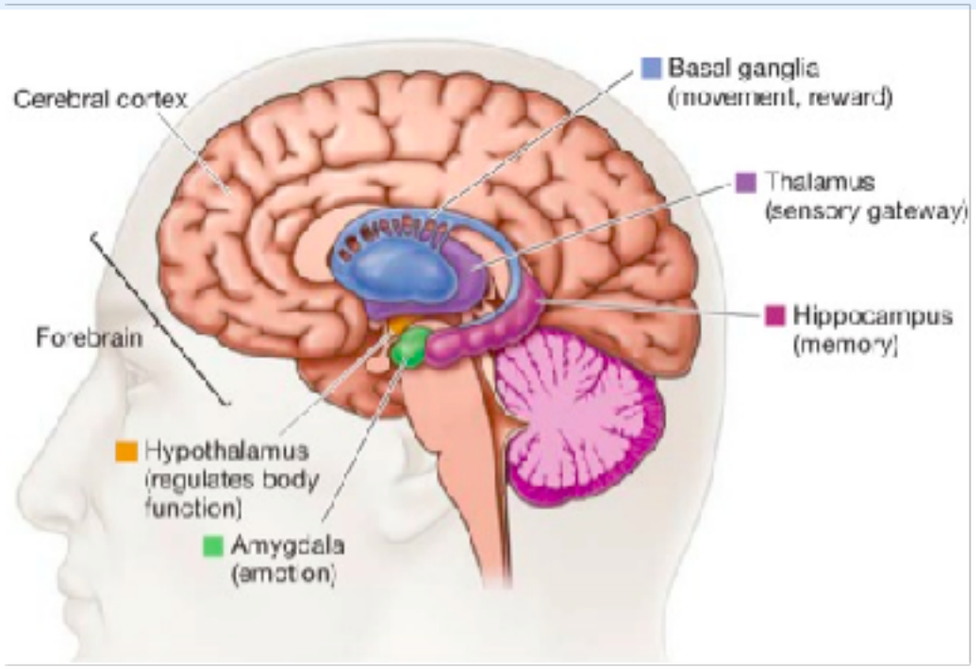


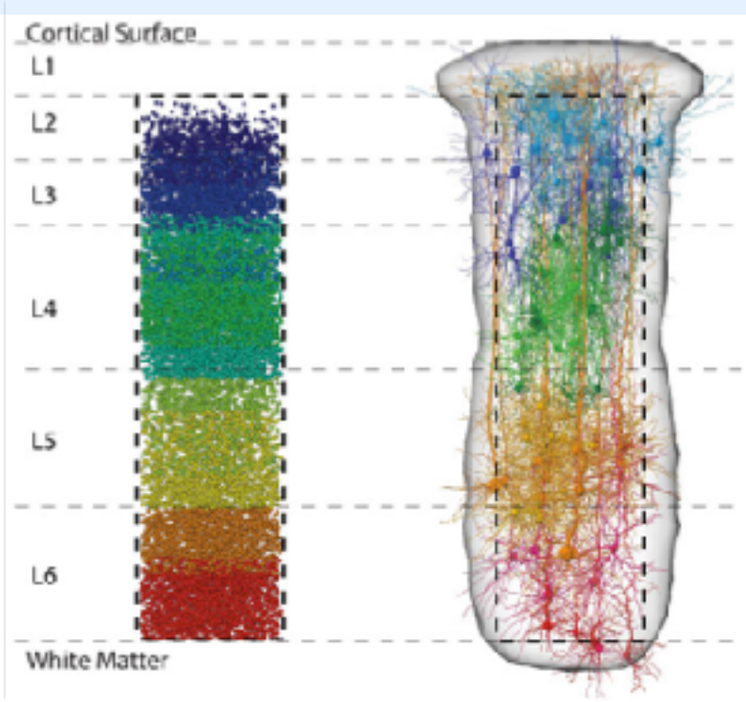


- Prediction of summary statistics: value functions.
- All machine learning based on prediction error.

Shared Principle: Sparsity

- Functional unit of the brain: sparse activation in L2/3.
- Overcompleteness in connections of Thalamic neurons to L4.
- Primates, rats, insects, rabbits, birds.





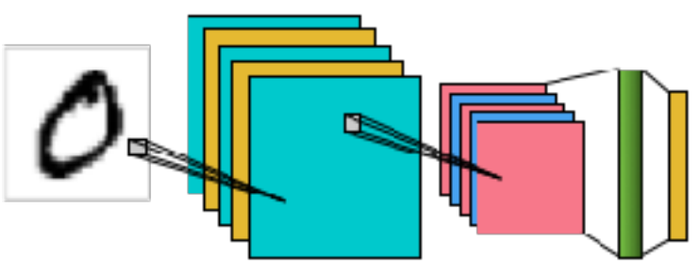
- Sparse representations as a general principle of regularisation and robustness.
- Penalised likelihood methods, simplicity of explanations.
- Optimal recovery guarantees.

Shared Principle: Complementary Systems

- Lesioned and epileptic patients: HM and KC, highlight that hippocampus for episodic memory and abstract representations.
- Early learning relies on episodic memory and hippocampus, then shifts to dopaminergic neurons in striatum.
- Complementary learning systems.

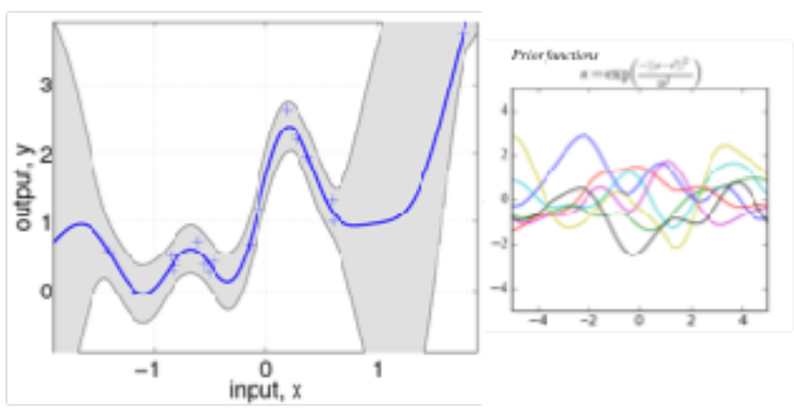
Parametric

Use parameters to represent functions and adjust with data.



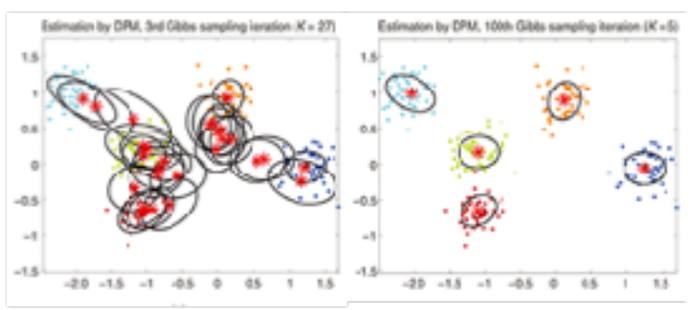
Non-parametric

Use infinite-dimensional parameter set. Store all data and use distances.



Semi-parametric

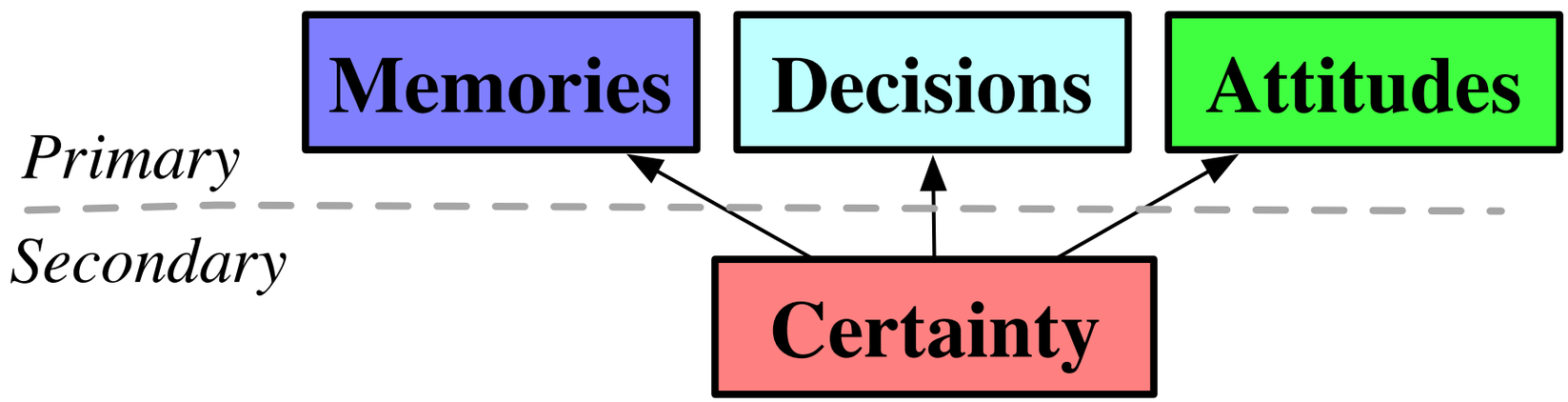
Combine parametric and non-parametric for complementary learning.



