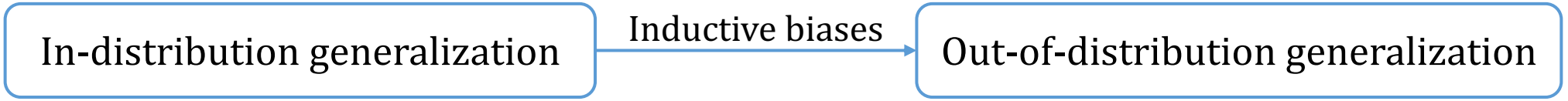
Inductive Biases for Deep Learning of Higher-Level Cognition

Inductive Biases for Deep Learning of Higher-Level Cognition



Inductive Biases

- Algorithms
- Data
- Ease of learning

Systematic Generalization

OOD Generalization

Agent

Policy

Non-stationary data streams

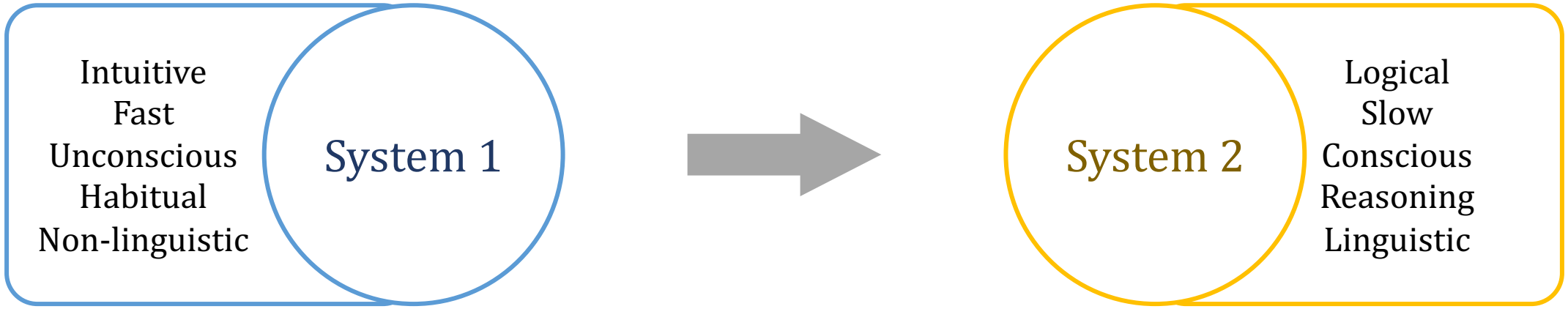
Transfer Learning

Continual Learning

Meta-learning

Inductive Biases for Deep Learning of Higher-Level Cognition

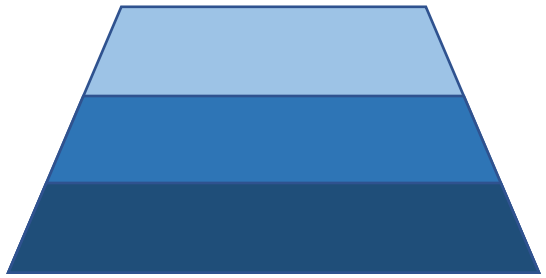
Cognitive Systems



Current DL → System 2 DL

Inductive Biases for Deep Learning of Higher-Level Cognition

High-level semantic variables
(concepts)



Low-level
(pixels, natural language, etc.)

Changes
Interventions

Sparse Factor Graph

Attention

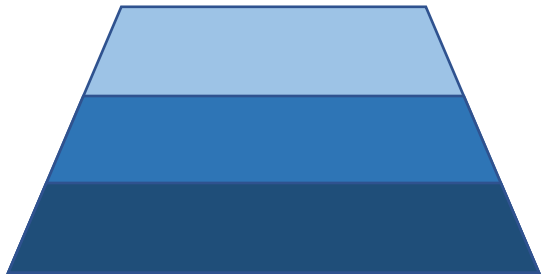
Dynamically recombined modules

Meta-learning

Causality

Inductive Biases for Deep Learning of Higher-Level Cognition

High-level semantic variables
(concepts)



Low-level
(pixels, natural language, etc.)

