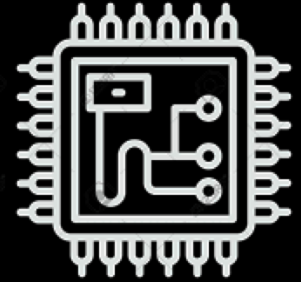


# **Abstraction: Symbolic Processing in the Brain**

# Background



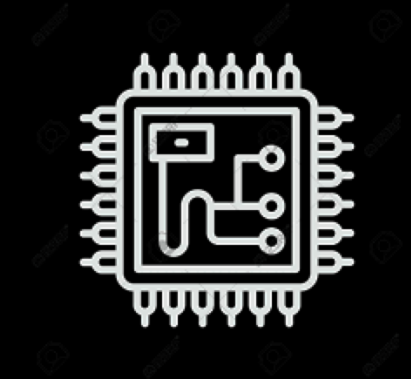
# Background

- **Miracle of traditional symbolic computing:**



**Computationally general: maximum flexibility**

♦ **existence proof of flexibility of human cognitive ability**



# Background

- **Miracle of traditional symbolic computing:**



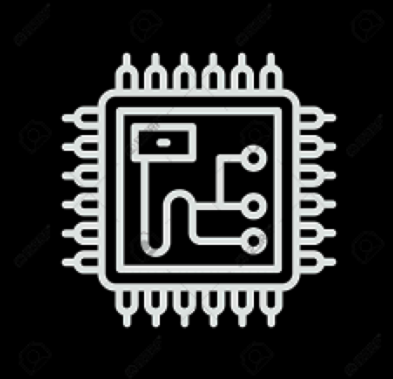
**Computationally general: maximum flexibility**

- ♦ **existence proof of flexibility of human cognitive ability**



**Inefficient and/or difficult to configure for complex domains**

- **vision, natural language**





# Background

- **Miracle of traditional symbolic computing:**



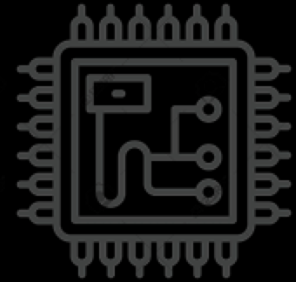
Computationally general: maximum flexibility

- ♦ existence proof of flexibility of human cognitive ability



Inefficient and/or difficult to configure for complex domains

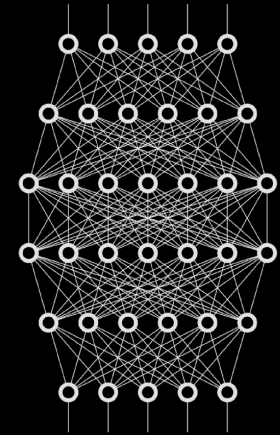
- vision, natural language



- **Miracle of deep learning:**



Computationally efficient: automated function approximation



# Background

- **Miracle of traditional symbolic computing:**



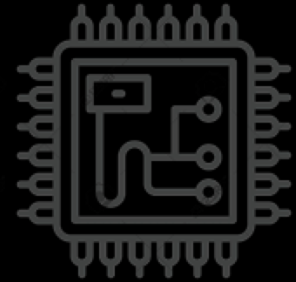
**Computationally general: maximum flexibility**

- ♦ existence proof of flexibility of human cognitive ability



**Inefficient and/or difficult to configure for complex domains**

- vision, natural language



- **Miracle of deep learning:**

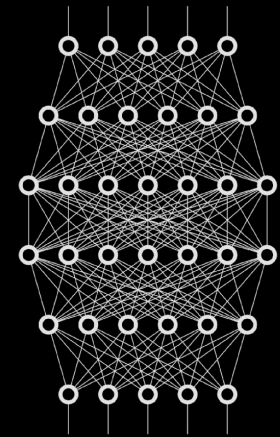


**Computationally efficient: automated function approximation**



**Inflexible:**

- ♦ sample-inefficient
- ♦ domain-specific



# Background

- **Miracle of traditional symbolic computing:**



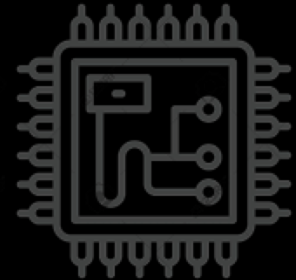
Computationally general: maximum flexibility

- ♦ existence proof of flexibility of human cognitive ability



Inefficient and/or difficult to configure for complex domains

- vision, natural language



- **Miracle of deep learning:**



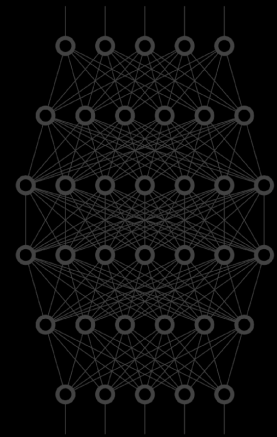
Computationally efficient: automated function approximation



Inflexible:

- ♦ sample-inefficient

- ♦ domain-specific



- **So where are we?**

# Clash of the Titans

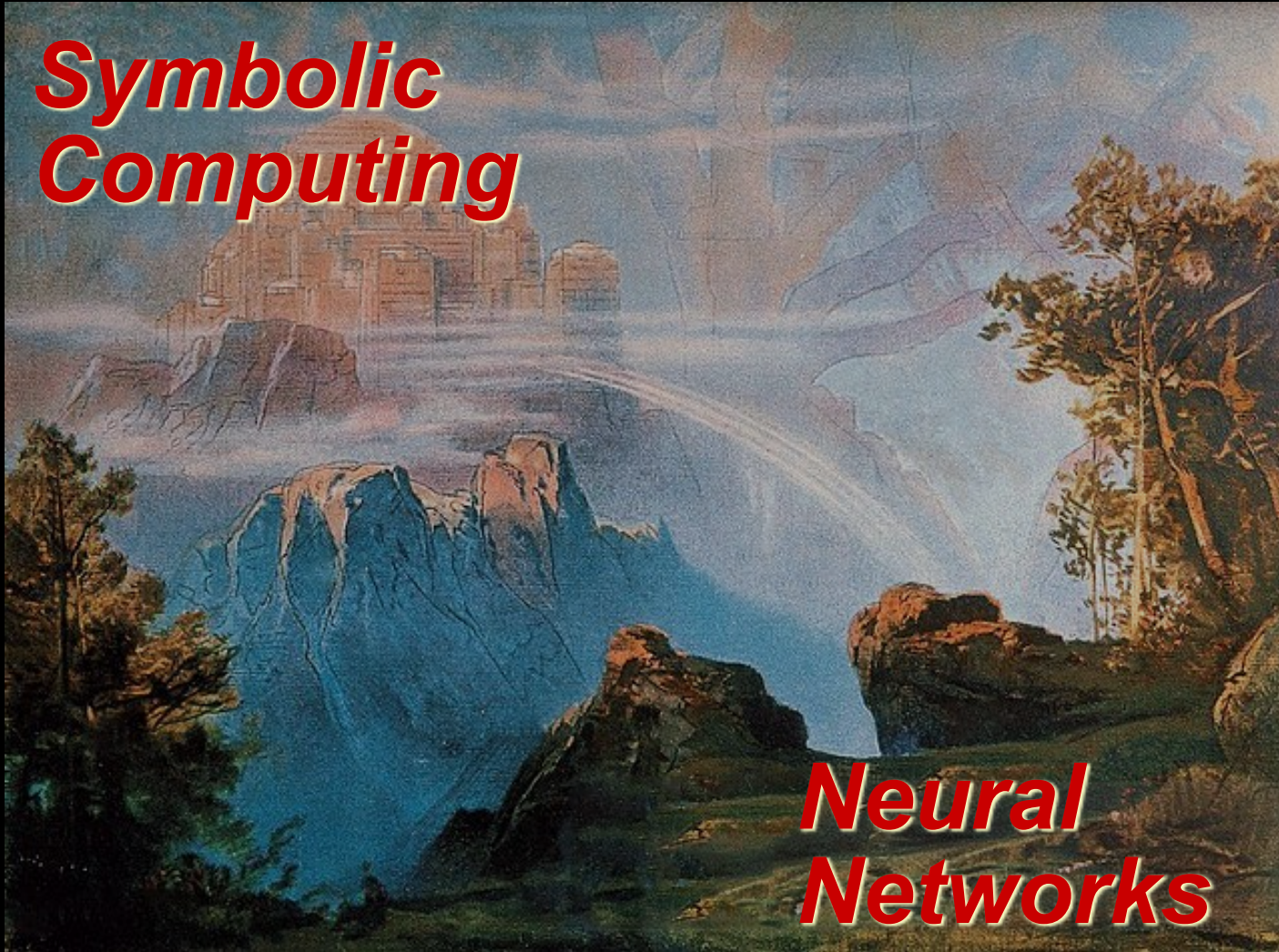




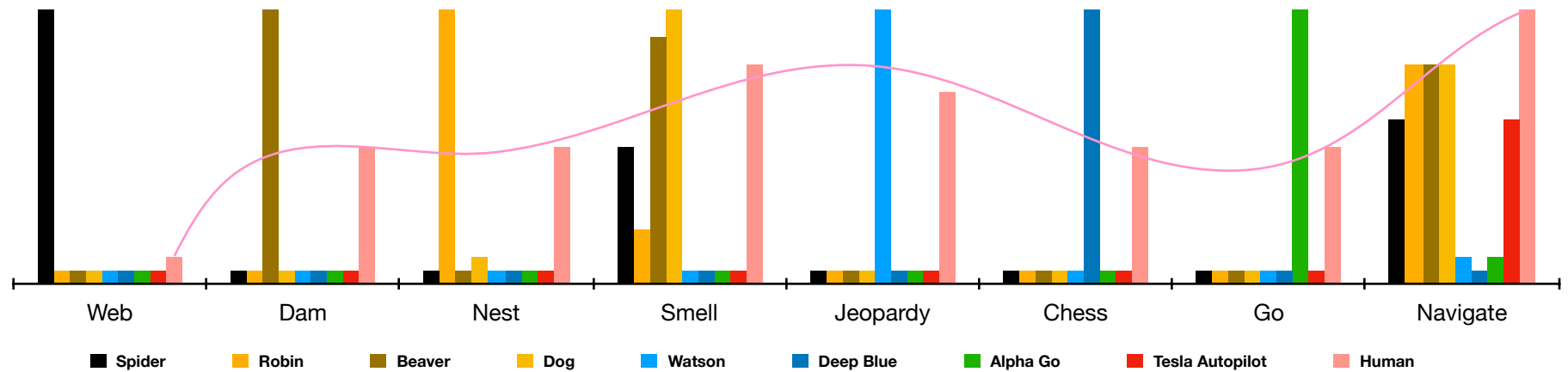
# Shangri-La

*Symbolic  
Computing*

*Neural  
Networks*

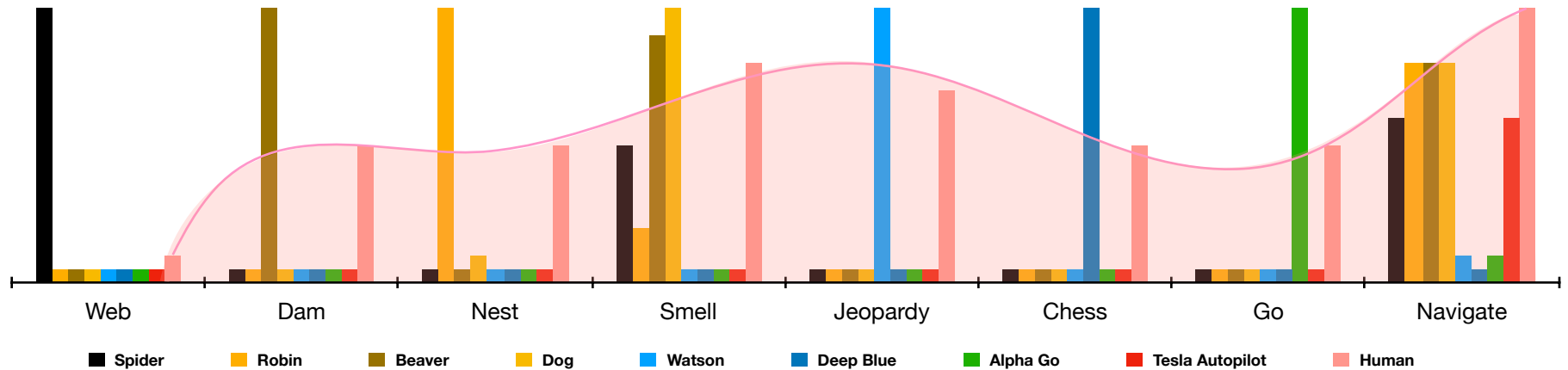


# Human Brain = Existence Proof





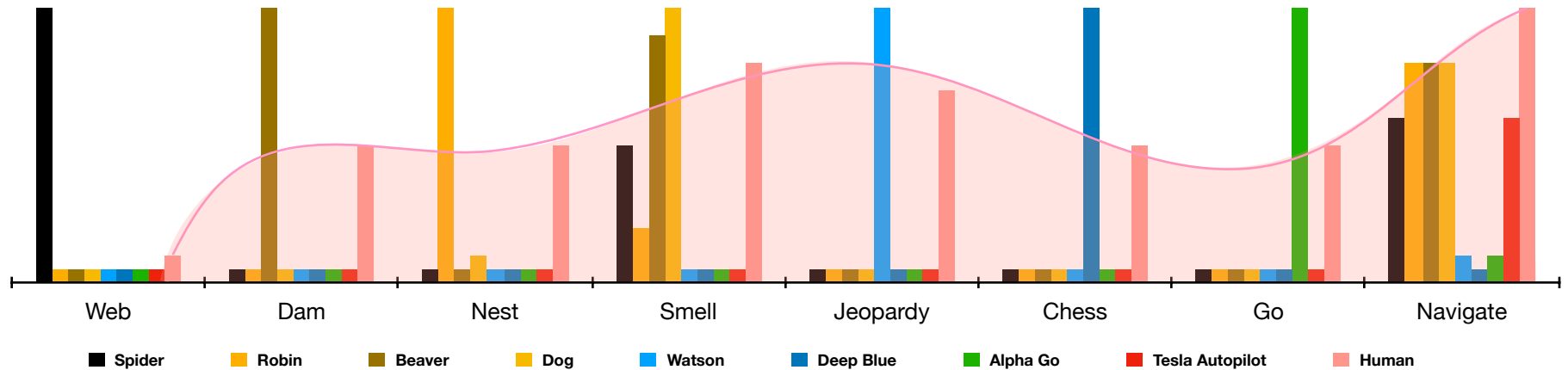
# Human Brain = Existence Proof



**“Sweet spot” between flexibility and efficiency**

- Near limitless range of tasks at adequate performance - flexibility

# Human Brain = Existence Proof

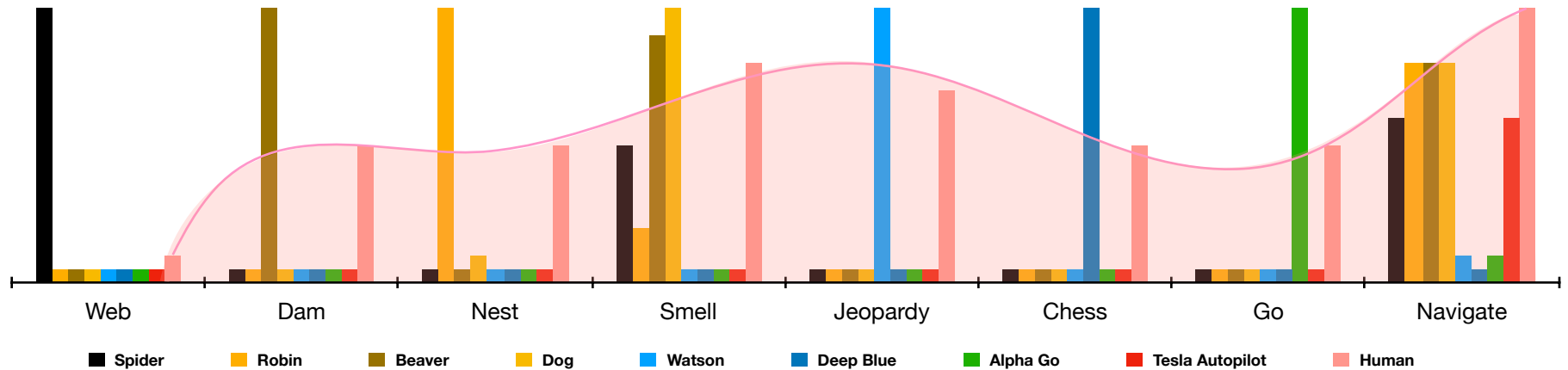


**“Sweet spot” between flexibility and efficiency**

- With *reasonable* amounts, and often little or *no* training - sample efficiency



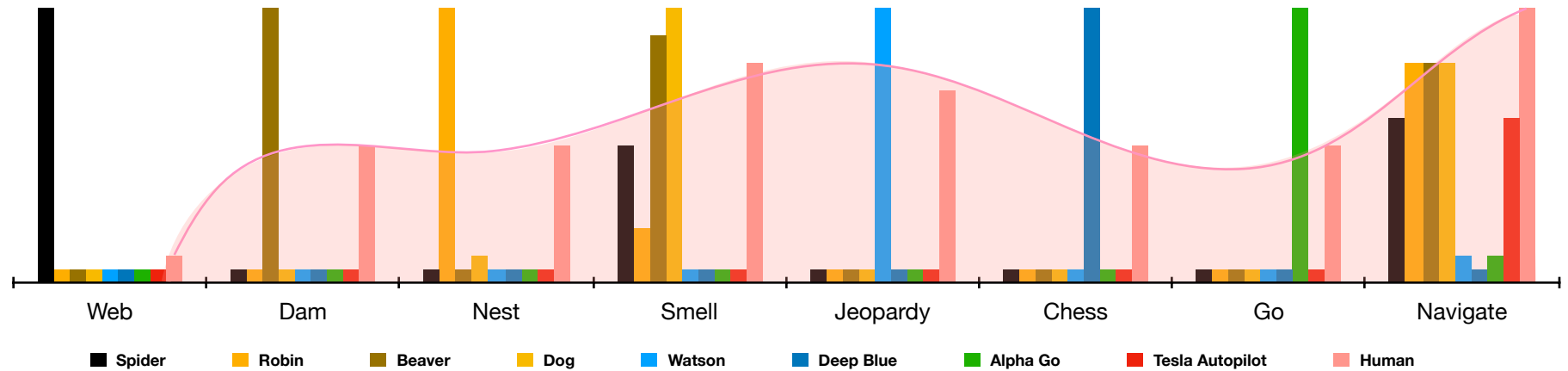
# Human Brain = Existence Proof



**“Sweet spot” between flexibility and efficiency**

- ~20 watts, often with parallel performance - processing efficiency

# Human Brain = Existence Proof



How does it accomplish this?