- Rapid, non-parametric systems, and slower parametric systems.
- Semi-parametric learning, with many possible variations.

Shared Principle: Uncertainty

- Young children can report confidence in their decisions and understanding.
- Recordings in rats, monkeys, choice-tasks in humans.
- People have the ability to represent and use confidence in memories, decisions attitudes.



- Wiener's cybernetics used the word chaos for uncertainty.
- Coverage and calibration, Bayesian analysis, uncertainty shapes learning, risk, value-at-risk and sensitivity.
- Impact on control, exploration and optimistic principles.

Shared Principles

- ***Modularity*** - motor system and action synergies and it's relation to hierarchical control.
- ***Explanation*** - causal mechanisms and categorisation. Causality and relational learning in machine learning.

Examples from our own work

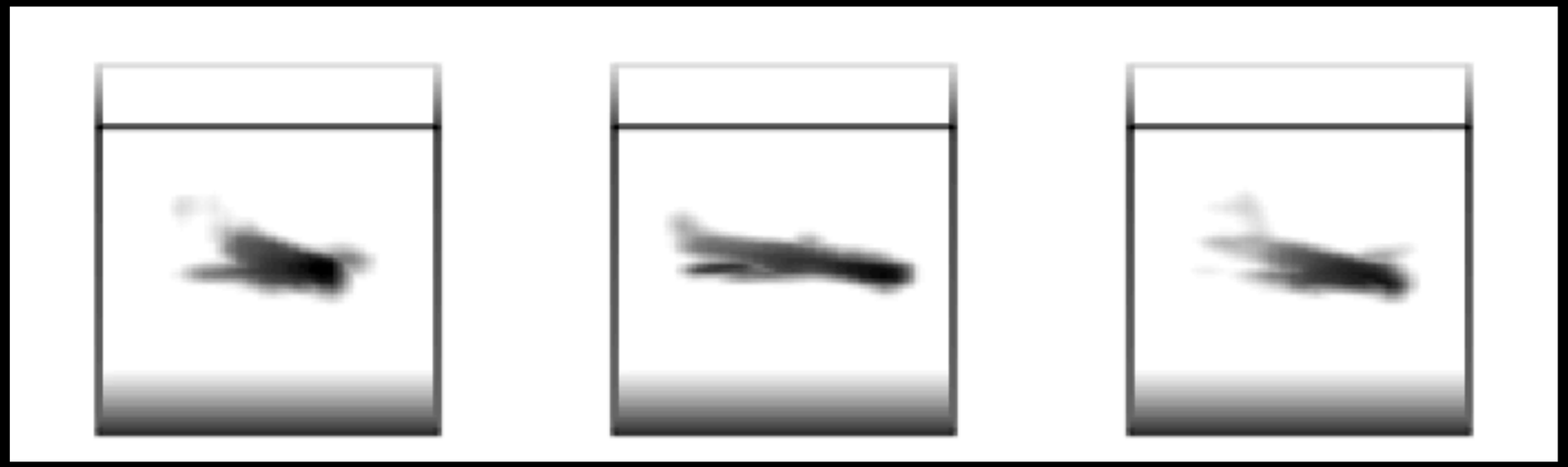
1. Perception and generalisation
2. Grounded cognition and future thinking
3. Reward and intrinsic motivation
4. Memory and coherence

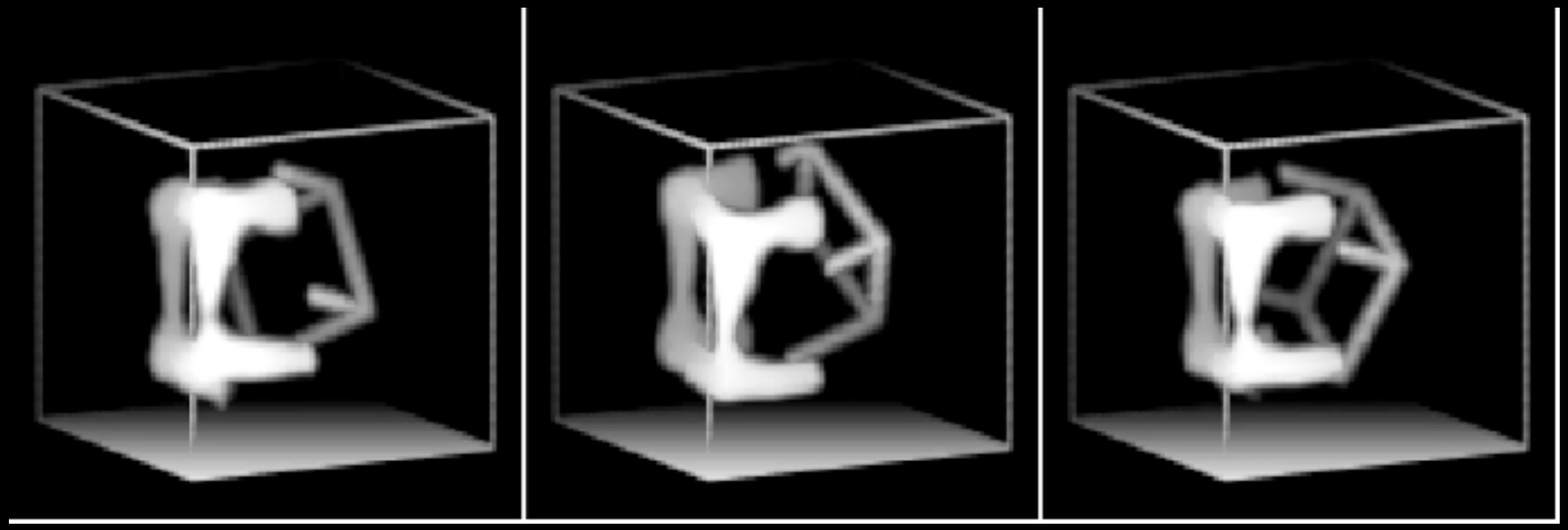
Perception and Generalisation

Cognitive Observation: Humans are able to generalise in remarkable ways: from scenes, with incomplete information, across diverse behaviours, and from limited amounts of data.

Cognitive Inspiration: Mental representations are formed that encode conceptual information, and capture generality and stochasticity of sources of information and allow for rapid transfer.