Changes
Interventions

Sparse Factor Graph

Attention

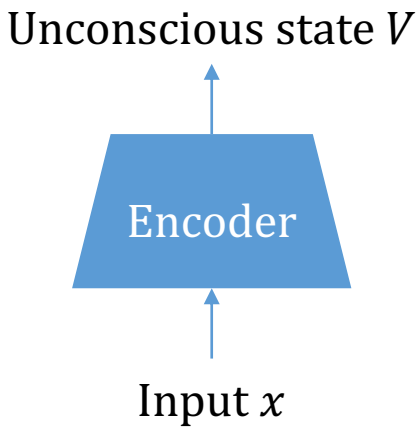
Causality

Dynamically recombined modules

Meta-learning

Inductive Biases for Deep Learning of Higher-Level Cognition

Sparse Factor Graph



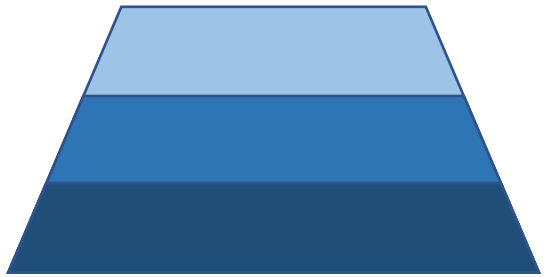
- Natural language is sparse (pixel space is not)
- Consciousness prior (put pressure on encoder)
-

$$P(V) \propto \prod_k \phi_k(V_{s_k})$$

- V_{s_k} is the subset of V with indices s_k, which involves **only a few** variables.

Inductive Biases for Deep Learning of Higher-Level Cognition

High-level semantic variables
(concepts)



Low-level
(pixels, natural language, etc.)

Changes
Interventions

Sparse Factor Graph

Attention

Causality

Dynamically recombined modules

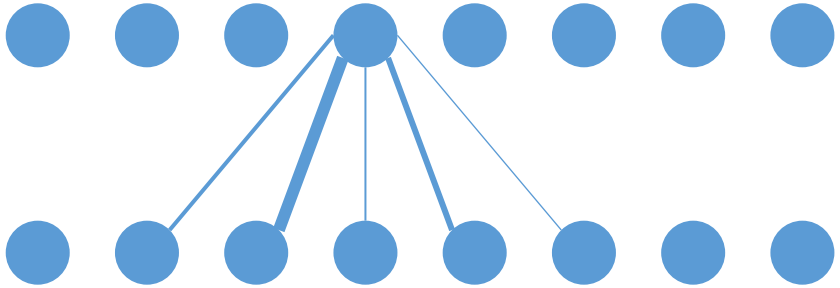
Meta-learning

Inductive Biases for Deep Learning of Higher-Level Cognition

Attention

- Content based soft attention
- Operate on sets
- Dynamic connection
- What is the object of attention? → named variable

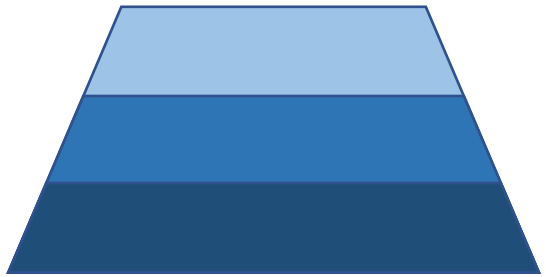
Inspiration: programming language



$$\text{Attention}(Q, K, V) = \text{softmax}(\frac{QK^T}{\sqrt{d_k}})V$$

Inductive Biases for Deep Learning of Higher-Level Cognition

High-level semantic variables
(concepts)



Low-level
(pixels, natural language, etc.)