# Shangri-La?





# Shangri-La?

- Challenge:

- Integrate flexibility of *symbolic processing* in traditional architectures
- with efficiency of *function approximation* in neural networks





# Shangri-La?

- Current efforts:

- **Neuro-symbolic** approaches:

- ♦ start with pre-specified symbolic primitives (“core knowledge”)
    - ♦ use deep learning to combine these (e.g., “program induction”)



# Shangri-La?

- Current efforts:

- “**Neo-connectionist**” approaches:

- ♦ use deep learning for “end-to-end” training of neural networks



# Shangri-La?

- Current efforts:

- “Neo-connectionist” approaches:

- ♦ inductive biases that favor abstraction
  - *training*: curricular learning, meta learning
  - *architecture & processing*: attention, external memory



# Shangri-La?

- Still not there...



# Two critical ingredients:

*Symbolic  
Modeling*

*Neural  
Networks*



# Two critical ingredients:

*Symbolic  
Modeling*

**Abstraction  
and  
Autonomy**

*Neural  
Networks*



# Two critical ingredients:

*Symbolic  
Modeling*

**Self-Reconfiguration**



*Neural  
Networks*



# Two critical ingredients:

*Symbolic  
Modeling*

**Relational Bottleneck  
and**

*Neural  
Networks*



# **Abstraction**

*The search for (low dimensional) structure*

# Abstraction

*The search for (low dimensional) structure*

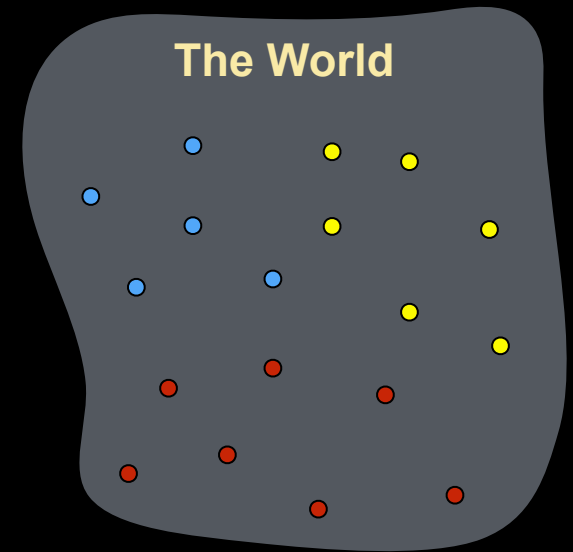
- Usually evaluated by capacity for generalization:

# Abstraction

*The search for (low dimensional) structure*

- Usually evaluated by capacity for generalization:

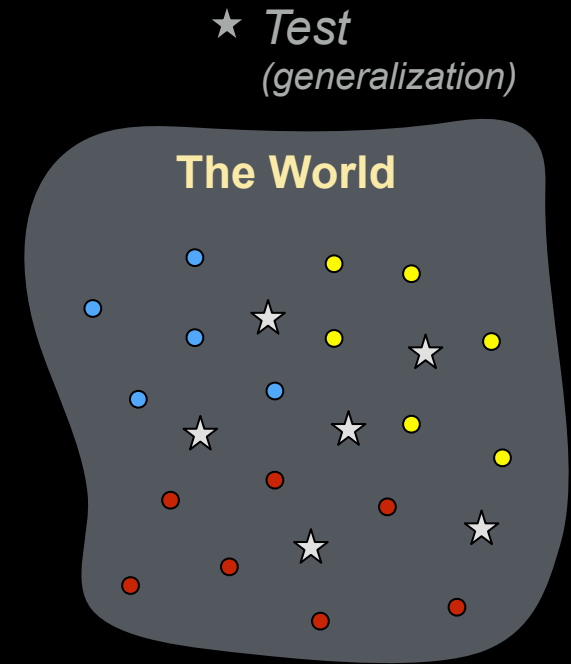
- Train  
(experience)



# Abstraction

*The search for (low dimensional) structure*

- Usually evaluated by capacity for generalization:

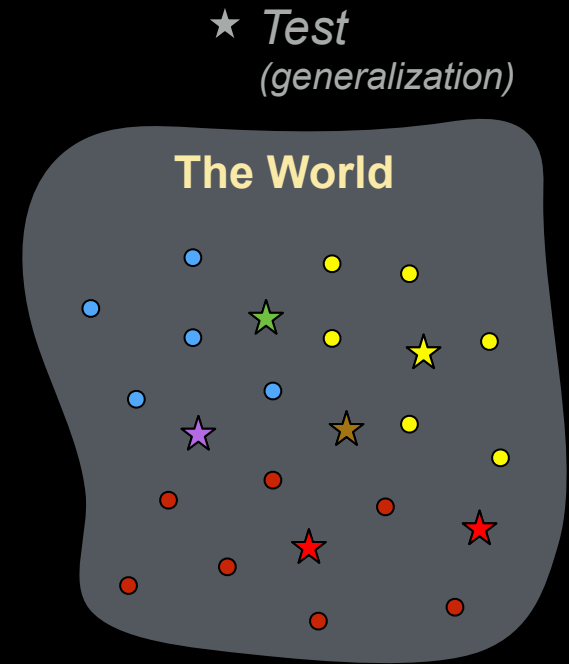




# Abstraction

*The search for (low dimensional) structure*

- Usually evaluated by capacity for generalization:
  - Interpolation  
*(out of sample)*



# Abstraction

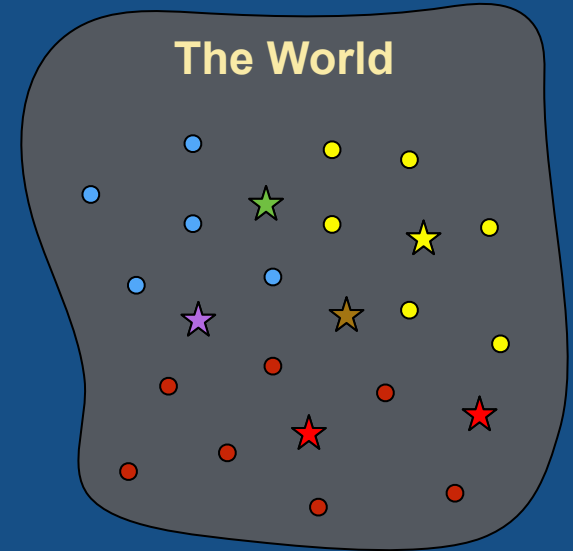
*The search for (low dimensional) structure*

- Usually evaluated by capacity for generalization:
  - Interpolation  
*(out of sample)*

## The Universe

★ Test  
*(generalization)*

The World



# Abstraction

*The search for (low dimensional) structure*

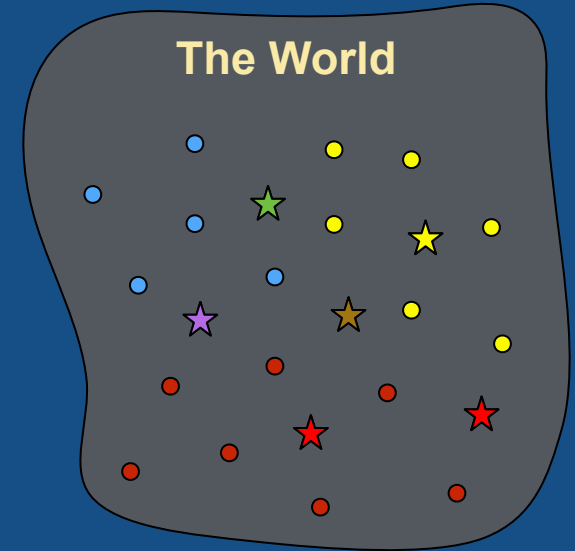
- Usually evaluated by capacity for generalization:
  - Interpolation  
*(out of sample)*
  - Extrapolation  
*(out of distribution)*

## The Universe

★ Test  
*(generalization)*

### The World

?  
★



# Abstraction

*The search for (low dimensional) structure*

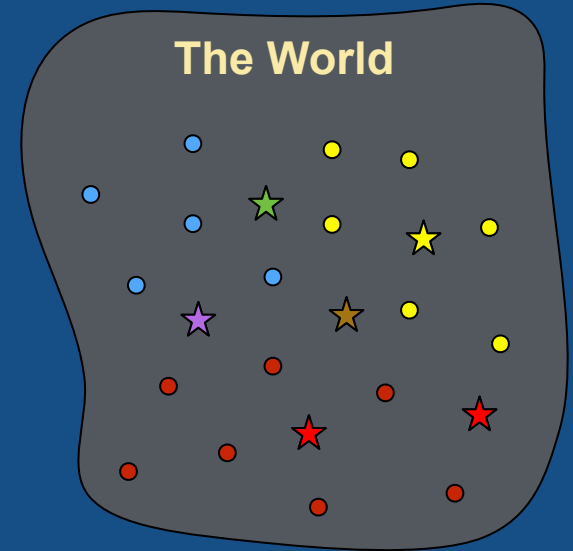
- Usually evaluated by capacity for generalization:
  - Interpolation  
*(out of sample)*
  - Extrapolation  
*(out of distribution)*

## The Universe

★ Test  
*(generalization)*

The World

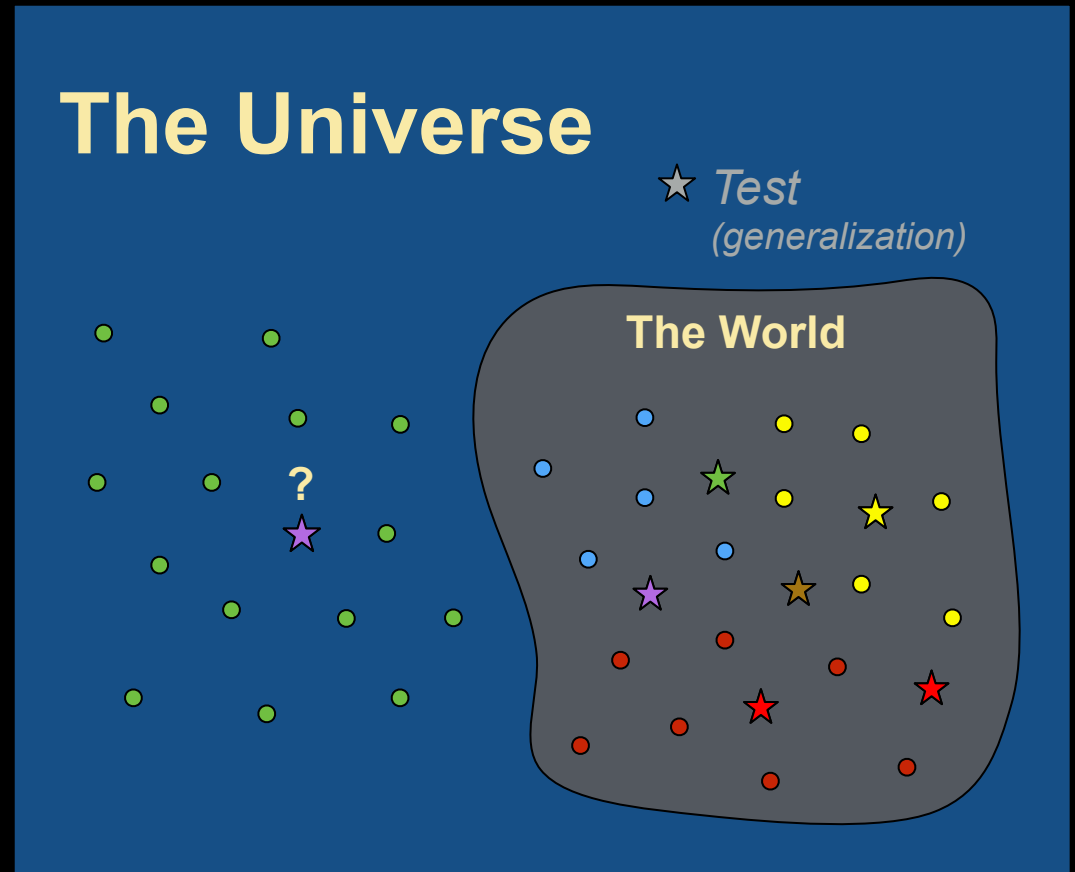
?



# Abstraction

*The search for (low dimensional) structure*

- Usually evaluated by capacity for generalization:
  - Interpolation  
*(out of sample)*
  - Extrapolation  
*(out of distribution)*



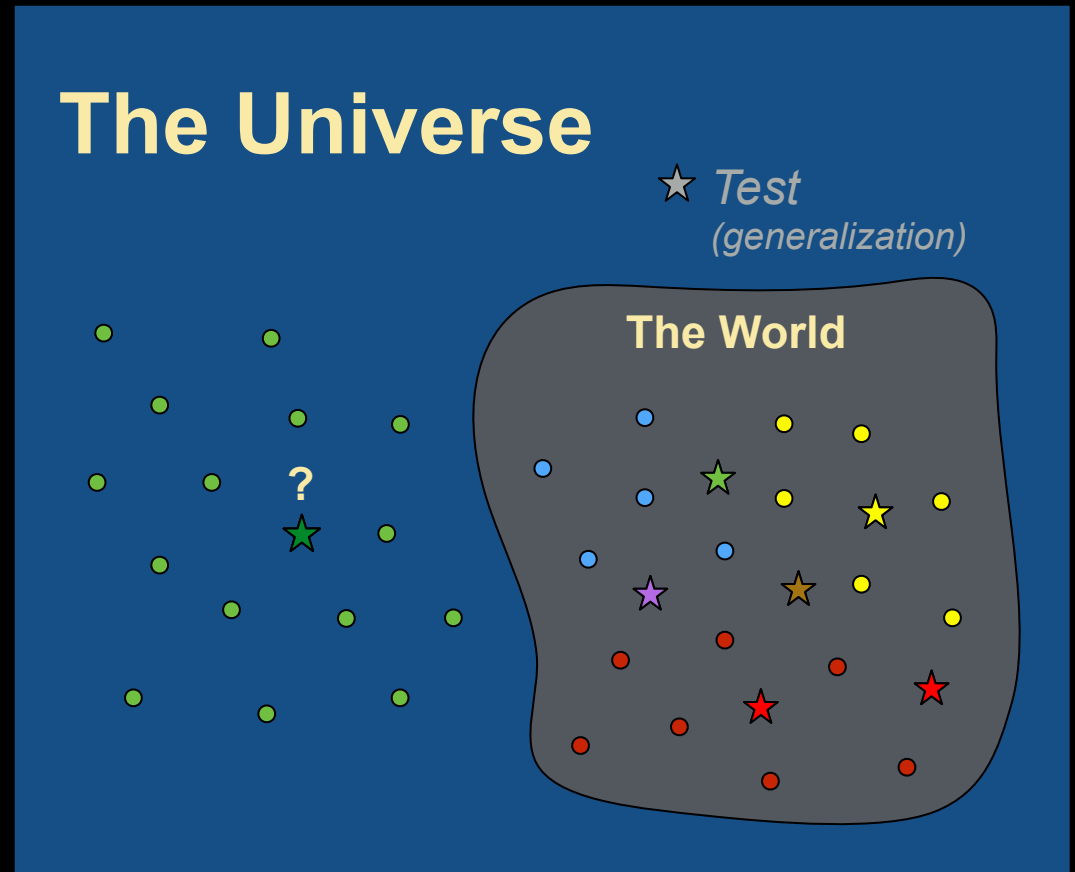
# Abstraction

*The search for (low dimensional) structure*

- Usually evaluated by capacity for generalization:

- **Interpolation**  
*(out of sample)*

- **Extrapolation**  
*(out of distribution)*



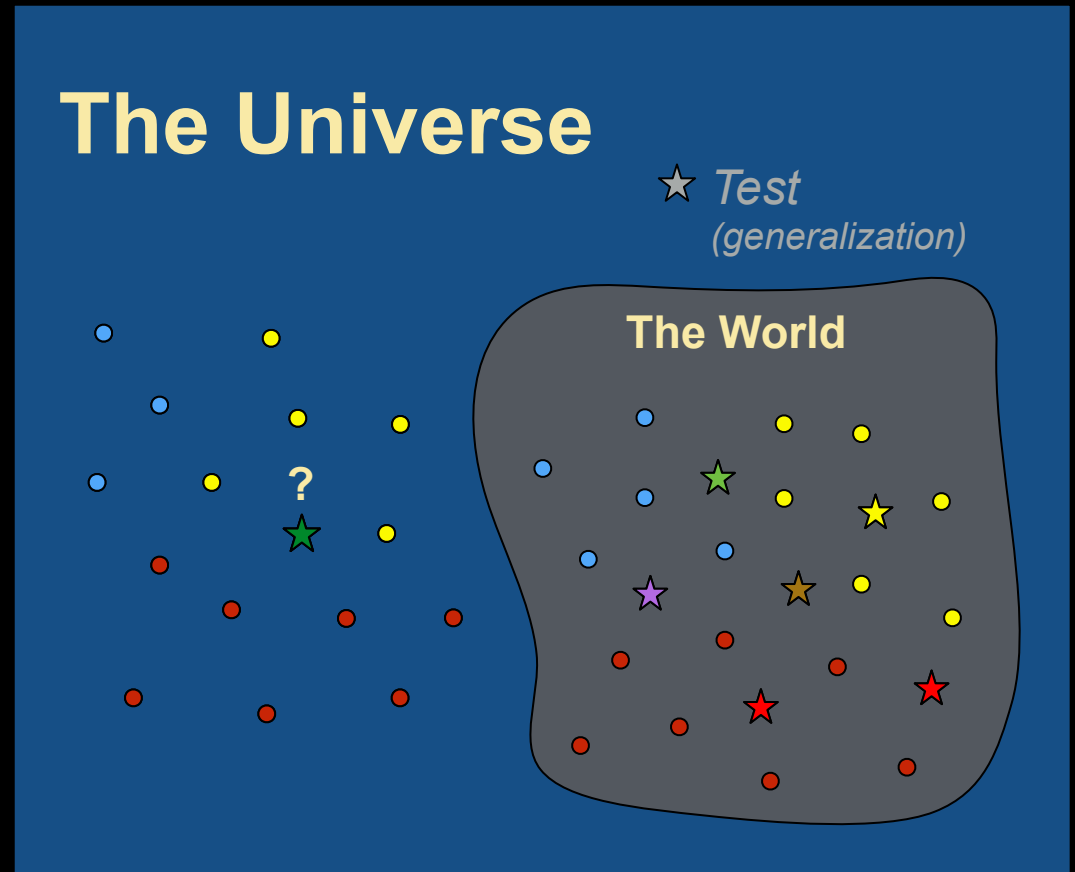
# Abstraction

*The search for (low dimensional) structure*

- Usually evaluated by capacity for generalization:

- **Interpolation**  
*(out of sample)*

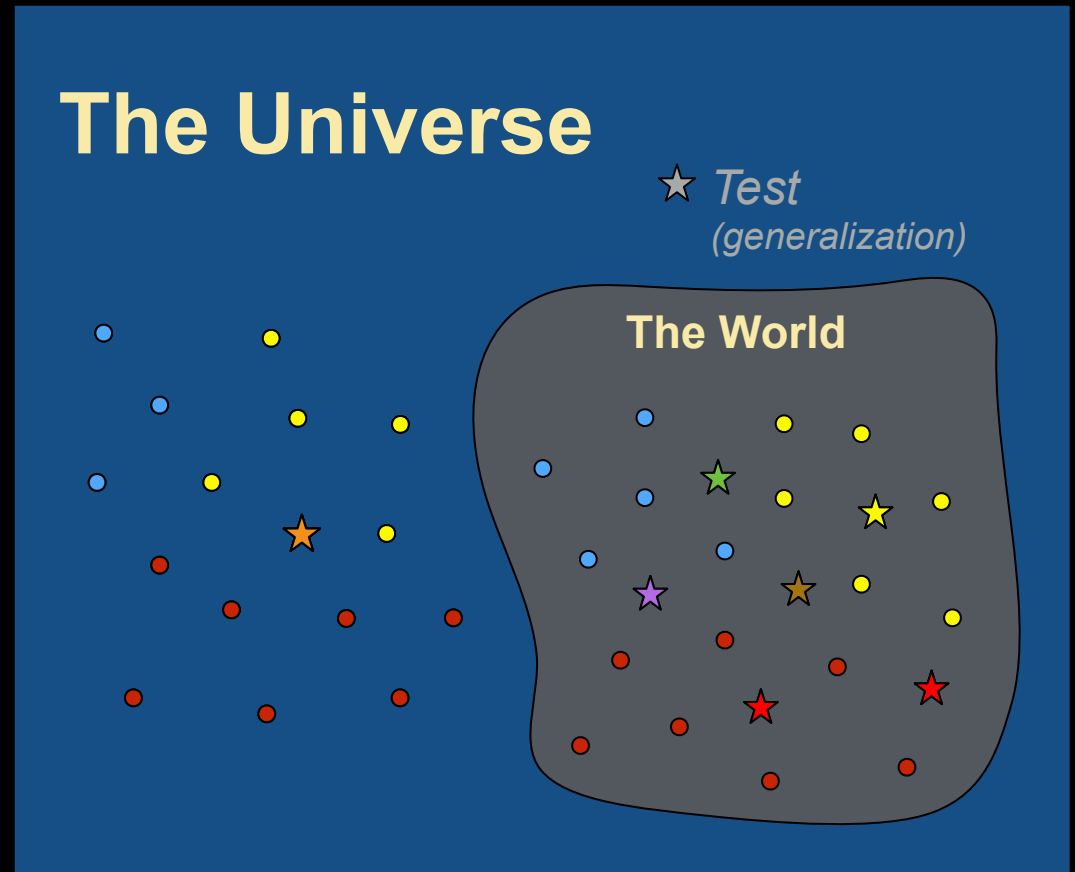
- **Extrapolation**  
*(out of distribution)*



# Abstraction

*The search for (low dimensional) structure*

- Usually evaluated by capacity for generalization:
  - Interpolation  
*(out of sample)*
  - Extrapolation  
*(out of distribution)*

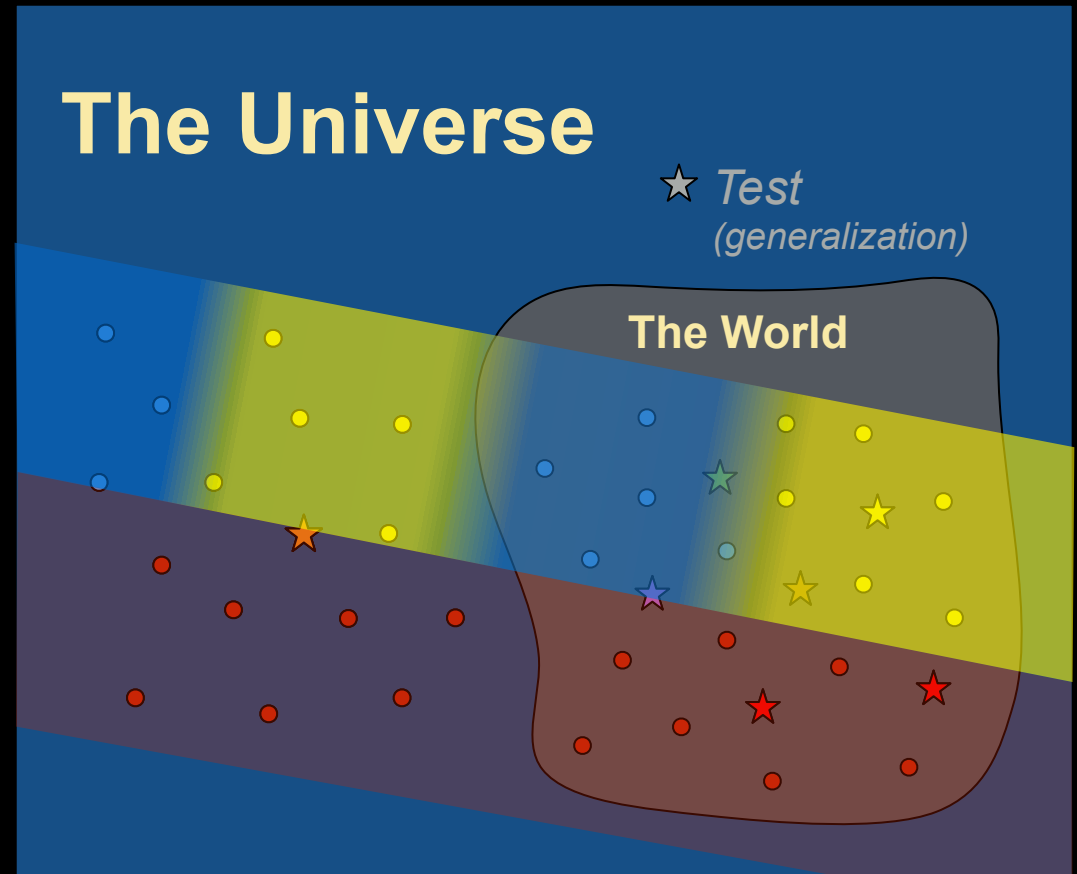




# Abstraction

*The search for (low dimensional) structure*

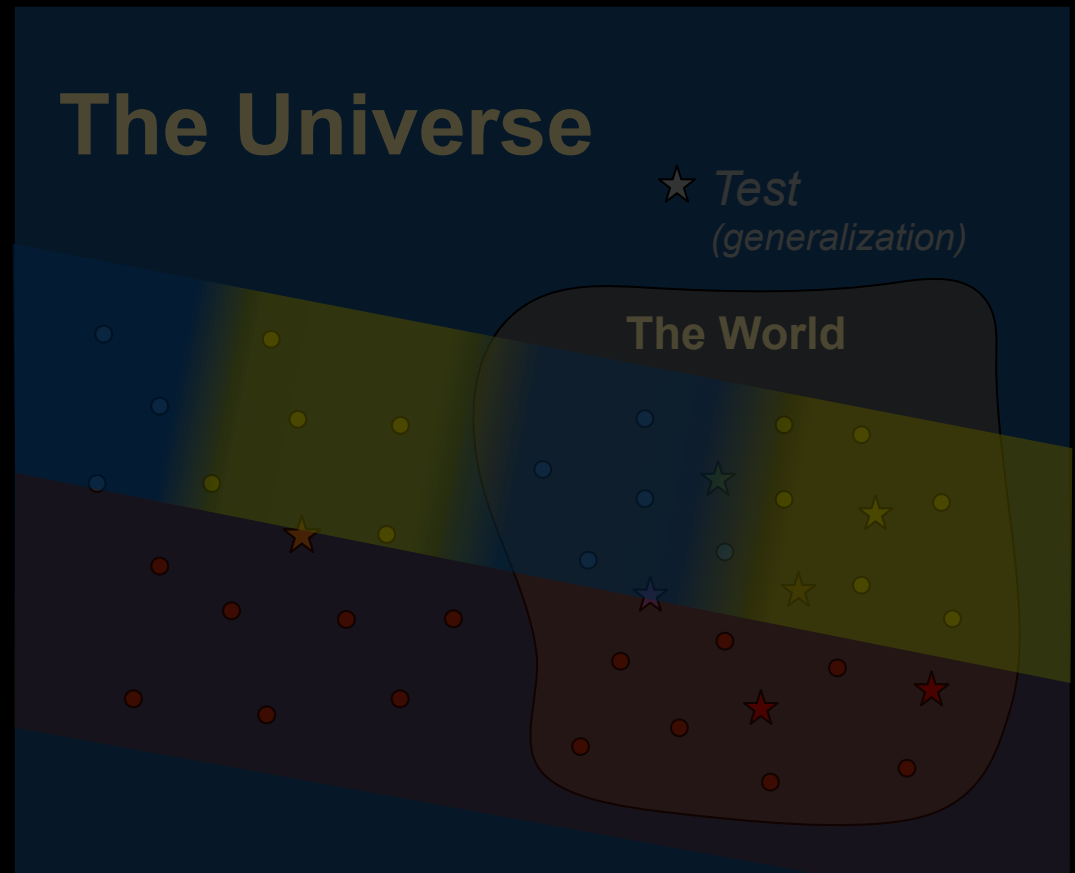
- Usually evaluated by capacity for generalization:
  - Interpolation  
*(out of sample)*
  - Extrapolation  
*(out of distribution)*
  - recognition of structure



# Abstraction

*The search for (low dimensional) structure*

- Usually evaluated by capacity for generalization:
  - **Interpolation**  
*(out of sample)*
  - **Extrapolation**  
*(out of distribution)*
    - recognition of structure
- “Cognitive” example...



# **Abstraction**

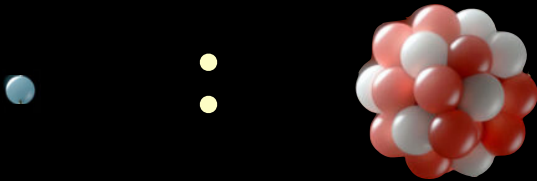
*The search for (low dimensional) structure*

***Analogy***

# Abstraction

*The search for (low dimensional) structure*

## *Analogy*

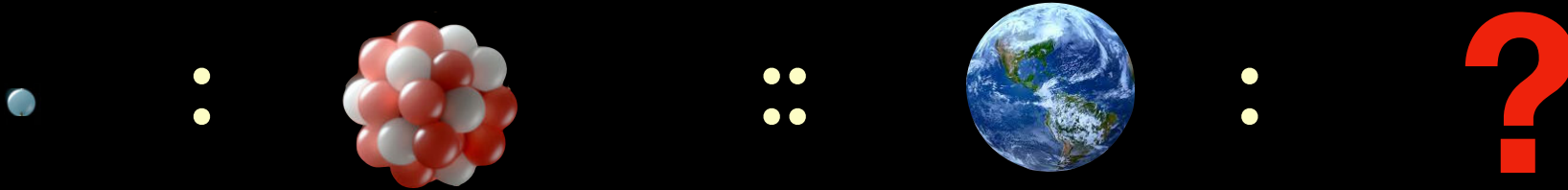


electron *is to* nucleus

# Abstraction

*The search for (low dimensional) structure*

## Analogy

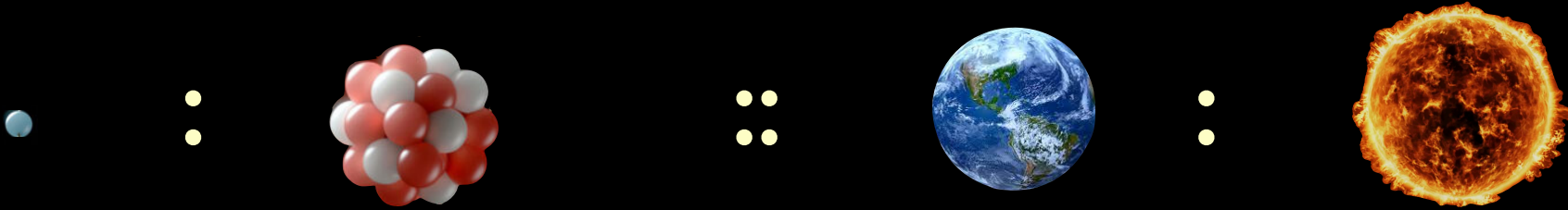


electron *is to* nucleus *as* earth *is to* ?

# Abstraction

*The search for (low dimensional) structure*

## *Analogy*



electron *is to* nucleus

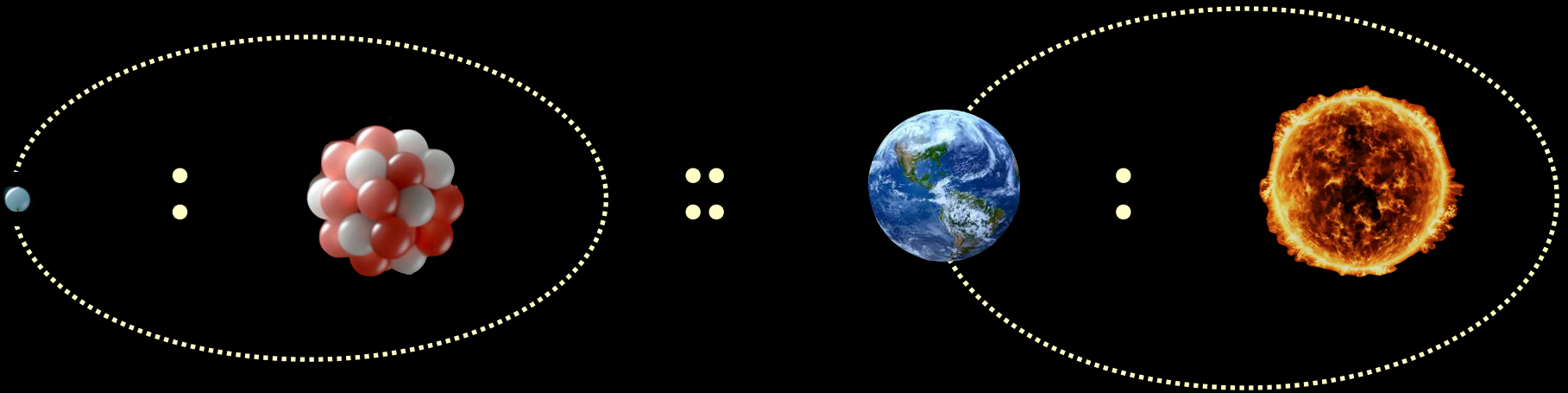
*as*

earth *is to* *Sun*

# Abstraction

*The search for (low dimensional) structure*

## Analogy



electron *is to* nucleus

*as*

earth *is to* sun

# Abstraction

*The search for (low dimensional) structure*

## Analogy



:



:



:



electron *is to* nucleus

*as*

earth

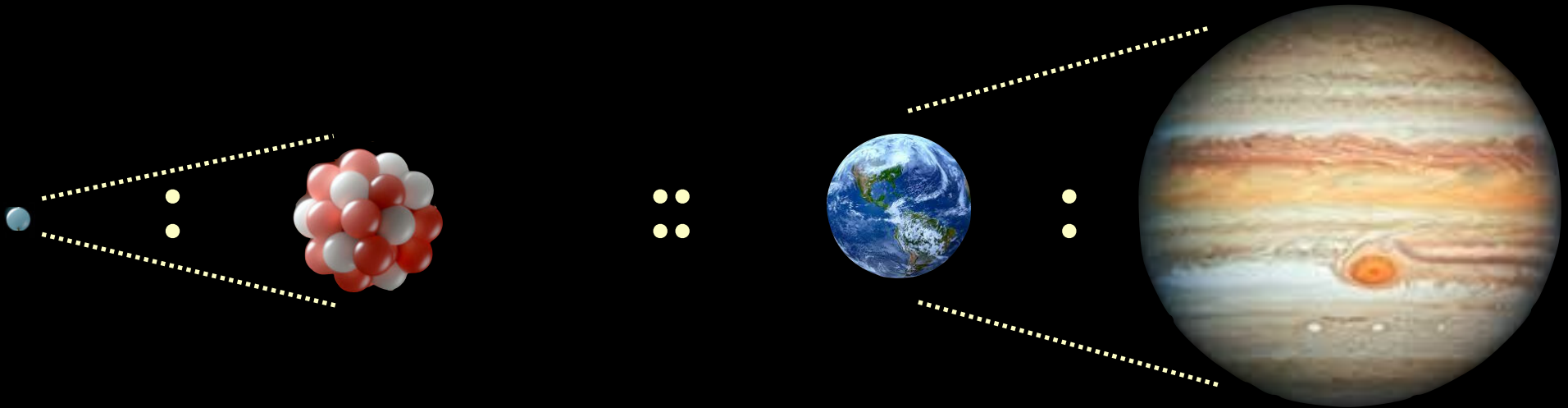
*is to Jupiter*



# Abstraction

*The search for (low dimensional) structure*

## Analogy



electron *is to* nucleus      *as*      earth *is to* *Jupiter*

# **Abstraction**

*The search for (low dimensional) structure*

***Analogy***

# Abstraction

*The search for (low dimensional) structure*

*Analogy*



electron : nucleus

# Abstraction

*The search for (low dimensional) structure*

## *Analogy*



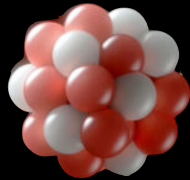
electron : nucleus :: earth : ?