Scene Understanding

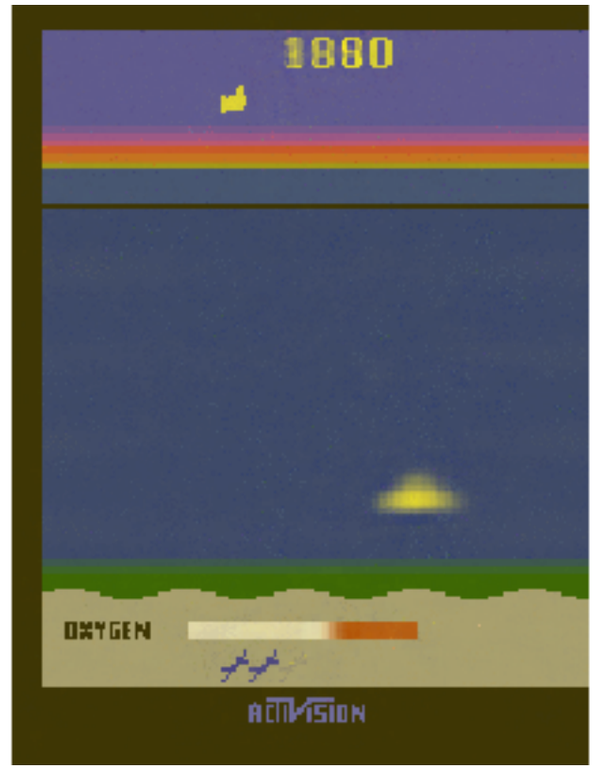




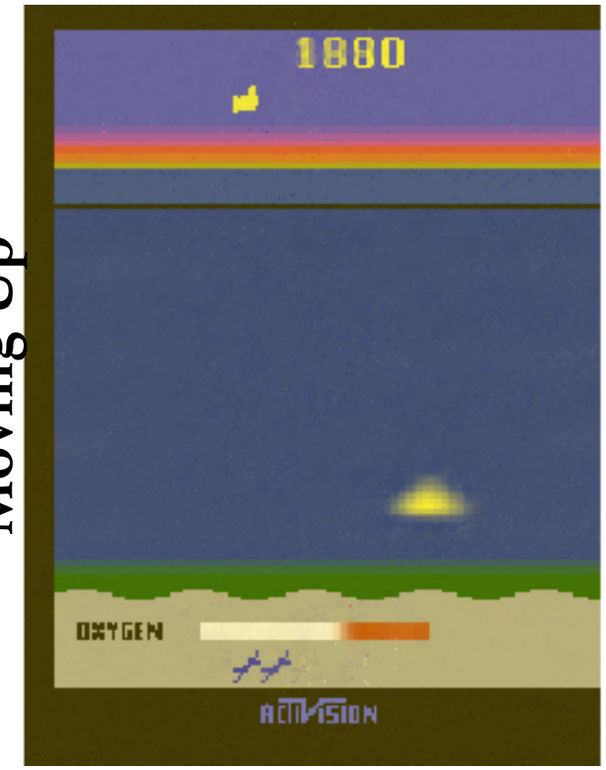
Concept Learning

Original

Score

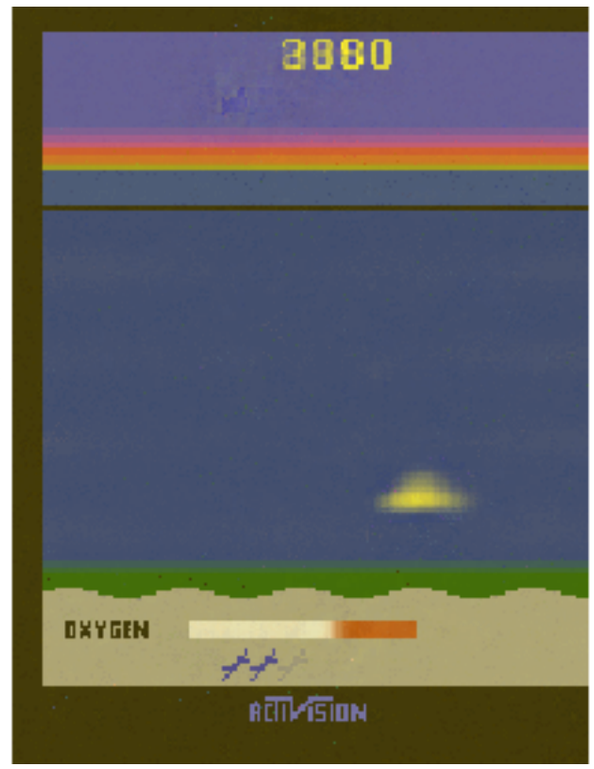


Moving Up

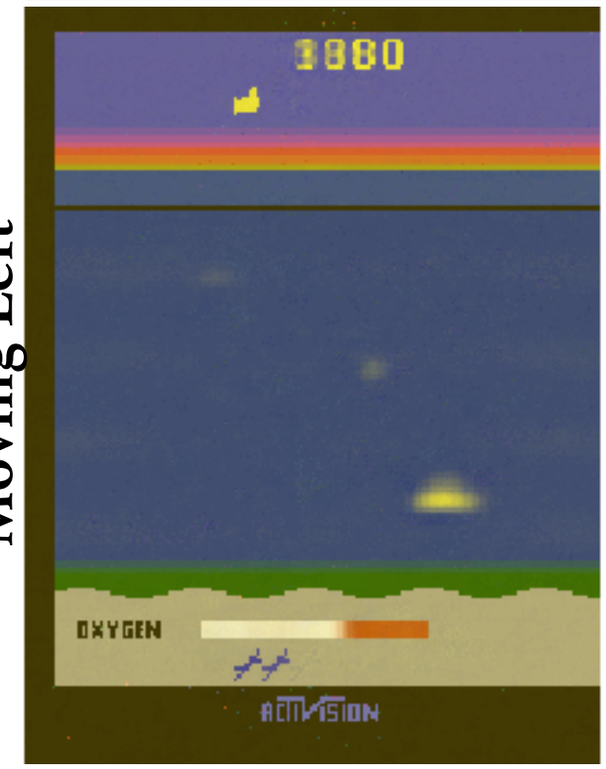


Oxygen/Swimmers

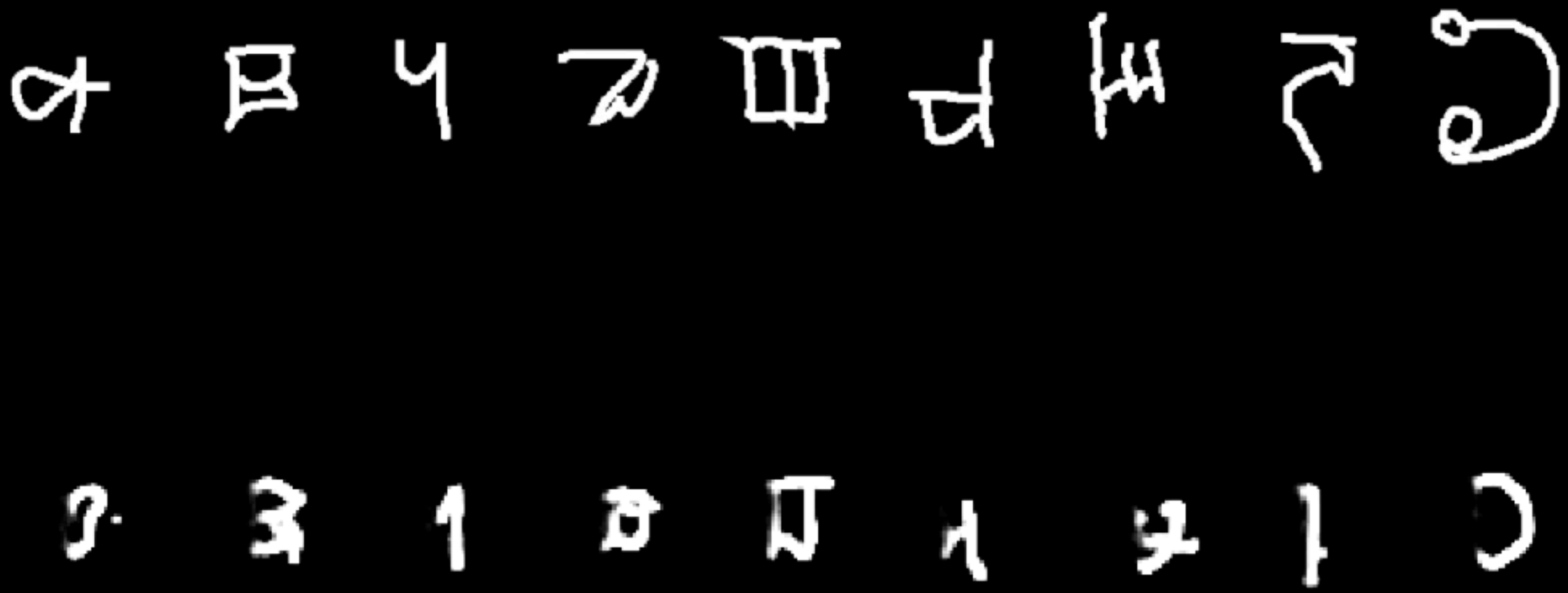