Changes
Interventions

Sparse Factor Graph

Attention

Causality

Dynamically recombined modules

Meta-learning

Inductive Biases for Deep Learning of Higher-Level Cognition

Dynamically recombined modules

- Global Workspace Theory

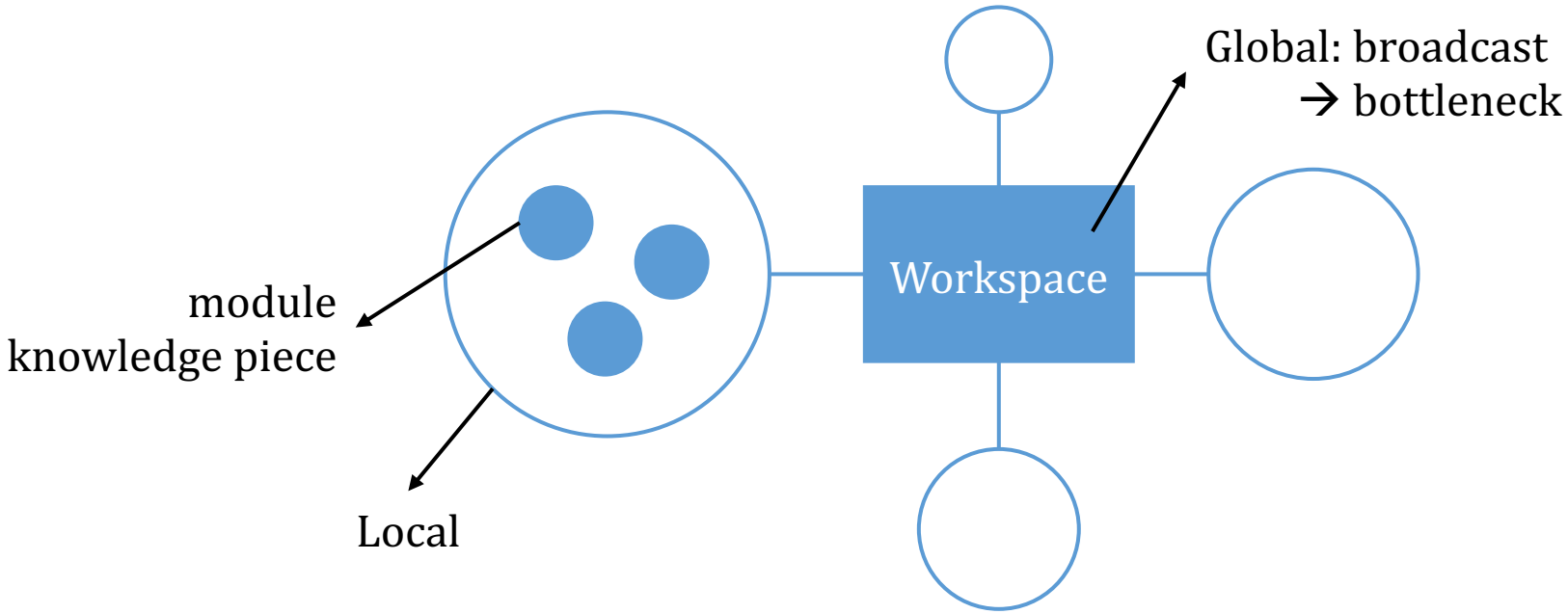
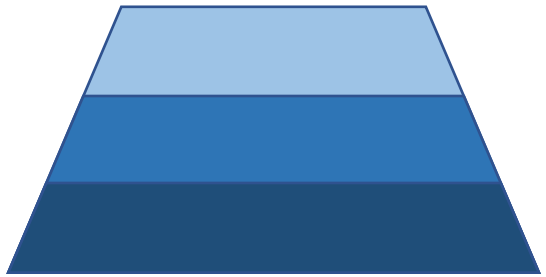


Image is taken directly from Bengio's presentation on NeurIPS2019

Inductive Biases for Deep Learning of Higher-Level Cognition

High-level semantic variables
(concepts)



Low-level
(pixels, natural language, etc.)

Changes
Interventions

Sparse Factor Graph

Attention

Causality

Dynamically recombined modules

Meta-learning

Inductive Biases for Deep Learning of Higher-Level Cognition

Meta-learning

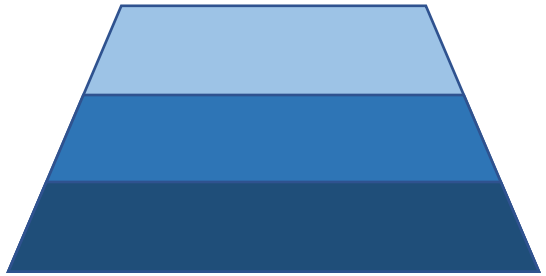
- Outer loop
  - Slow learning
  - Stable and stationary
- Inner loop
  - Fast learning
  - Adapt to new changes
  - Task-specific

- OOD Generalization
  - Intervention cause distribution changes
  - Change is localized

Image is taken directly from Bengio's presentation on NeurIPS2019

Inductive Biases for Deep Learning of Higher-Level Cognition

High-level semantic variables
(concepts)



Low-level
(pixels, natural language, etc.)

Changes
Interventions

Sparse Factor Graph

Attention

Dynamically recombined modules

Meta-learning

Causality

Inductive Biases for Deep Learning of Higher-Level Cognition

Causality

- Statistical correlation ≠ Causality

- Independent Causal Mechanisms (ICM)
- ❌ Bayesian Network
- ❌ Structural Causal Models
- ✅ Causal Factor Graph

- Nature doesn't shuffle examples.

- Challenge: Joint discovery of
  1. High-level representations
  2. Causal structure at the high level