# Abstraction

*The search for (low dimensional) structure*

## *Analogy*

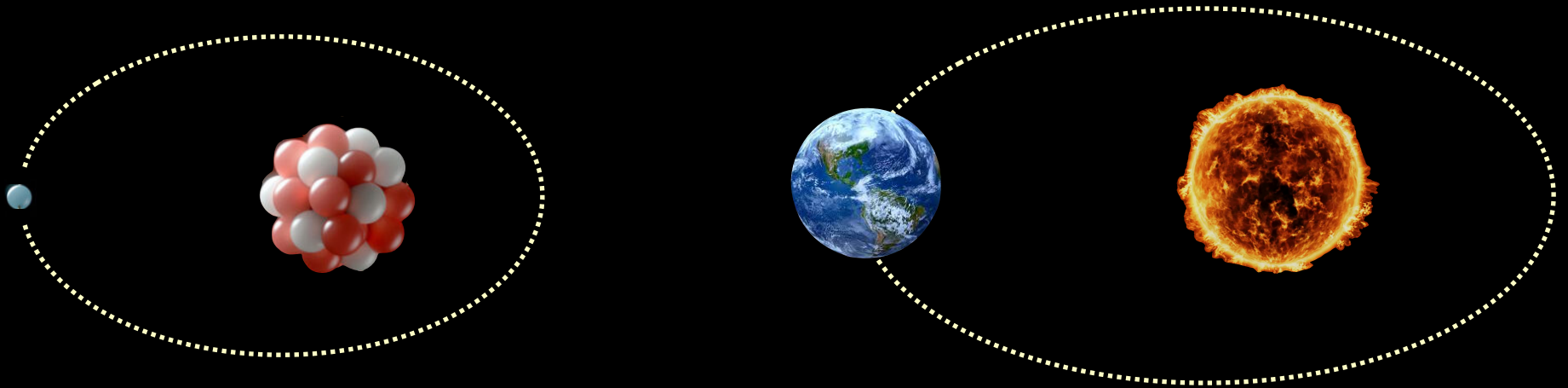


electron : nucleus :: earth : *sun*

# Abstraction

*The search for (low dimensional) structure*

## Analogy

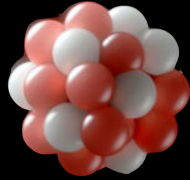


electron : nucleus :: earth : *sun*

# Abstraction

*The search for (low dimensional) structure*

## *Analogy*

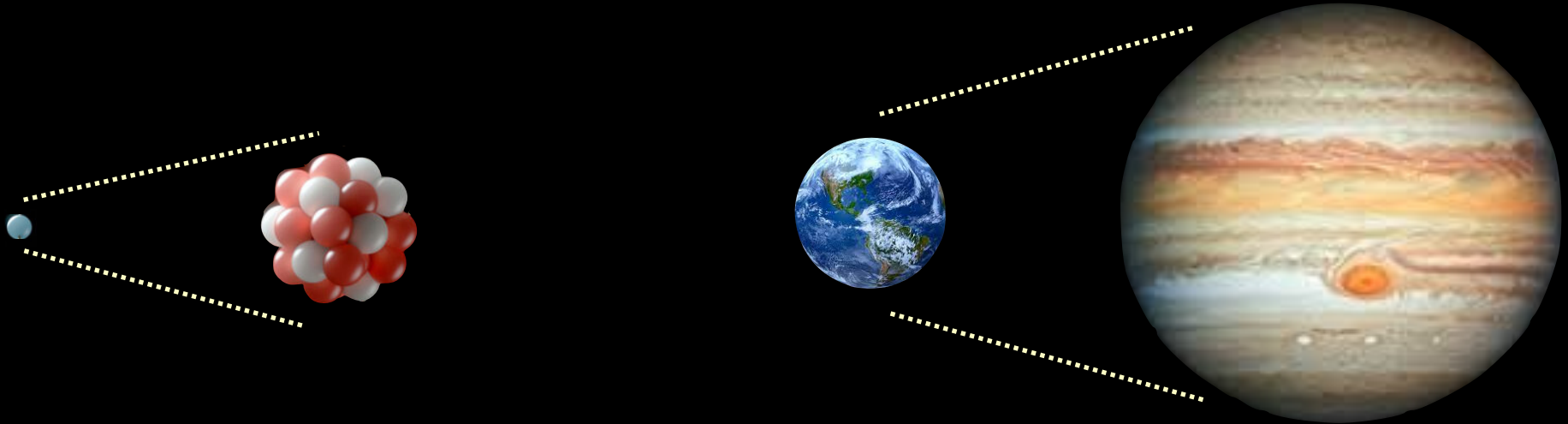


electron : nucleus :: earth : *Jupiter*

# Abstraction

*The search for (low dimensional) structure*

## Analogy



electron : nucleus :: earth : *Jupiter*



# Abstraction

*The search for (low dimensional) structure*

## Analogy

Structure

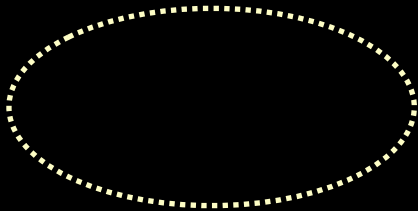
Relation

Function

.....

Linear

$$y = mx + b$$

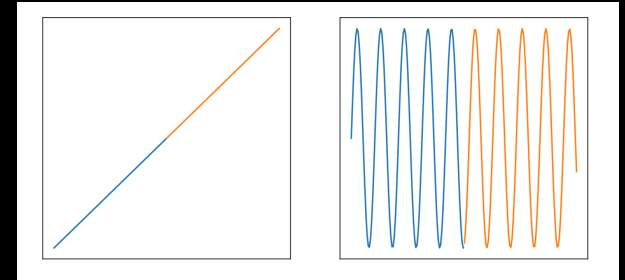


Orbital

$$a^2 - b^2 = c^2$$

# Abstraction

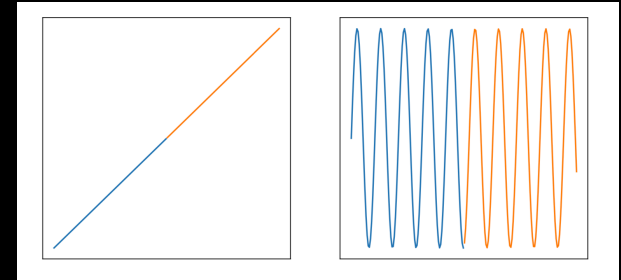
*The search for (low dimensional) structure*



# Abstraction

*The search for (low dimensional) structure*

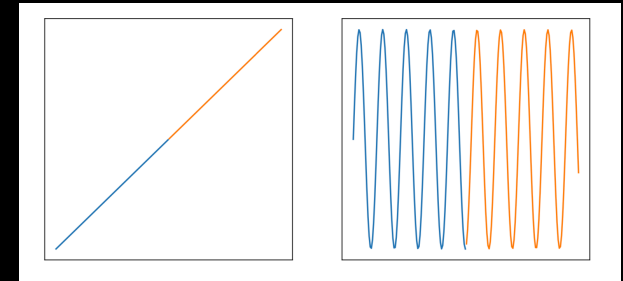
- Function learning and extrapolation



# Abstraction

*The search for (low dimensional) structure*

- **Function learning and extrapolation**
  - Neuro-symbolic approach:
    - Gaussian process models  
(Schultz et al., Cog Pay 2017)
      - must pre-specify basis functions

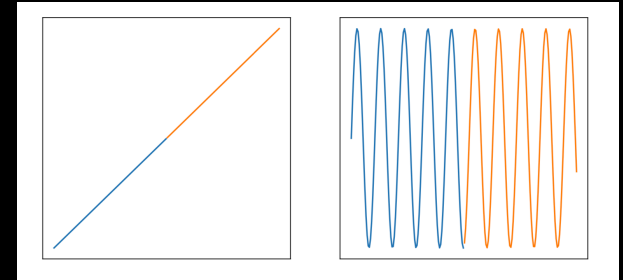


# Abstraction

## *The search for (low dimensional) structure*

- **Function learning and extrapolation**

- Neuro-symbolic approach:
  - Gaussian process models  
(Schultz et al., Cog Pay 2017)
    - must pre-specify basis functions
- Connectionist approach:
  - Autoregressive / Iterator (RNN) models  
(Segert & Cohen, TMLR 2022)



# Abstraction

## *The search for (low dimensional) structure*

- **Function learning and extrapolation**

- Neuro-symbolic approach:

- Gaussian process models

- (Schultz et al., Cog Pay 2017)

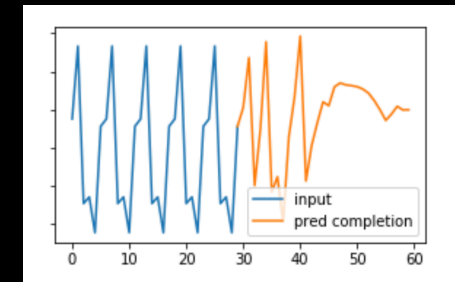
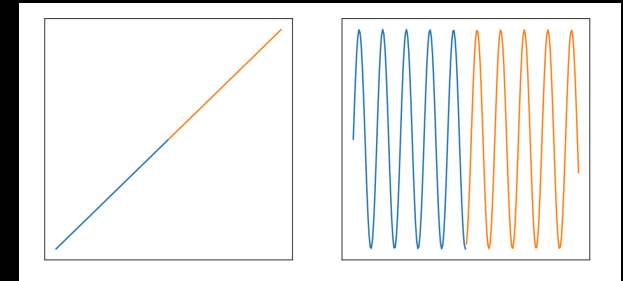
- must pre-specify basis functions

- Connectionist approach:

- Autoregressive / Iterator (RNN) models

- (Segert & Cohen, TMLR 2022)

- function approximation - accumulation of errors





# Abstraction

## *The search for (low dimensional) structure*

- **Function learning and extrapolation**

- Neuro-symbolic approach:

- Gaussian process models

- (Schultz et al., Cog Pay 2017)*

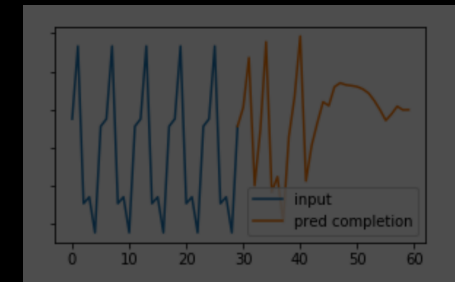
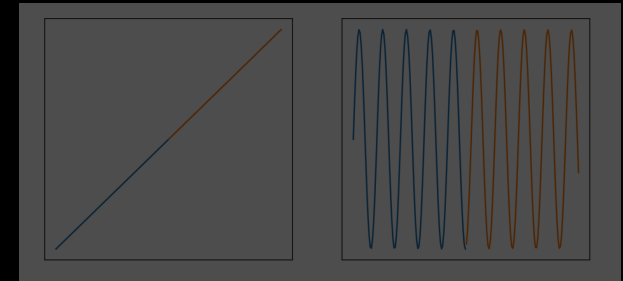
- must pre-specify basis functions

- Connectionist approach:

- Autoregressive / Iterator (RNN) models

- (Segert & Cohen, TMLR 2022)*

- function approximation - accumulation of errors



- **What we really want is the discovery of *symmetry* functions...**

# **Abstraction**

*The search for (low dimensional) structure*

# Abstraction

*The search for (low dimensional) structure*

- The Search **Symmetry**:

*“It is only slightly overstating the case to say that  
physics is the study of symmetry”*

*(Anderson, Science 1978)*

# Abstraction

*The search for (low dimensional) structure*

- The Search **Symmetry**:

*“It is only slightly overstating the case to say that  
physics is the study of symmetry”*

*(Anderson, Science 1978)*

- It is doing so only a bit more to say so for cognitive science...

# Abstraction

*The search for (low dimensional) structure*

- The Search **Symmetry**:

*“It is only slightly overstating the case to say that  
physics is the study of symmetry”*

*(Anderson, Science 1978)*

- It is doing so only a bit more to say so for cognitive science...

- function learning, analogies, stereotypes, schemas  $\Rightarrow$

*Invariance / equivariance over transformations*

# Abstraction

*The search for (low dimensional) structure*

- The Search **Symmetry**:

*“It is only slightly overstating the case to say that  
physics is the study of symmetry”*

*(Anderson, Science 1978)*

- It is doing so only a bit more to say so for cognitive science...
- function learning, analogies, stereotypes, schemas  $\Rightarrow$   
*Invariance / equivariance over transformations*
- Genuine extrapolation requires the discovery of symmetry



# Abstraction

*The search for (low dimensional) structure*

- The Search **Symmetry**:

*“It is only slightly overstating the case to say that  
physics is the study of symmetry”*

*(Anderson, Science 1978)*

- It is doing so only a bit more to say so for cognitive science...
- function learning, analogies, stereotypes, schemas  $\Rightarrow$   
*Invariance / equivariance over transformations*
- Genuine extrapolation requires the discovery of symmetry
- What inductive bias in learning will promote such discovery?

# Relations as Symmetry

# Relations as Symmetry

- Symmetries  $\approx$

“relations” that obtain over the fundamental domain of a function,  
that apply indefinitely (periodically) beyond it

# Relations as Symmetry

- Symmetries  $\approx$

“relations” that obtain over the fundamental domain of a function,  
that apply indefinitely (periodically) beyond it

– *properly formalized in Group Theory (more on that if there is time)*

# Relations as Symmetry

- Symmetries  $\approx$

“relations” that obtain over the fundamental domain of a function,  
that apply indefinitely (periodically) beyond it

– *properly formalized in Group Theory (more on that if there is time)*

key point: inducing the *learning of relations* →  
promote the *discovery of symmetries*

# Relations as Symmetry

- Symmetries  $\approx$

“relations” that obtain over the fundamental domain of a function,  
that apply indefinitely (periodically) beyond it

– *properly formalized in Group Theory (more on that if there is time)*

key point: inducing the *learning of relations*  $\rightarrow$   
promote the *discovery of symmetries*

- **Similarities in data are a place to start...** (*“correlations are all you need”*)

	$X_1$	$X_2$	$X_3$
$Y_1$			
$Y_2$			
$Y_3$			


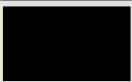
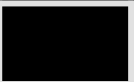






Identity map

	$X_1$	$X_2$	$X_3$
$Y_1$			
$Y_2$			
$Y_3$			

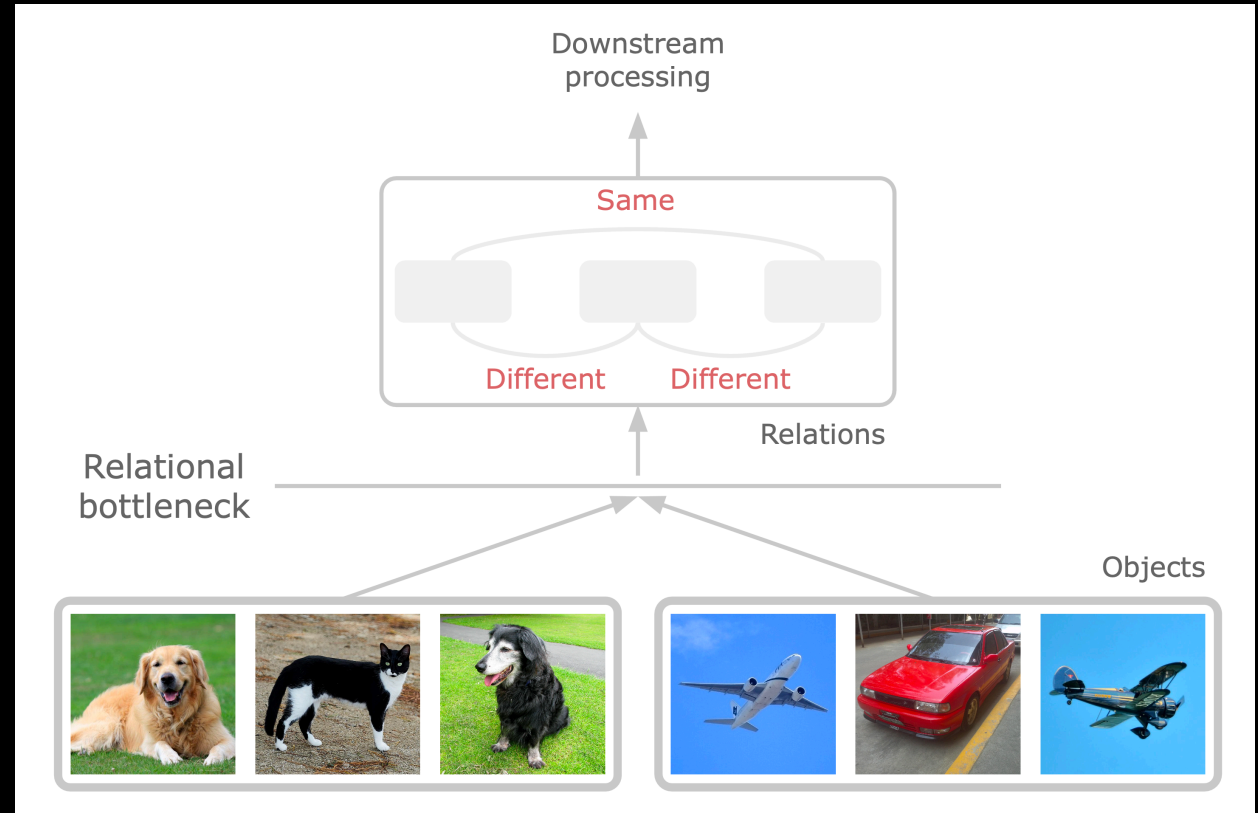
Sequence map (for ABA)