Score/Lives



Moving Left



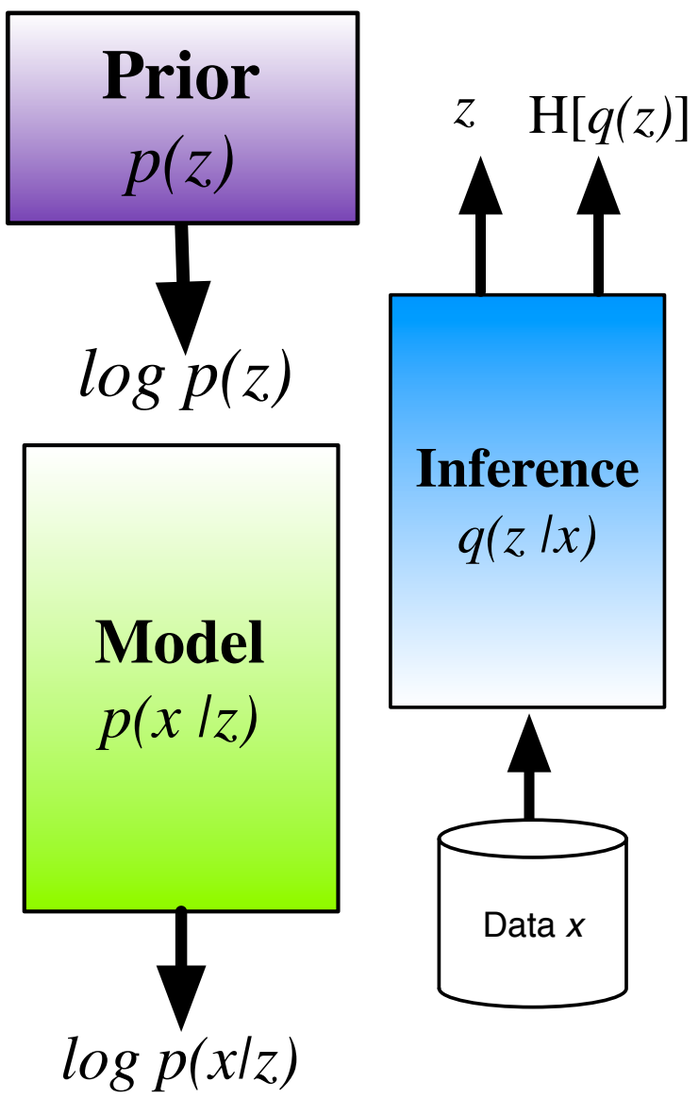
One-shot Generalisation

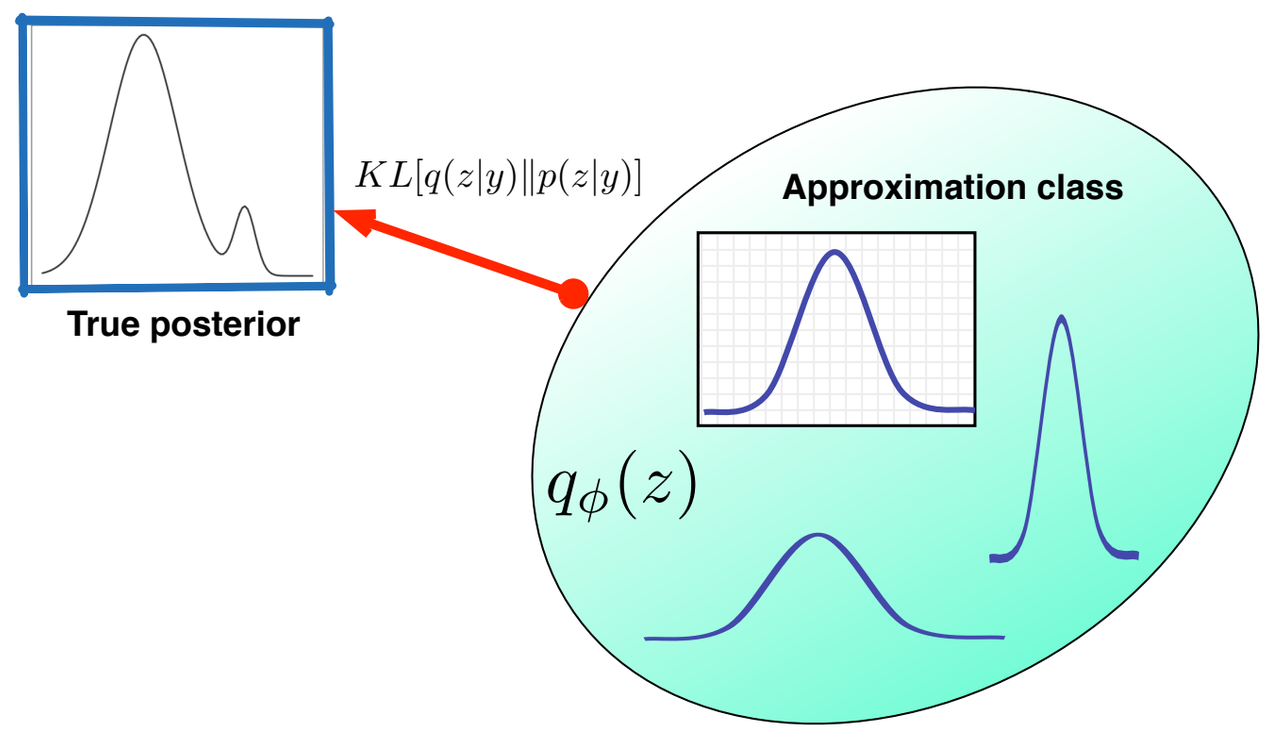


Latent Variable Models

Reconstruction — Penalty

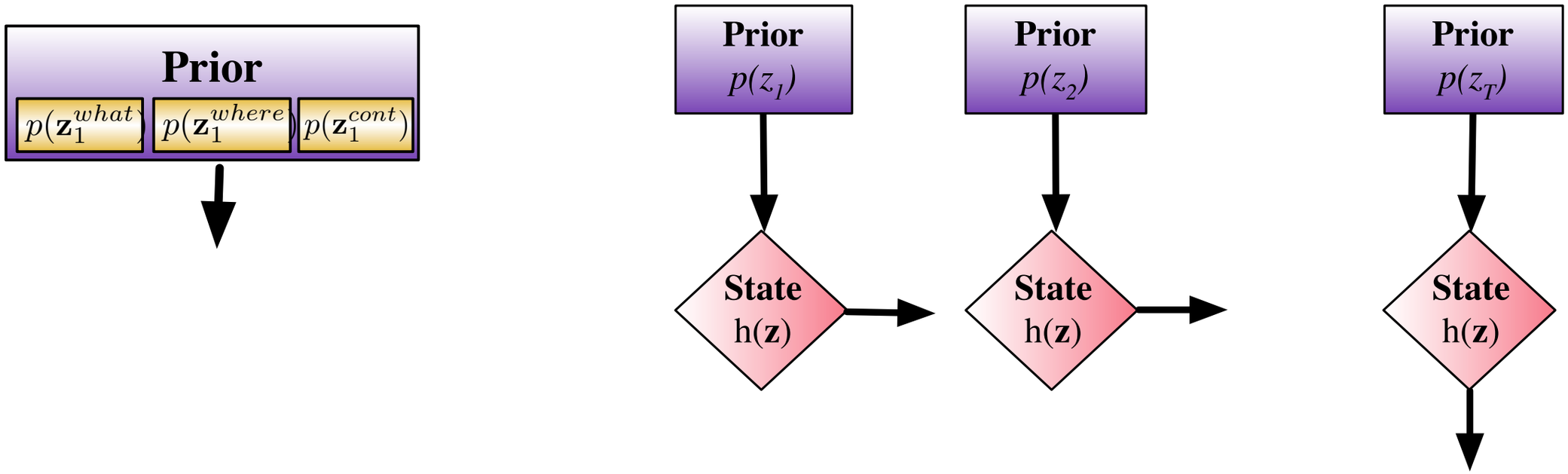
$$\mathcal{F}(y,q) = \mathbb{E}_{q(z)}[\log p(y|z)] - KL[q(z)\|p(z)]$$



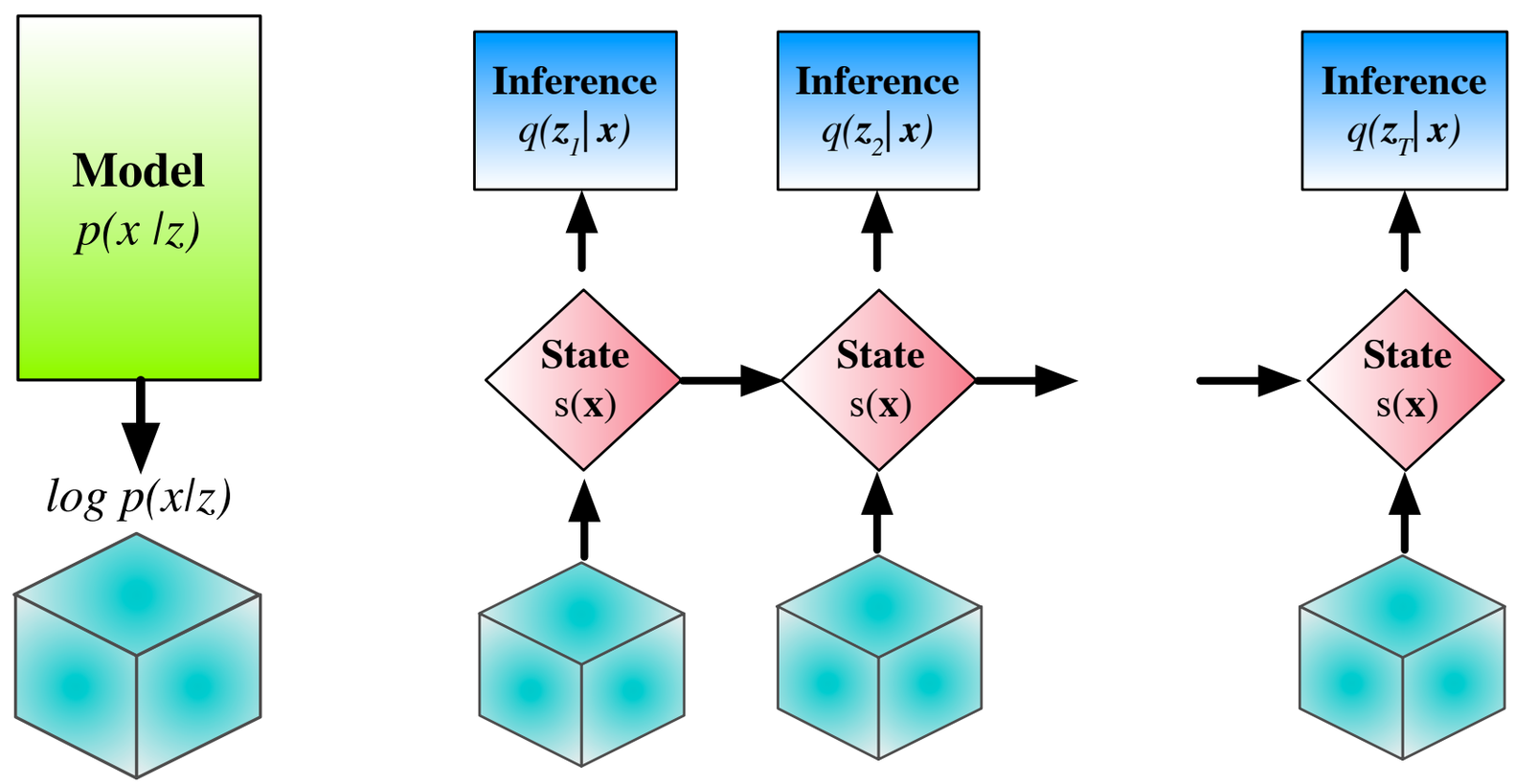


Variational inference is scalable and robust as a default approach for inference in deep probabilistic models.

Structured Models



- Model can be non-differentiable, like a graphics engine.
- Volume can represent colour channels, volumes, time.
- Use volumetric convolutions.

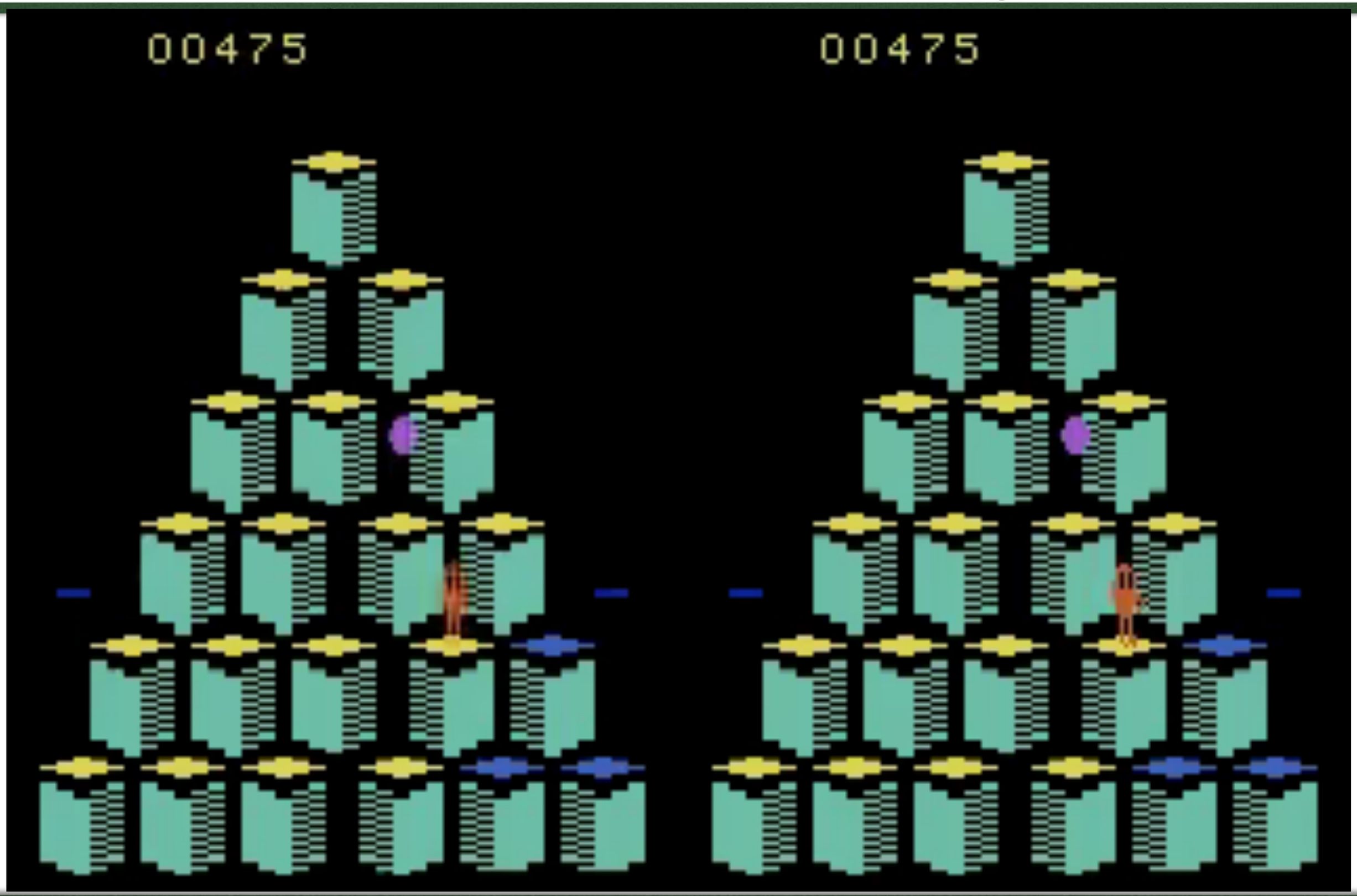


Grounded Cognition

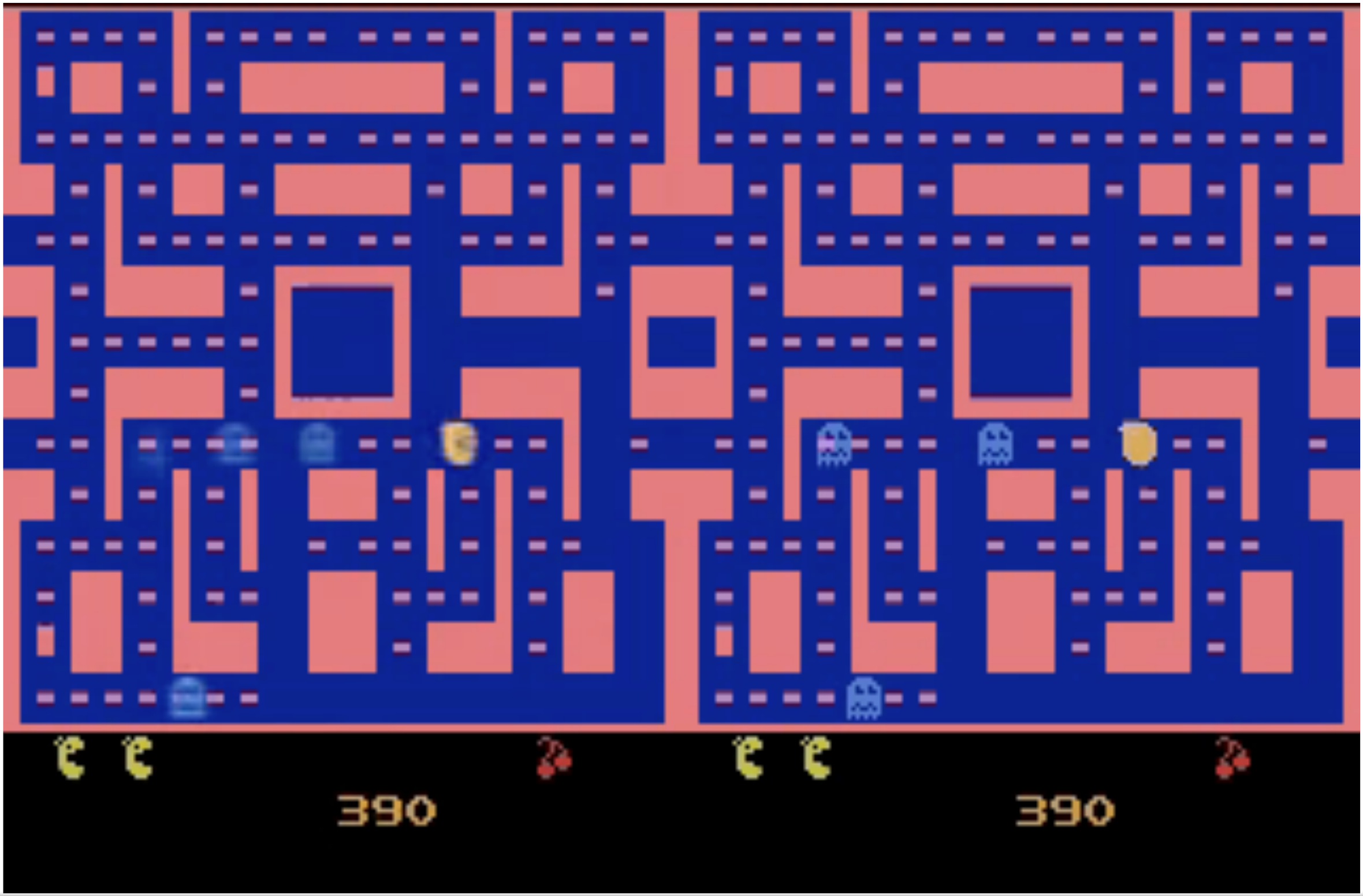
Cognitive Observation: People understand their environments and can make plans about the future in rapid and flexible ways.