# Relational Bottleneck

	$X_1$	$X_2$	$X_3$
$Y_1$			
$Y_2$			
$Y_3$			

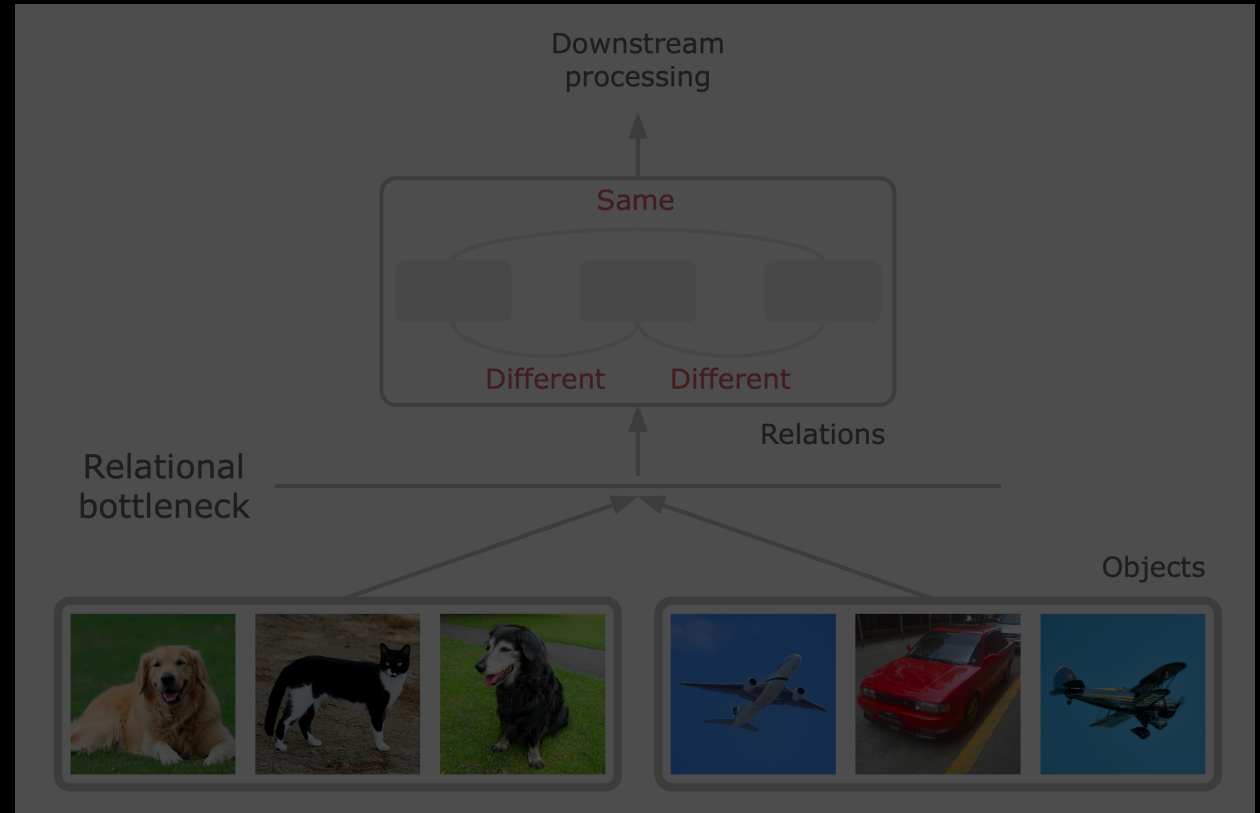
Identity map



# Relational Bottleneck

	$X_1$	$X_2$	$X_3$
$Y_1$			
$Y_2$			
$Y_3$			

Identity map



How do we build this in a neural network?

# Relational Bottleneck

# Relational Bottleneck

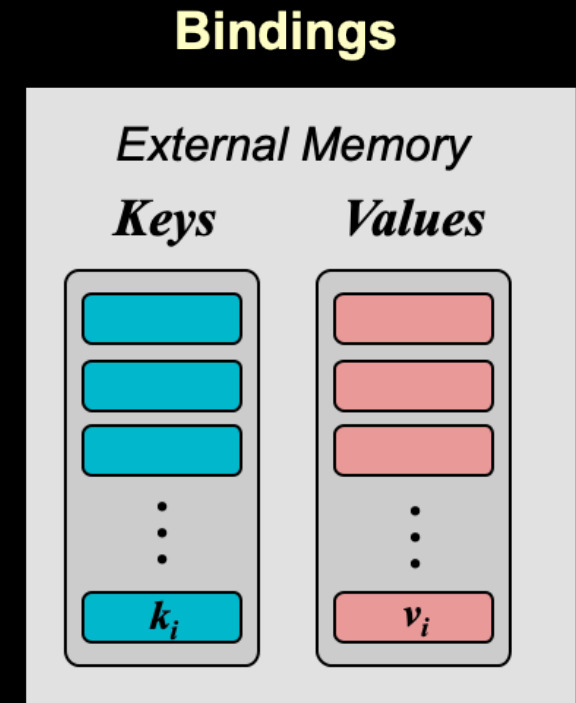
- Build on use of “external memory” in deep learning networks  
(*Neural Turing Machine, Graves et al., 2014*)

# Relational Bottleneck

- Build on use of “external memory” in deep learning networks  
(*Neural Turing Machine, Graves et al., 2014*)

- External memory:

Form of “dictionary” (*key-value pairs*):





# Relational Bottleneck

- Build on use of “**external memory**” in deep learning networks  
(*Neural Turing Machine, Graves et al., 2014*)

- **External memory:**

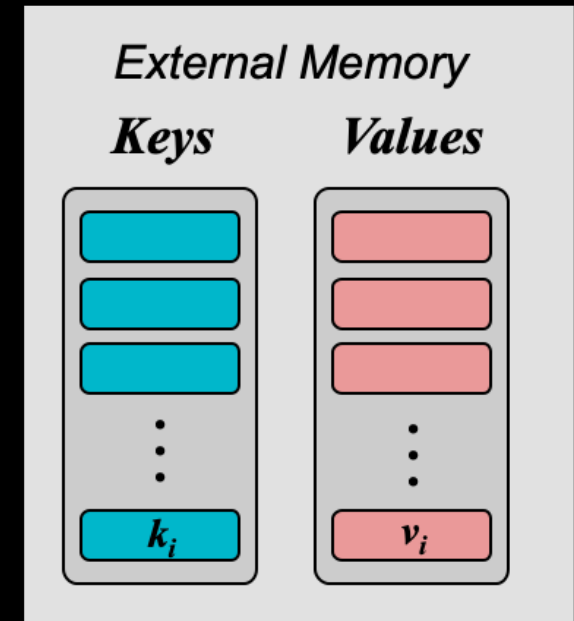
Form of “dictionary” (*key-value pairs*):

**Rapid binding** of arbitrary pieces of information

+

**Similarity-based retrieval** (e.g., *use of inner products*)

## Bindings



# Relational Bottleneck

- Build on use of “**external memory**” in deep learning networks  
(*Neural Turing Machine, Graves et al., 2014*)

- **External memory:**

Form of “dictionary” (*key-value pairs*):

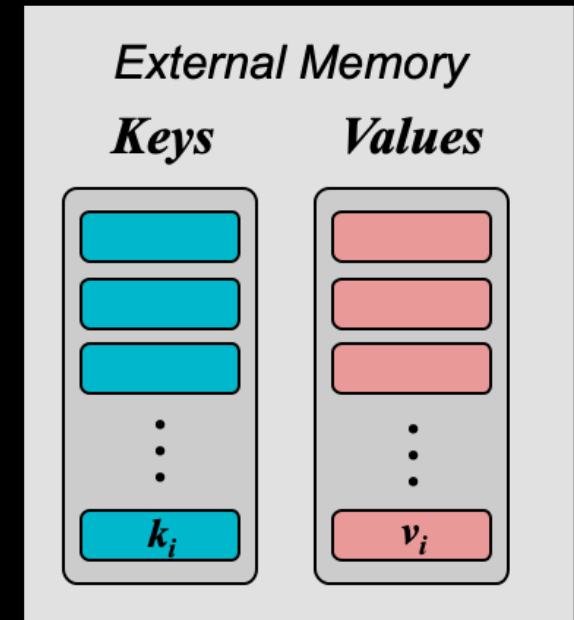
**Rapid binding** of arbitrary pieces of information

+

**Similarity-based retrieval** (*e.g., use of inner products*)

⇒ implements **episodic memory** function of  
medial temporal cortex (*e.g., hippocampus*)  
(*Complementary Memory Systems, McClelland et al., 1995*)

## Bindings



# Relational Bottleneck

- Build on use of “**external memory**” in deep learning networks

(*Neural Turing Machine, Graves et al., 2014*)

- **External memory:**

Form of “dictionary” (key-value pairs):

**Rapid binding** of arbitrary pieces of information

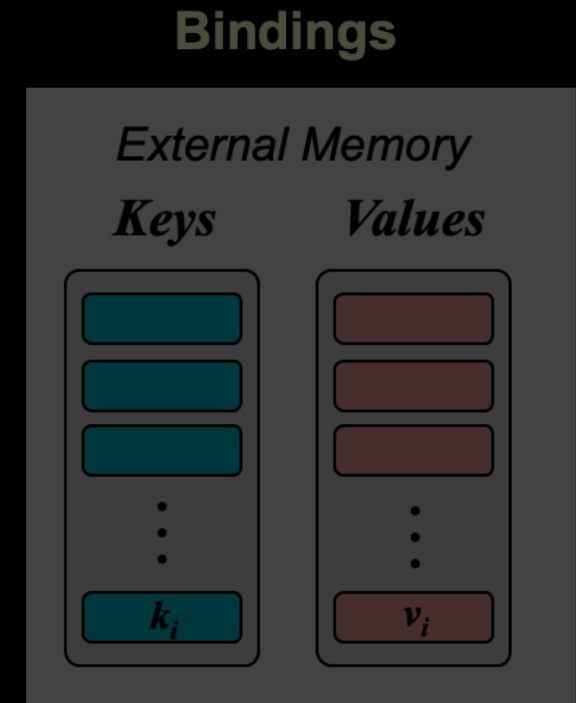
+

**Similarity-based retrieval** (e.g., use of inner products)

⇒ implements **episodic memory** function of  
medial temporal cortex (e.g., hippocampus)

(*Complementary Memory Systems, McClelland et al., 1995*)

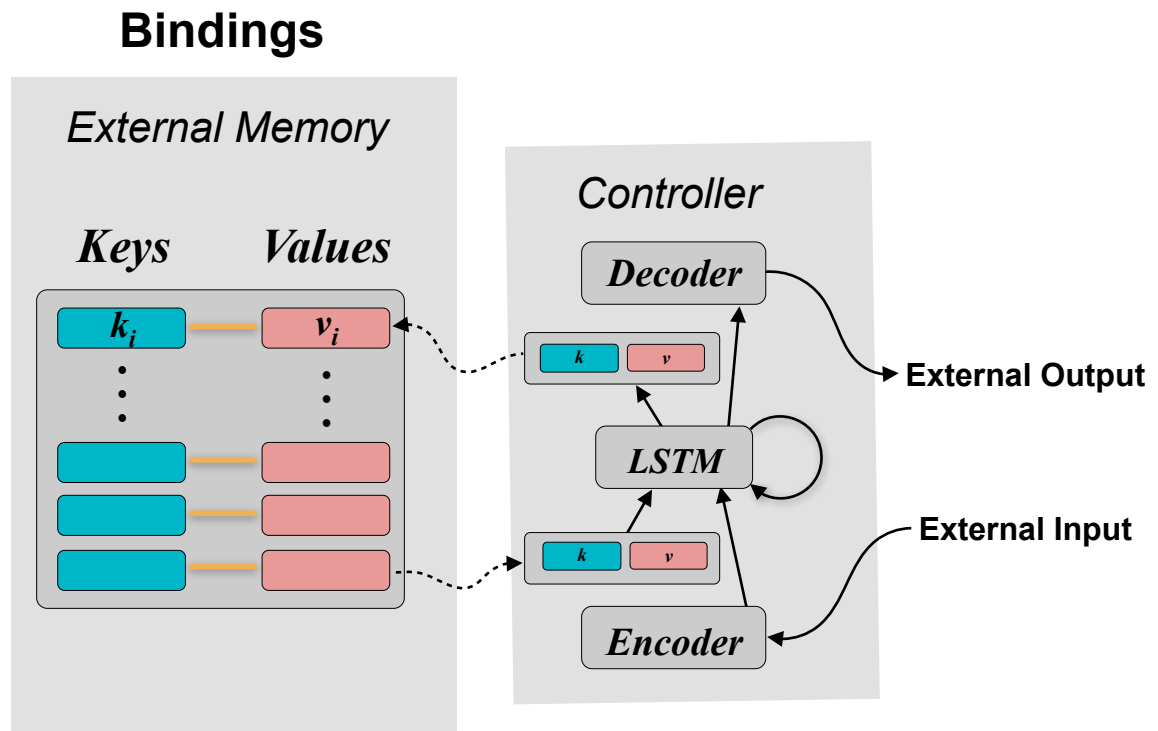
- **Example...**



# Neural Network with External Memory

## Neural Turing Machine (NTN)

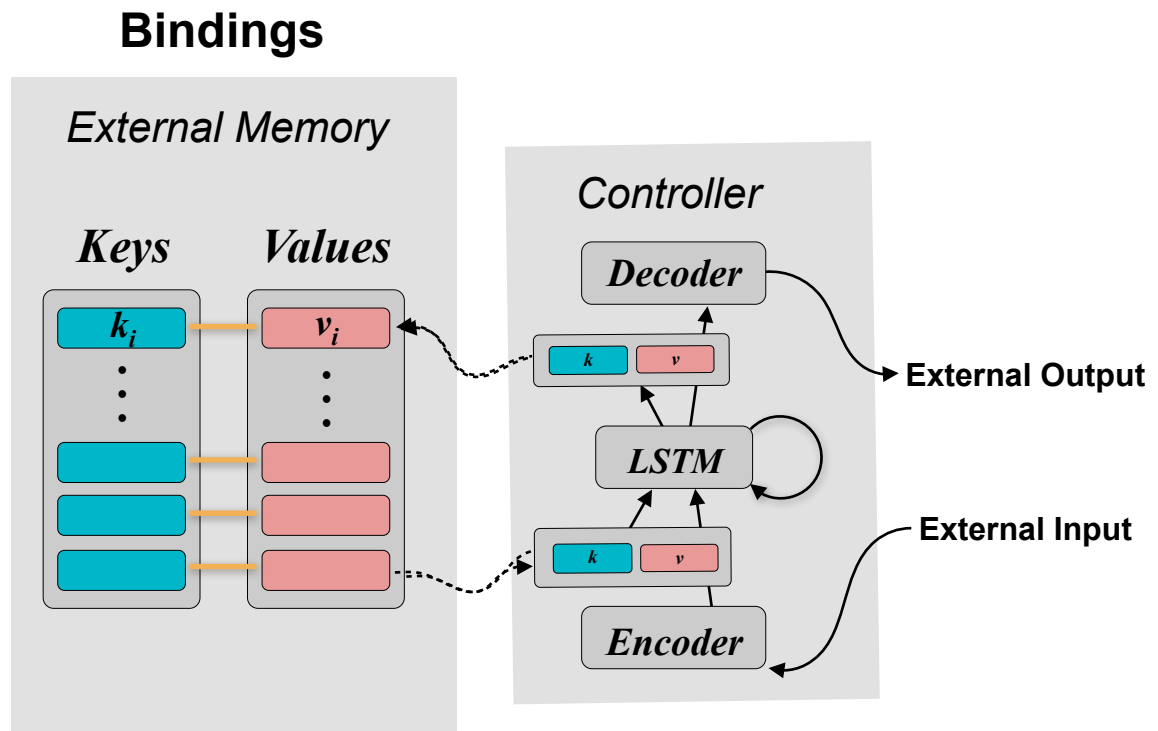
(Graves et al., 2014)



# Neural Network with External Memory

## Emergent Symbols Through Binding Network (ESBN)

(Webb et al., ICLR 2021)