Cognitive Inspiration: Simulations of environments are constructed and used to give grounded understanding of decisions, explanations and judgements.

Future Thinking



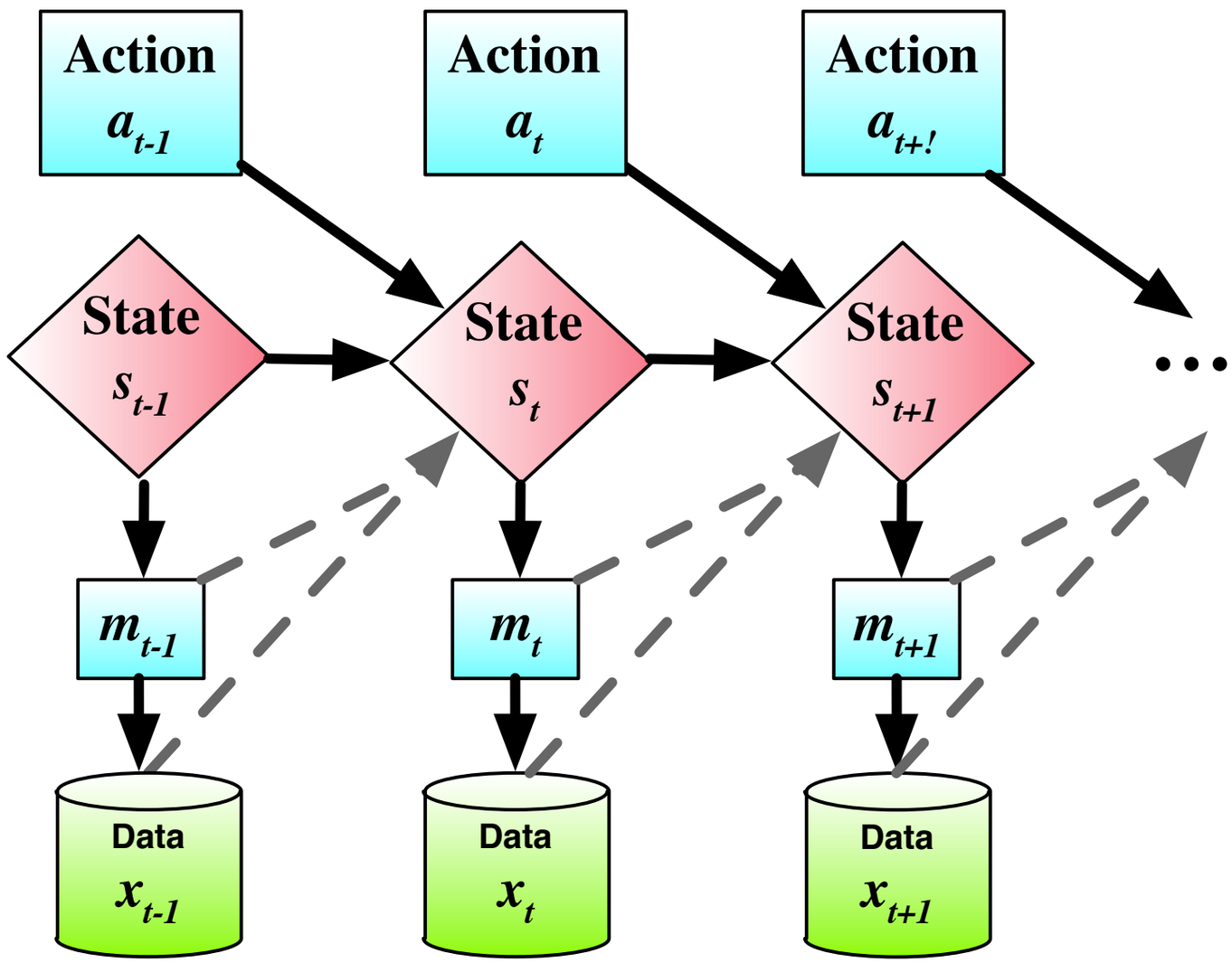
Future Thinking



Environment Simulation

Action-conditional and latent-only transitions.

Grounded representations in actions and observations, using simulation to support grounding.



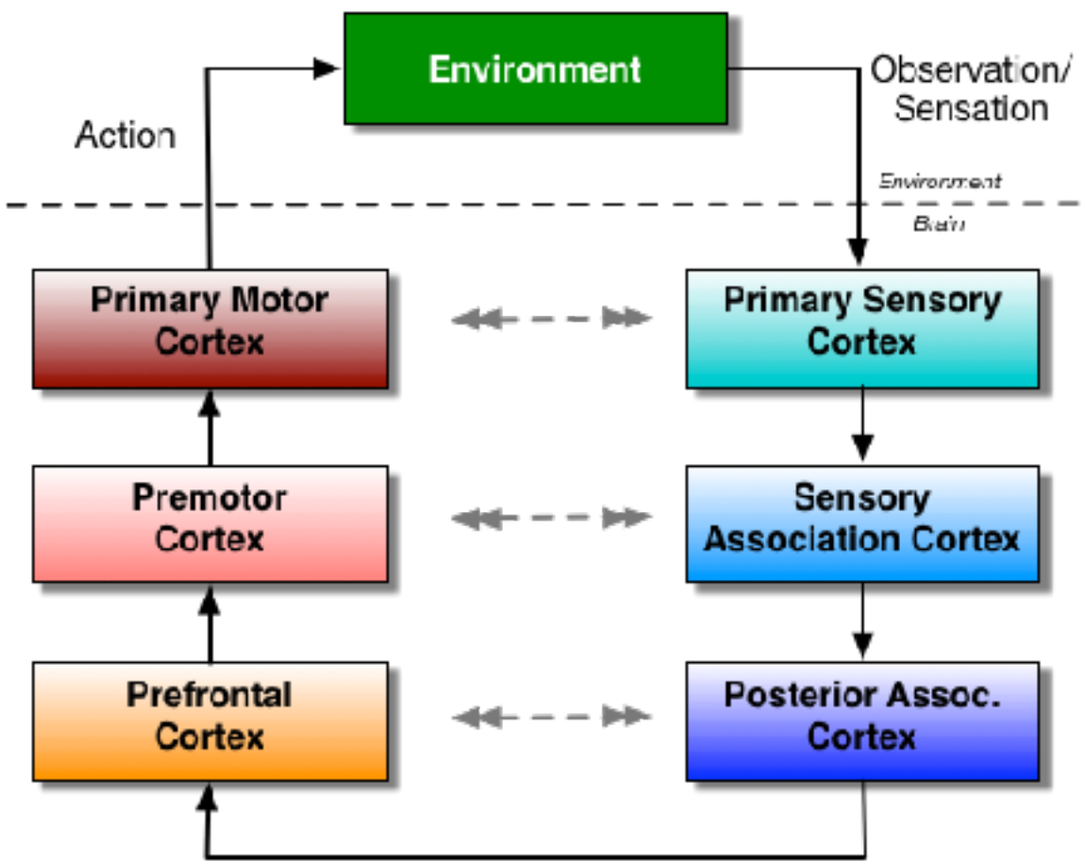
Intrinsic Motivation

Cognitive Observation: People don't always receive external rewards from their environments. Instead they engage in play, have fears, pain, joy, and are curious, which are internal rewards.