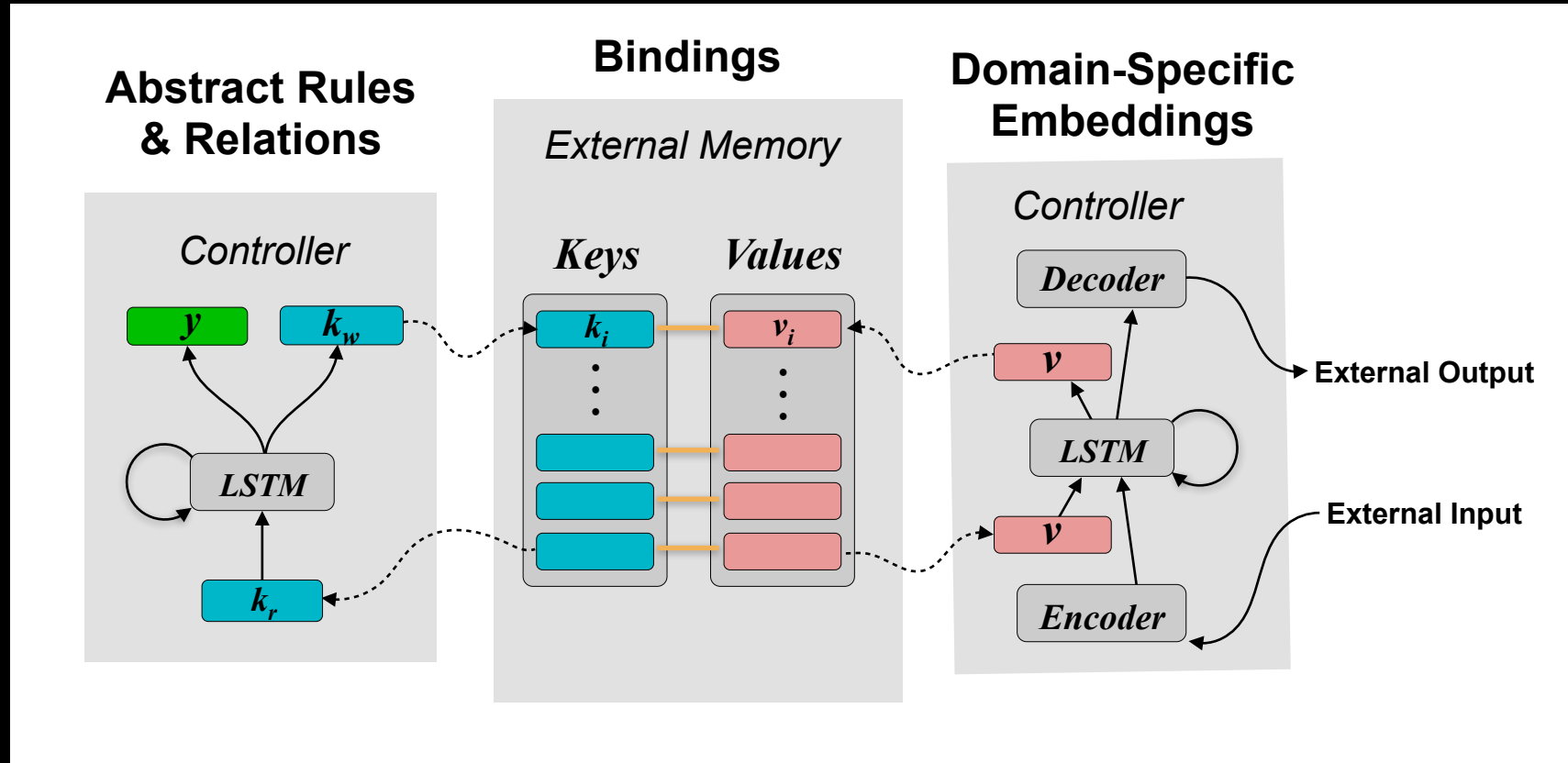




# Neural Network with External Memory

## Emergent Symbols Through Binding Network (ESBN)

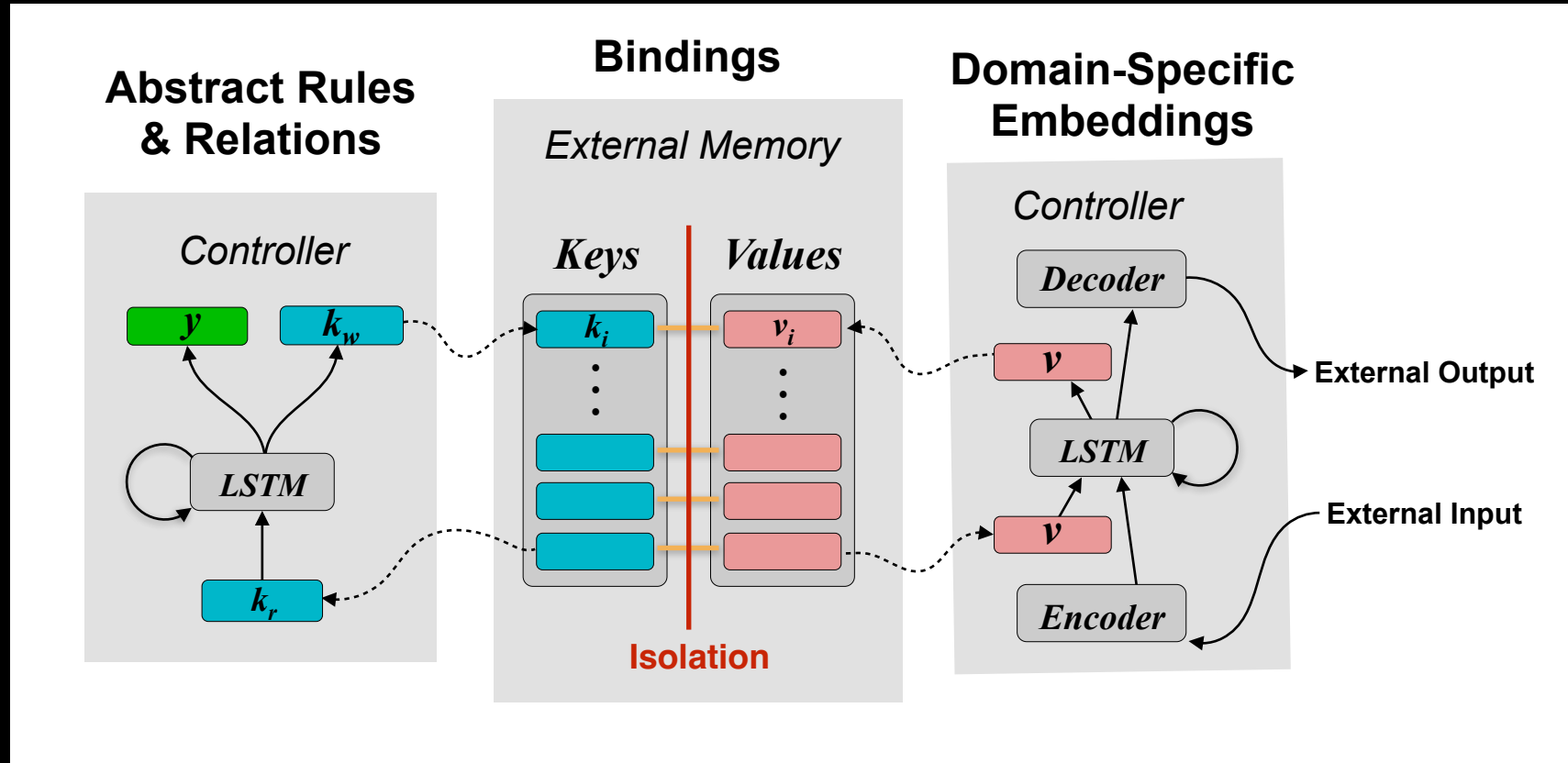
(Webb et al., ICLR 2021)



# Neural Network with External Memory

## Emergent Symbols Through Binding Network (ESBN)

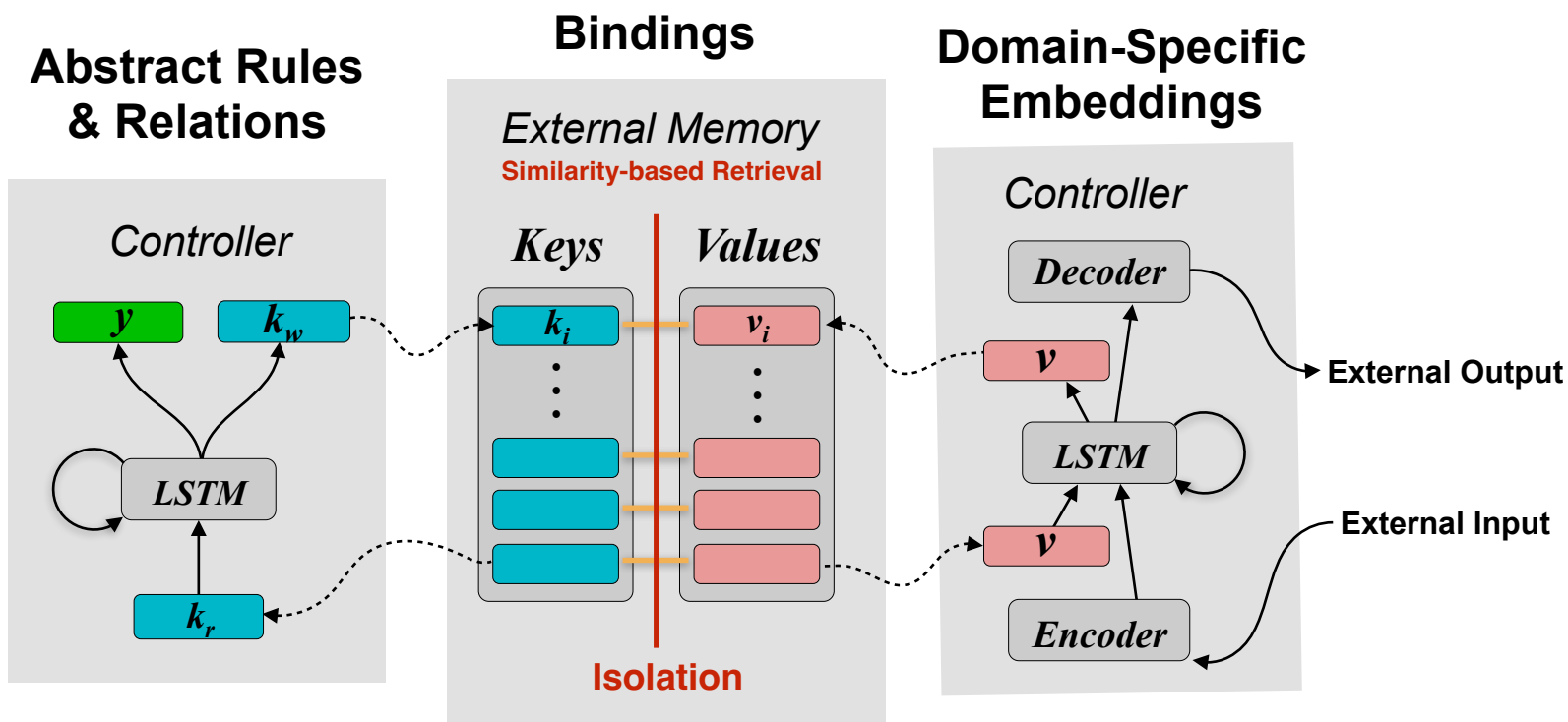
(Webb et al., ICLR 2021)



# Relational Bottleneck

## Emergent Symbols Through Binding Network (ESBN)

(Webb et al., ICLR 2021)



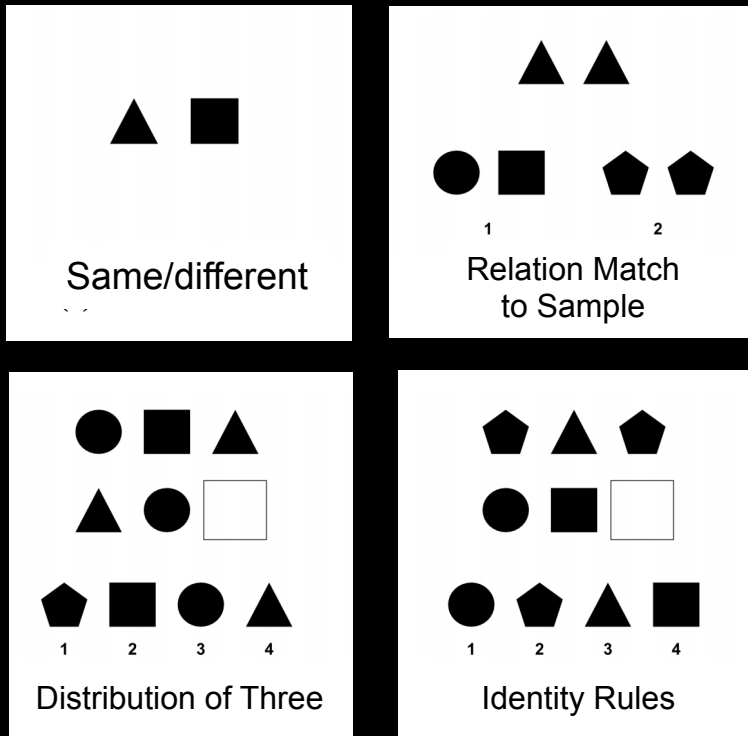
Isolation + similarity-based retrieval  $\Rightarrow$   
“Relational bottleneck”

# ESBN: Training

(Webb et al., 2021)

## Tasks

### Simple relations



*from Ravens Progressive Matrices*

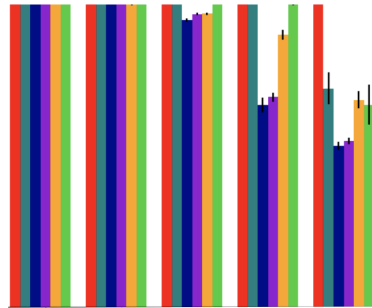


# ESBN: Results

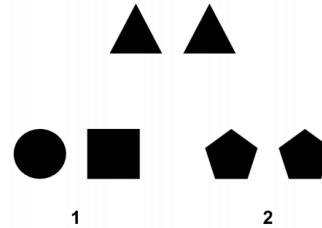
(Webb et al., 2021)



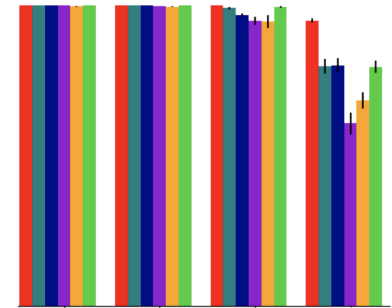
Same/different



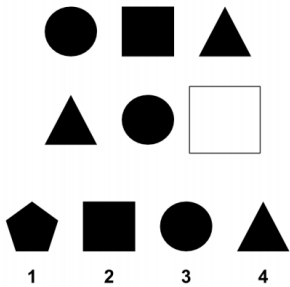
Same/different



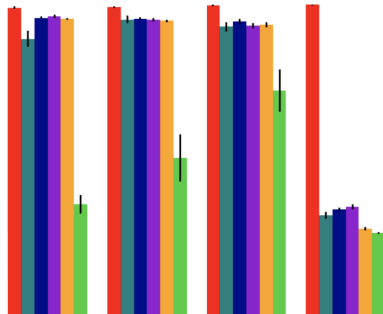
Relation Match  
to Sample



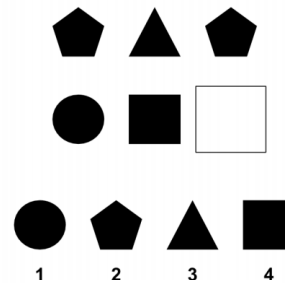
Relation Match to Sample



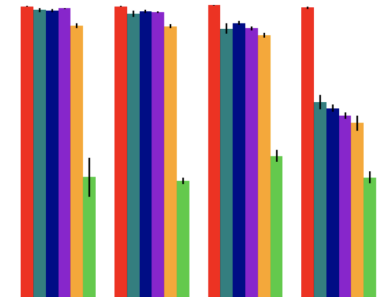
Distribution of Three



Distribution of Three



Identity Rules

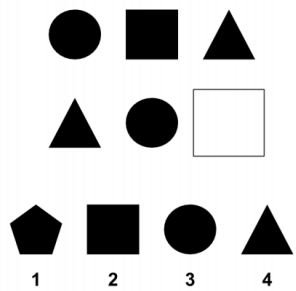


Identity Rules

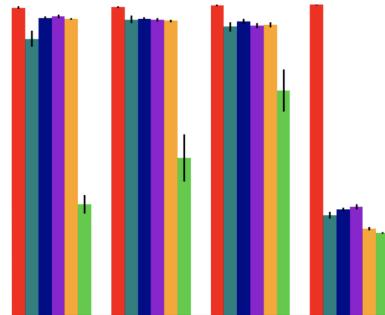


# ESBN: Results

*(Webb et al., 2021)*



Distribution of Three

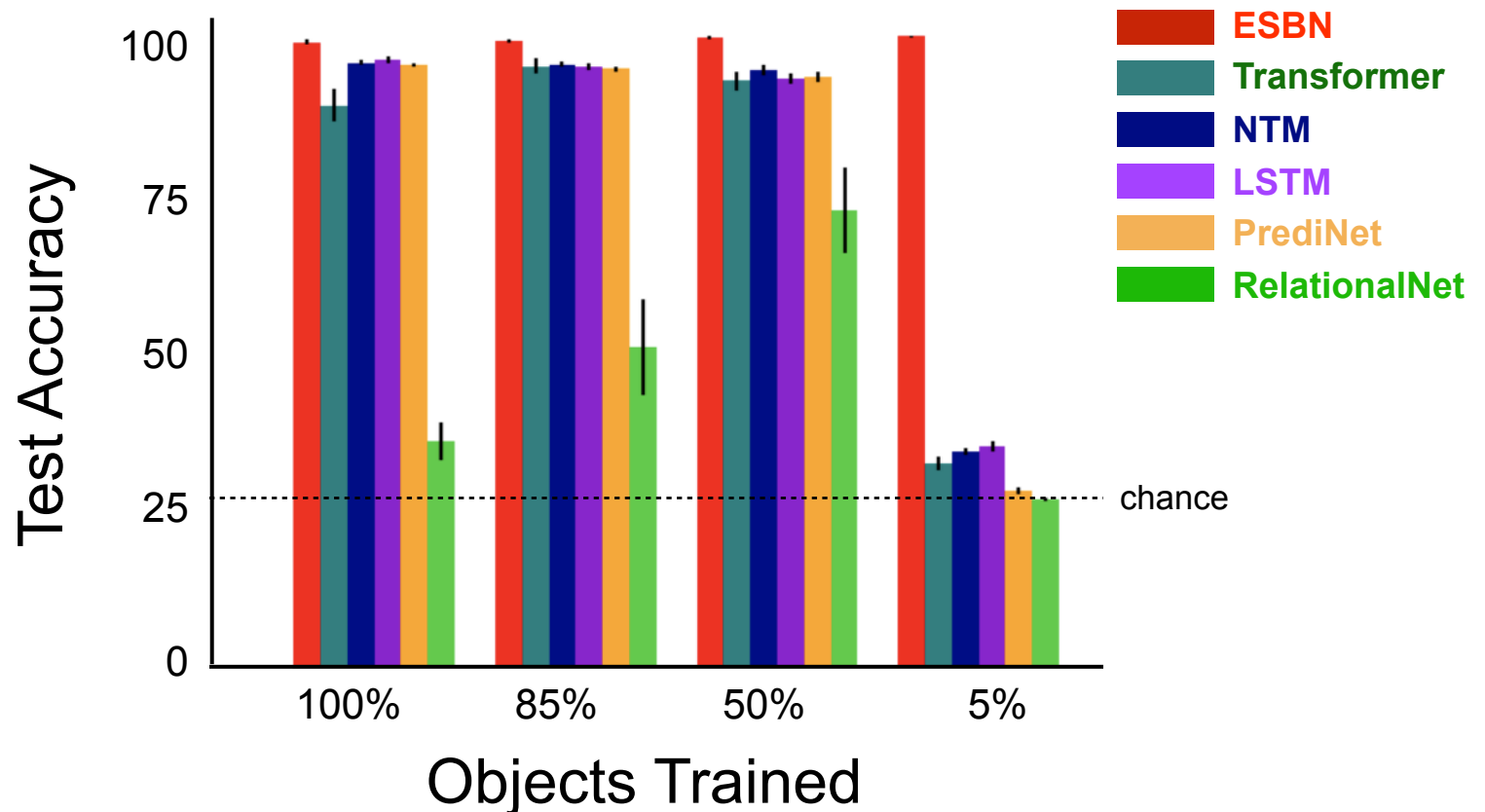
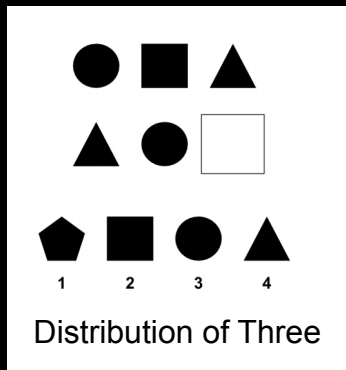