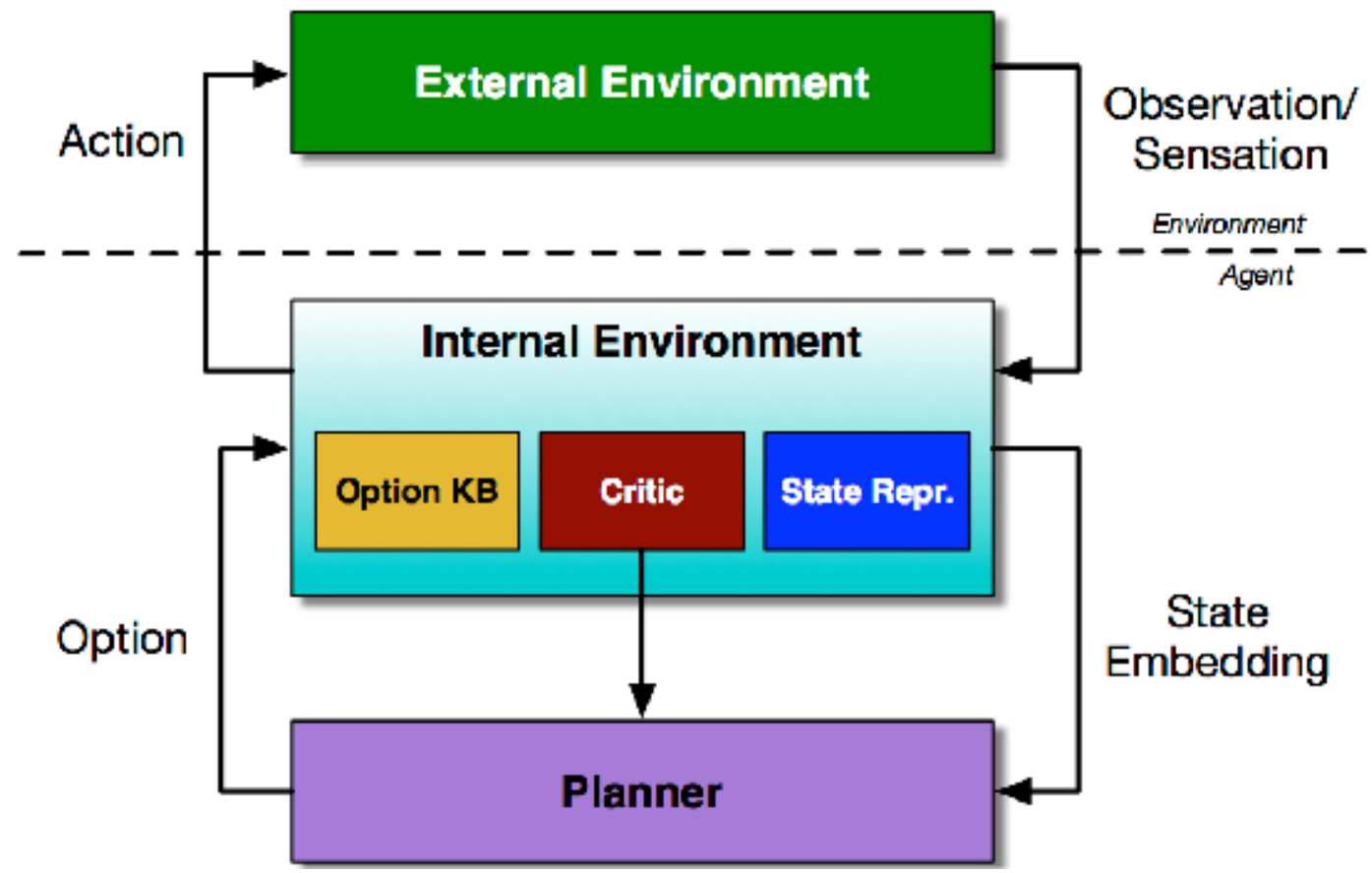
Cognitive Inspiration: Equip agents with mechanisms to produce and learn from internal rewards that can guide behaviour, when external rewards are absent.

Intrinsic Motivation

Biological perception-action loop

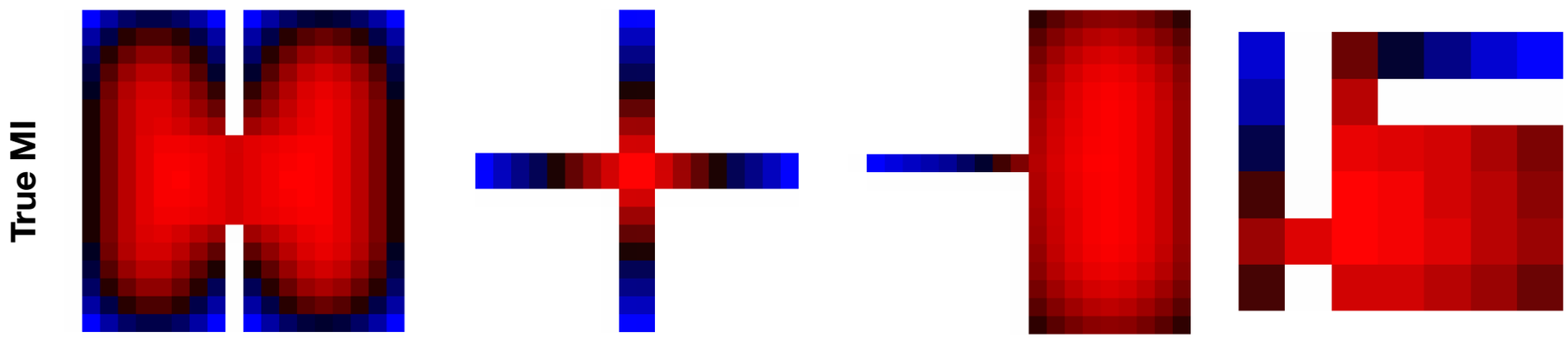


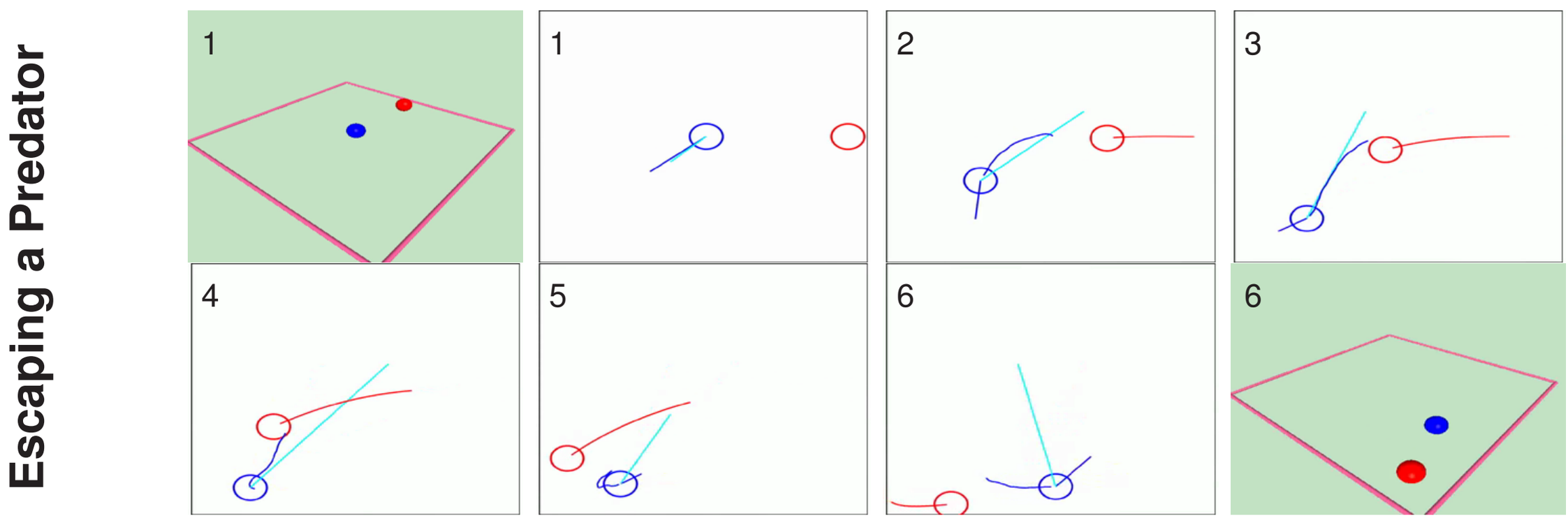
Computational perception-action loop



Empowerment

$$\mathcal{E}(\mathbf{s}) = \max_{\omega} \mathcal{I}^{\omega}(\mathbf{a}, \mathbf{s}'|\mathbf{s}) \qquad \max_{\omega} \mathbb{E}_{p(s'|a,s)\omega(a|s)} \left[\log \frac{p(\mathbf{s}', \mathbf{a}|\mathbf{s})}{p(\mathbf{s}'|\mathbf{s})\omega(\mathbf{a}|\mathbf{s})} \right]$$