Distribution of Three

# ESBN: Results

(Webb et al., 2021)

## Extrapolation Performance

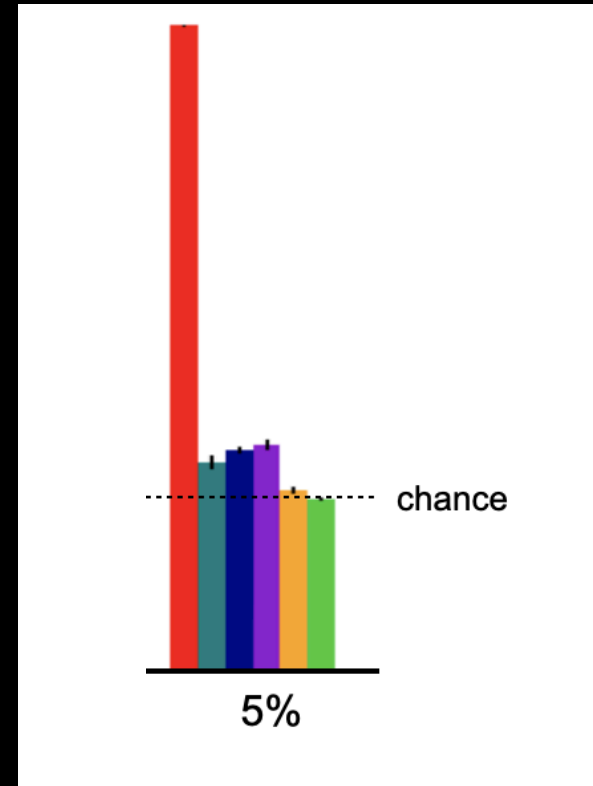


# ESBN: Results

(Webb et al., 2021)

## Extrapolation Performance

- Trained on *fewest number of items* needed to *exemplify the rule*

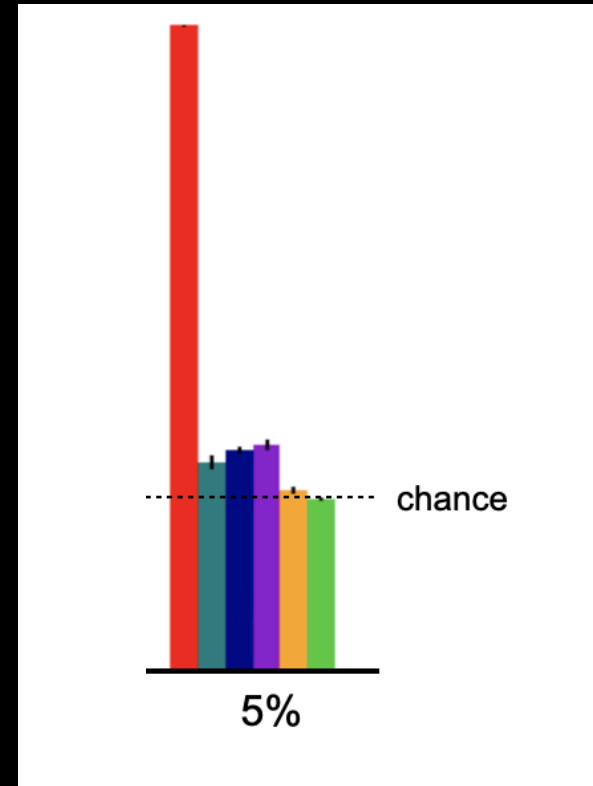


# ESBN: Results

(Webb et al., 2021)

## Extrapolation Performance

- Can *extrapolate* use of rule to *any set of items* it can encode

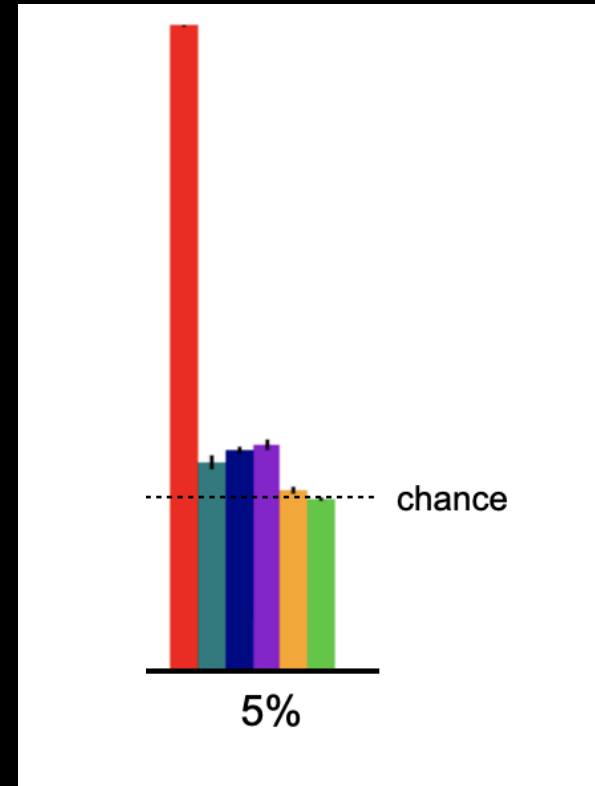


# ESBN: Results

(Webb et al., 2021)

## Extrapolation Performance

- *Learns* single set of keys (*roles*)  
used for any set of items (*fillers*)

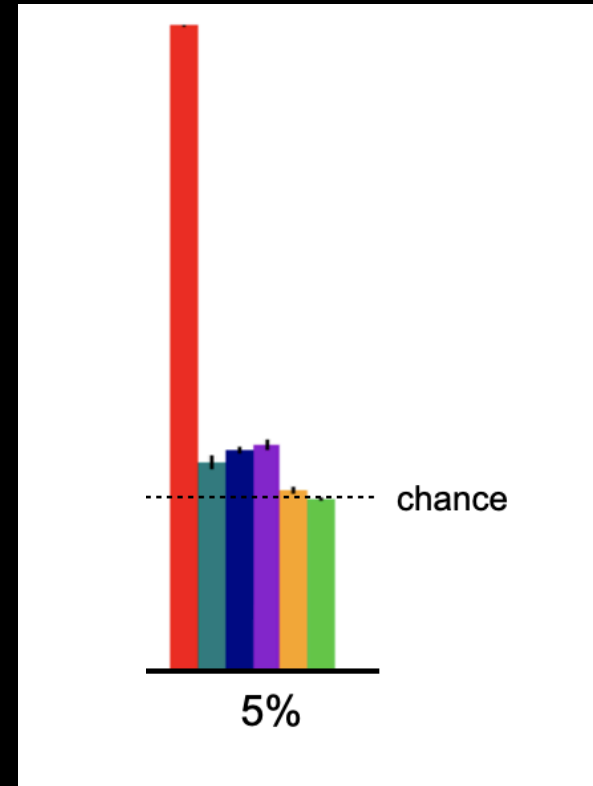


# ESBN: Results

(Webb et al., 2021)

## Extrapolation Performance

⇒ Genuinely *symbolic processing*  
using external (episodic memory)  
for *variable binding*

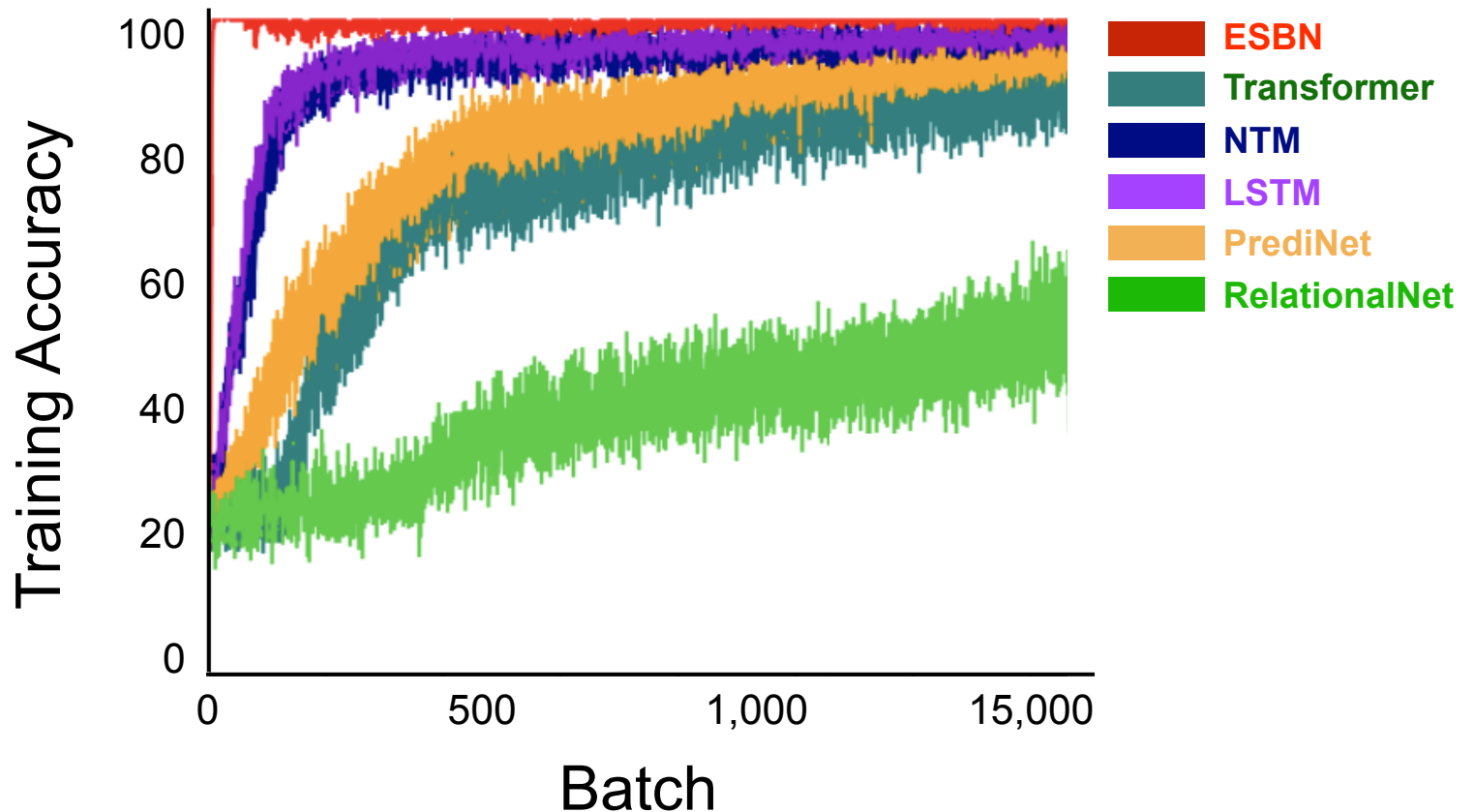
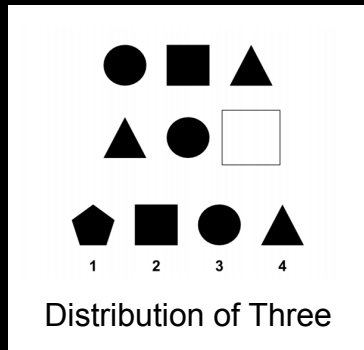




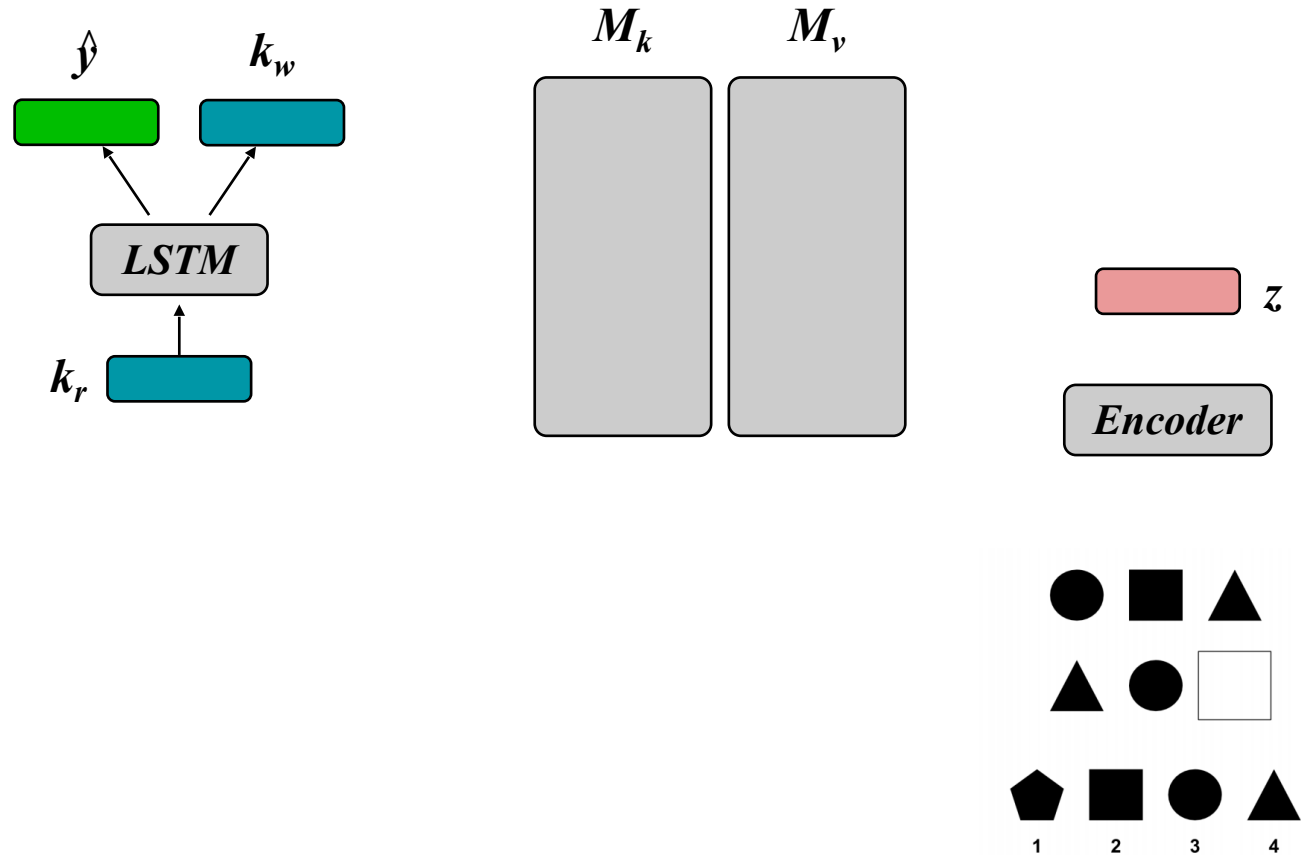
# ESBN: Results

(Webb et al., 2021)

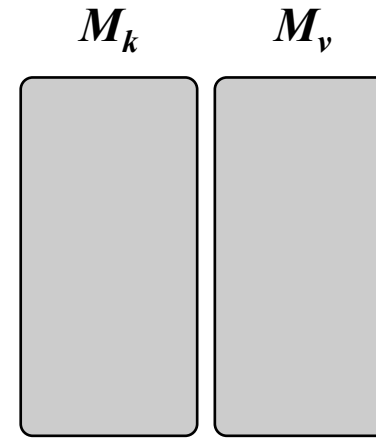
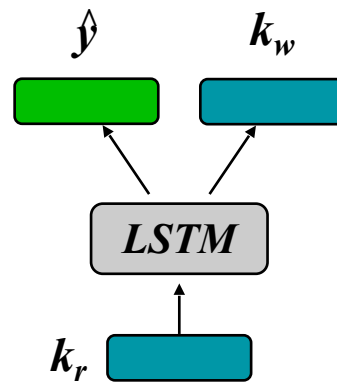
## Sample Efficiency



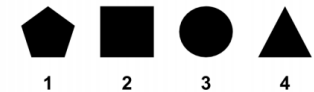
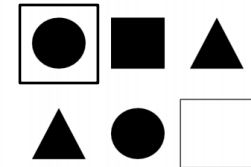
# Example Trial: Distribution of Three



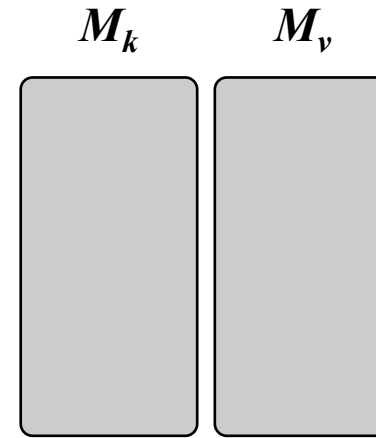
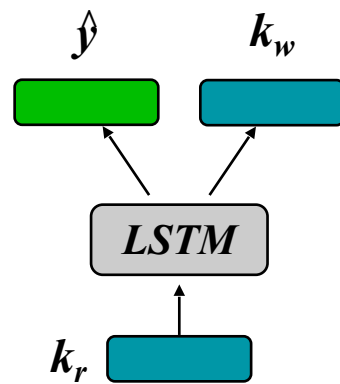
# Example Trial: Distribution of Three