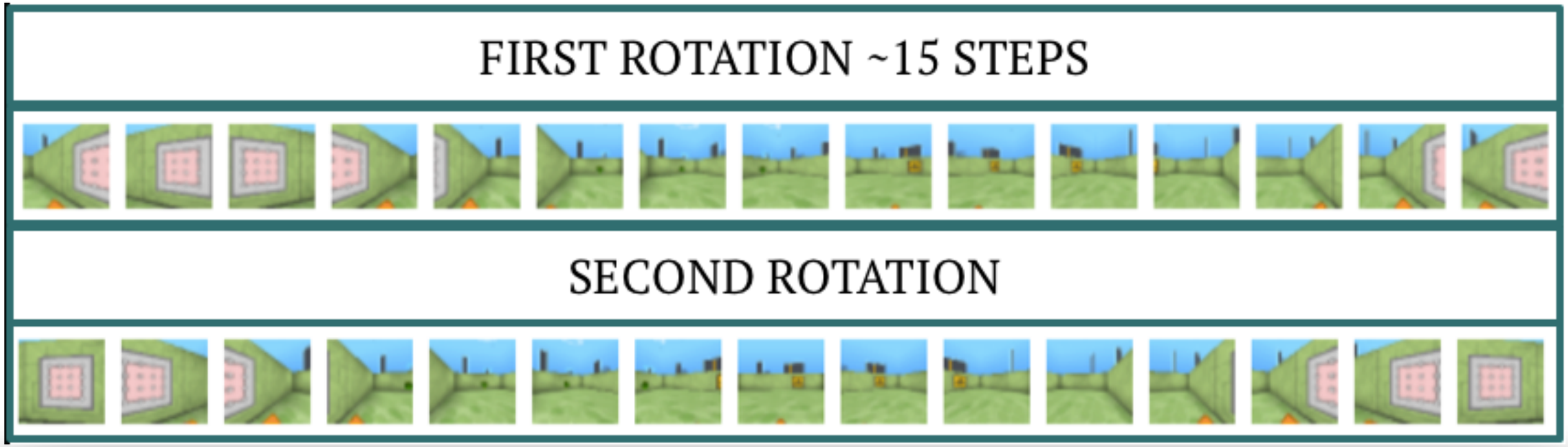


Memory and Coherence

Cognitive Observation: People are able to form associations between objects that are temporally distant. In addition they can introspect on aspects of their past, using their memories.

Cognitive Inspiration: How can we equip agents with external memory systems to allow for temporally coherent reasoning, and introspection.

Temporal Coherence

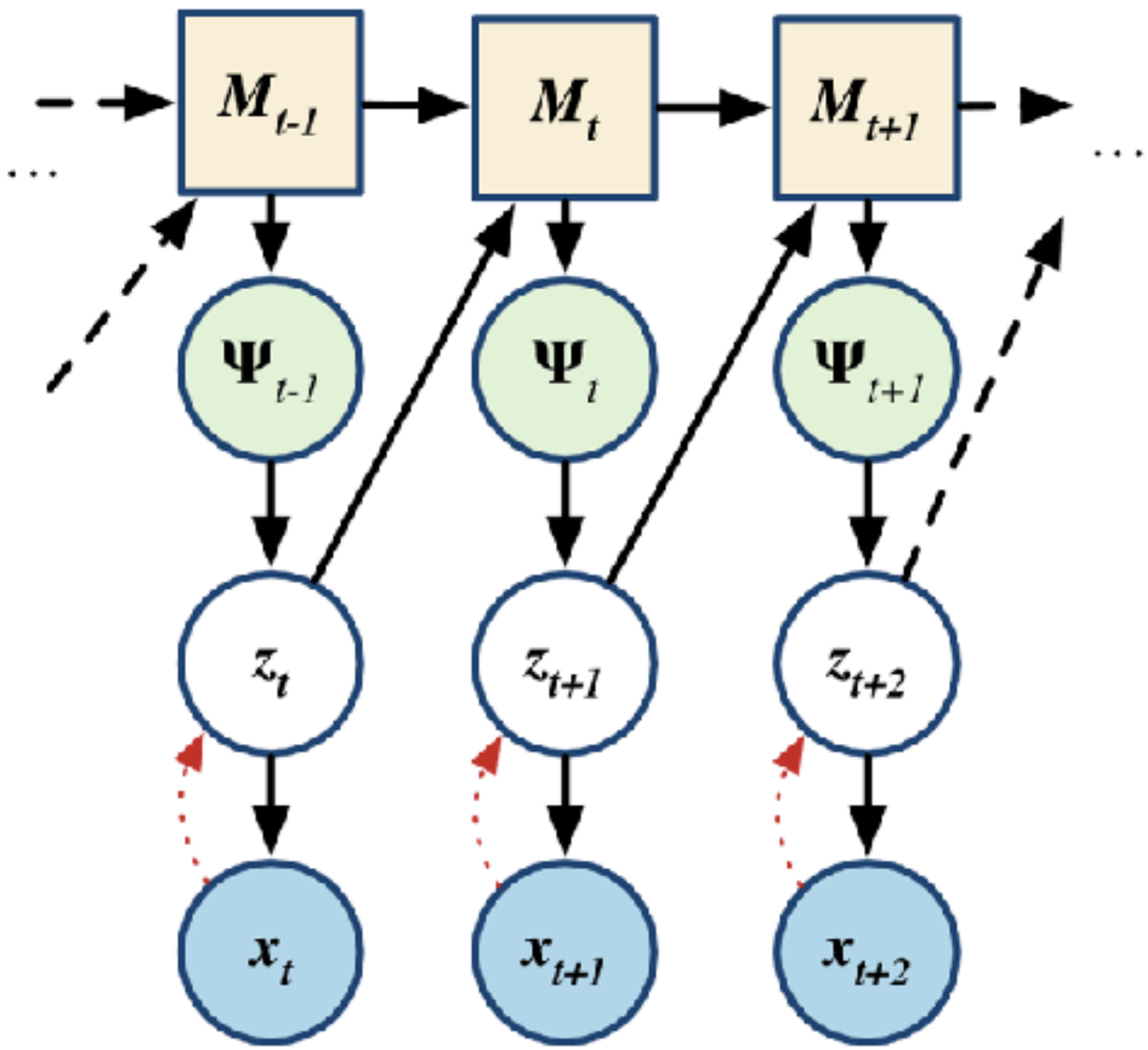


Memory and Recall

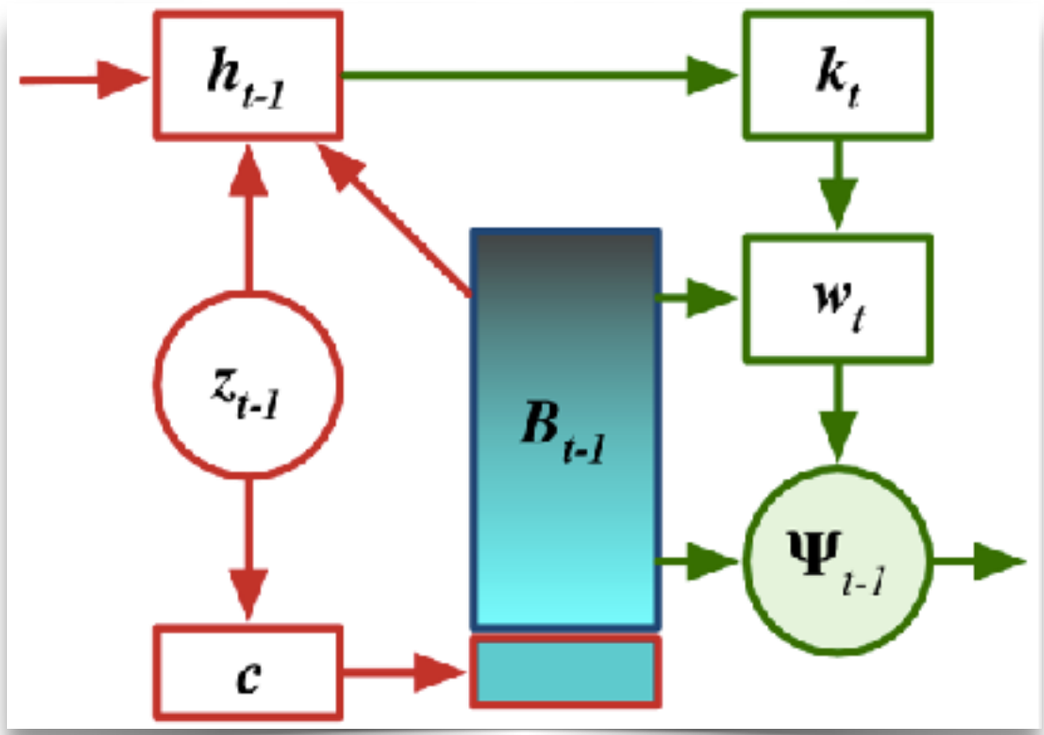
Recall

One-shot generalisation

Memory-augmented Models



Extend temporal latent variable models to include external memory and variational inference.



Common-sense Reasoning

World Simulation

Data-efficient Learning

Visual Concept Learning

Causal Reasoning

Complementary Learning

Hypothesis formation

Macro-actions and Planning

Exploration

Relational learning

Observations and Inspirations

Mutual Inspirations between Cognitive and Statistical Sciences

Shakir Mohamed

Research Scientist, DeepMind

shakir@deepmind.com

@shakir_za

Sheffield Machine Learning Research Retreat 2017