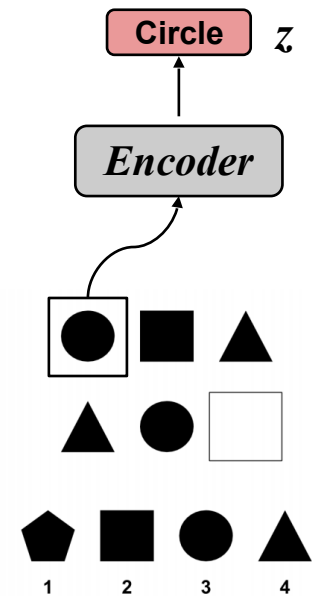
Encode



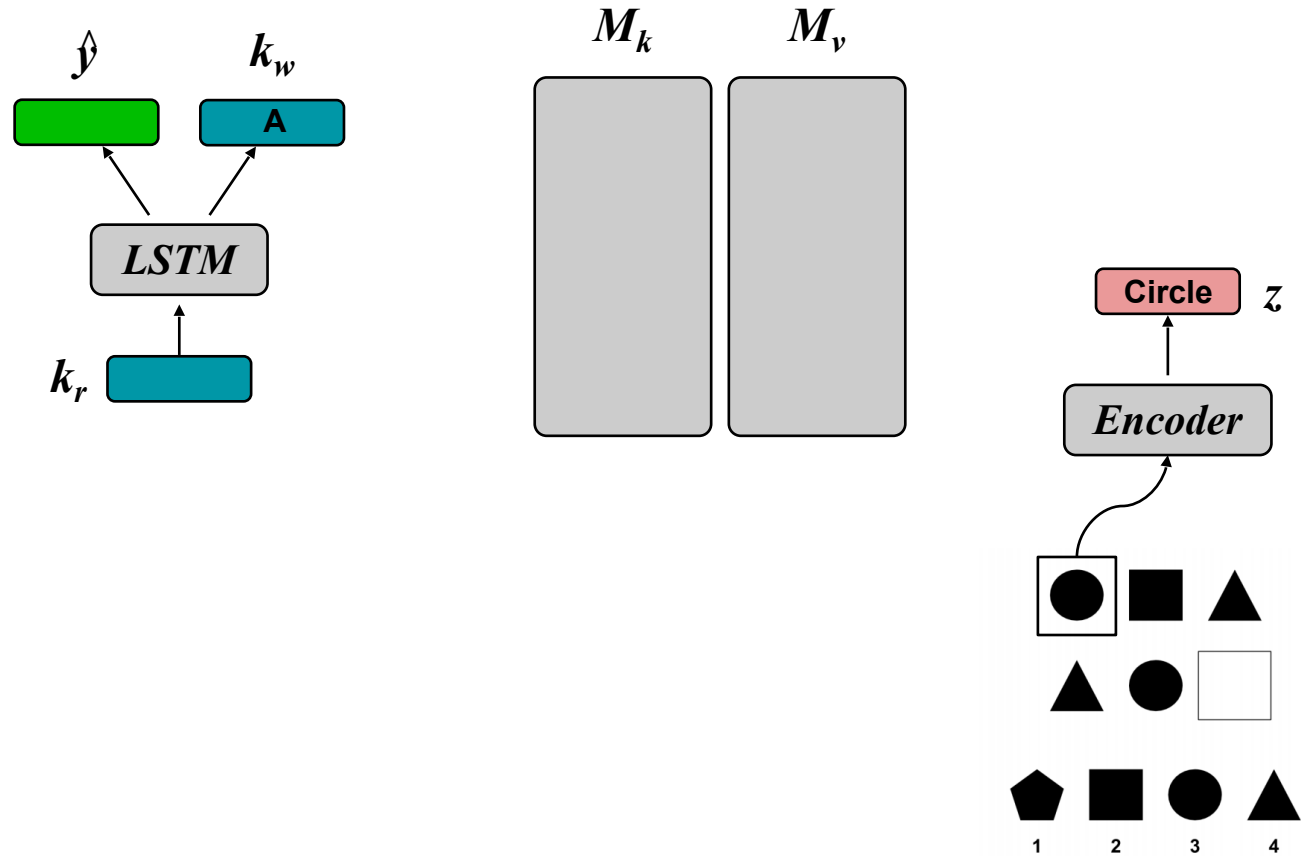
# Example Trial: Distribution of Three



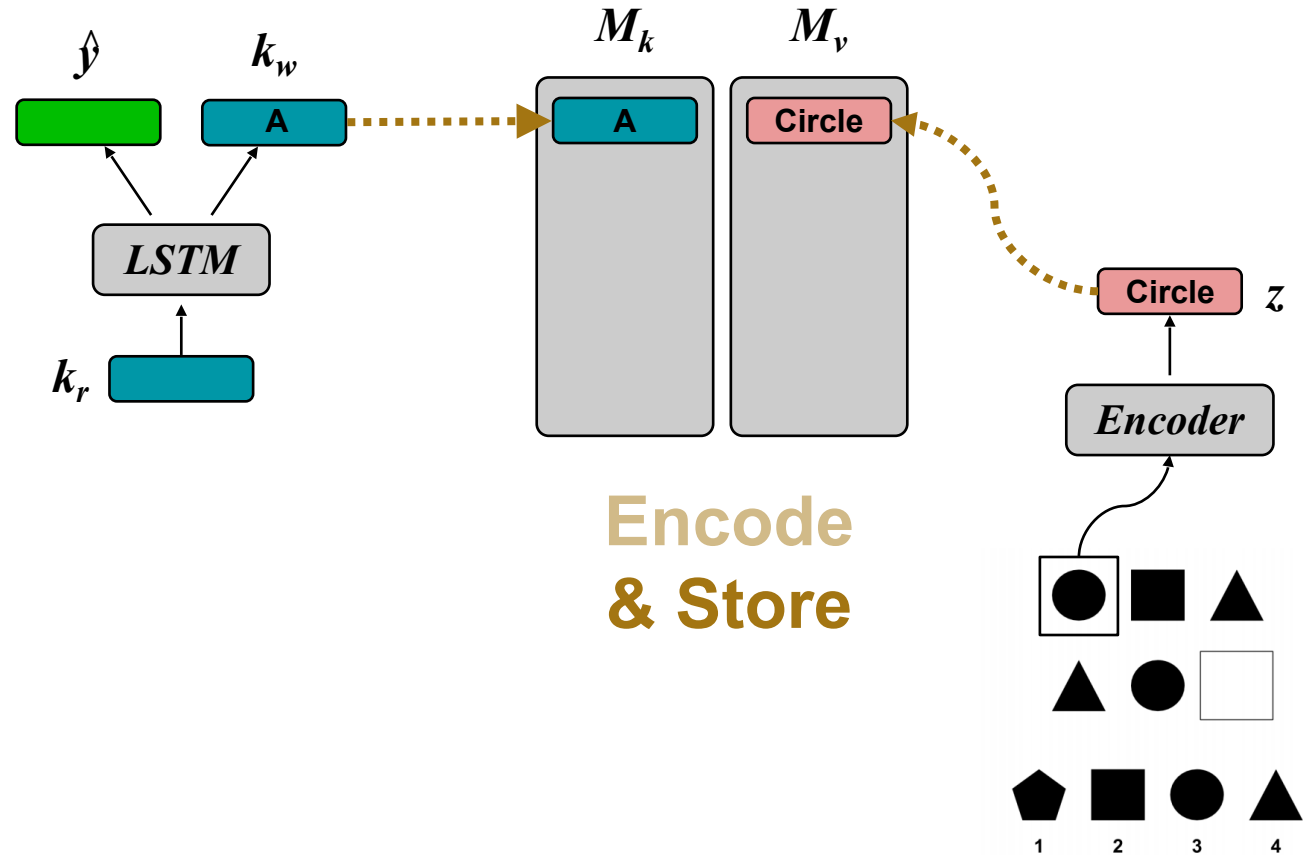
Encode



# Example Trial: Distribution of Three

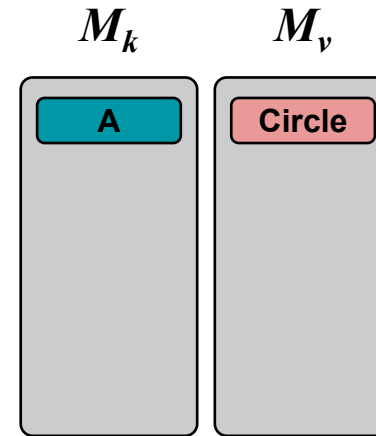
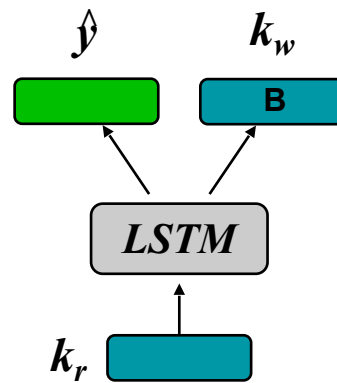


# Example Trial: Distribution of Three

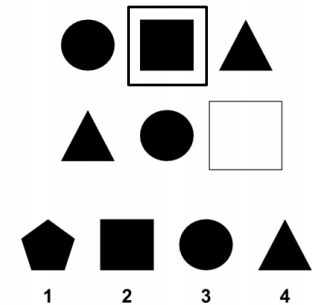
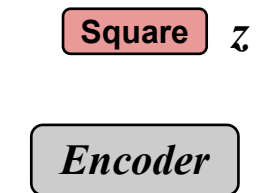




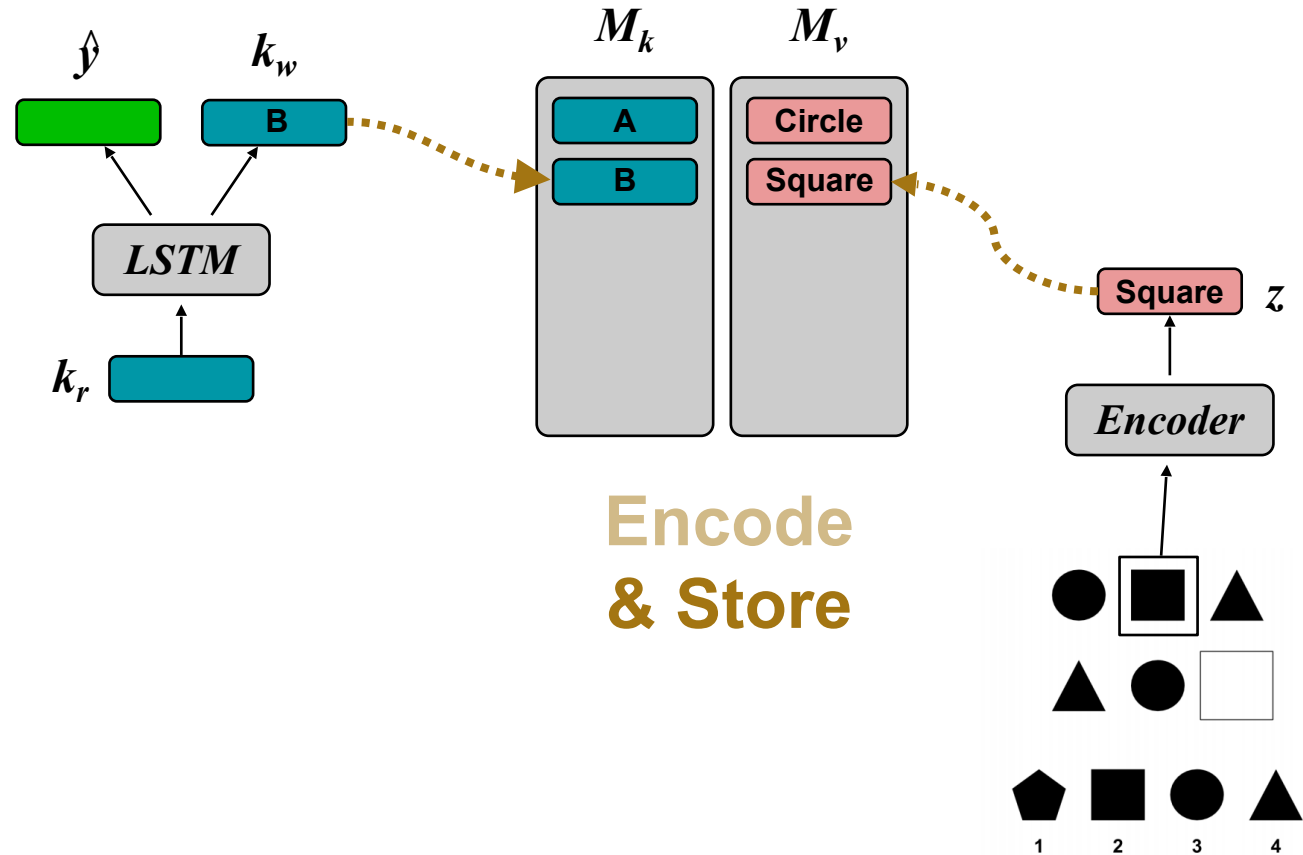
# Example Trial: Distribution of Three



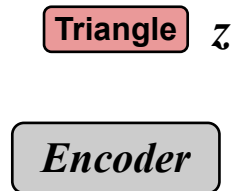
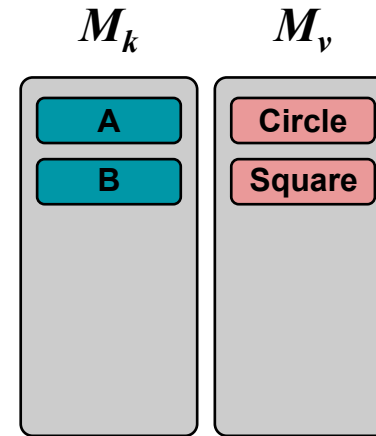
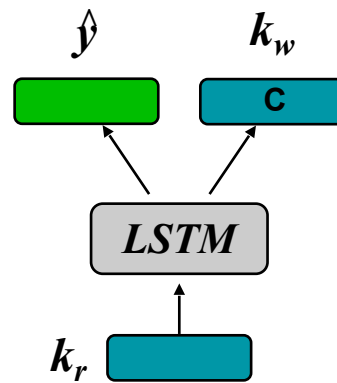
Encode  
& Store



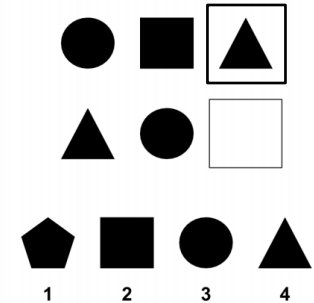
# Example Trial: Distribution of Three



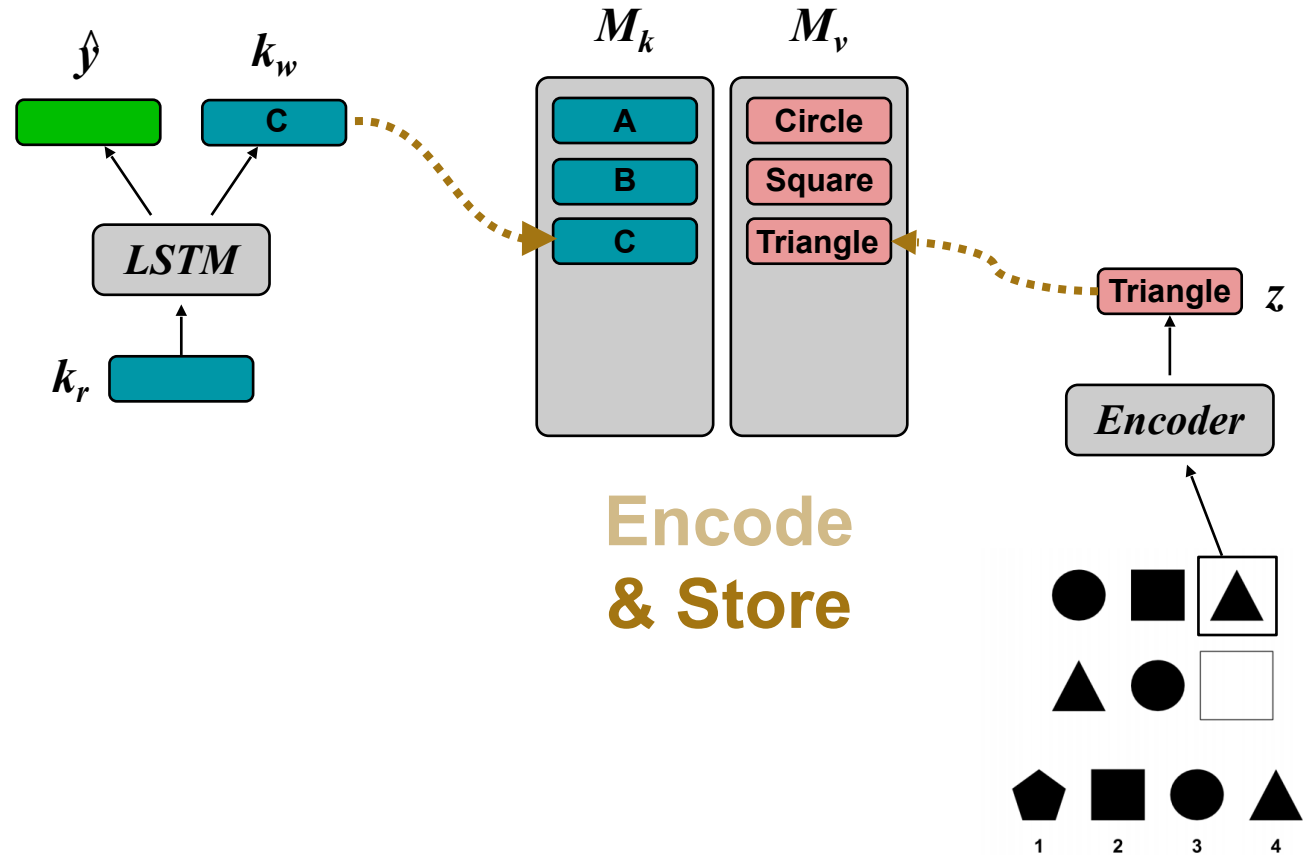
# Example Trial: Distribution of Three



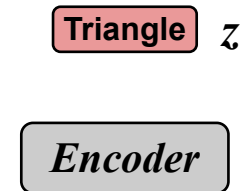
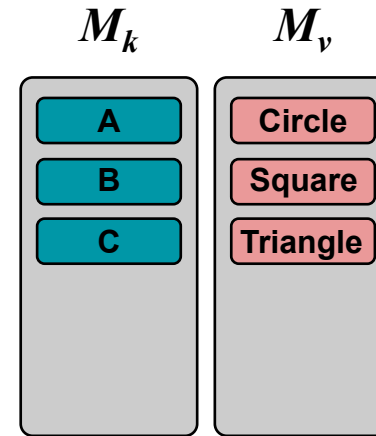
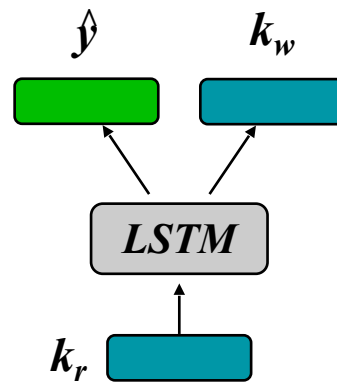
Encode  
& Store



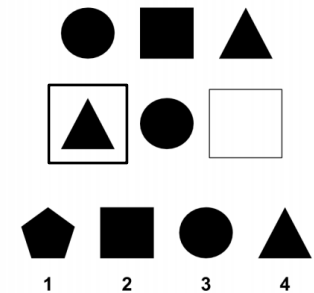
# Example Trial: Distribution of Three



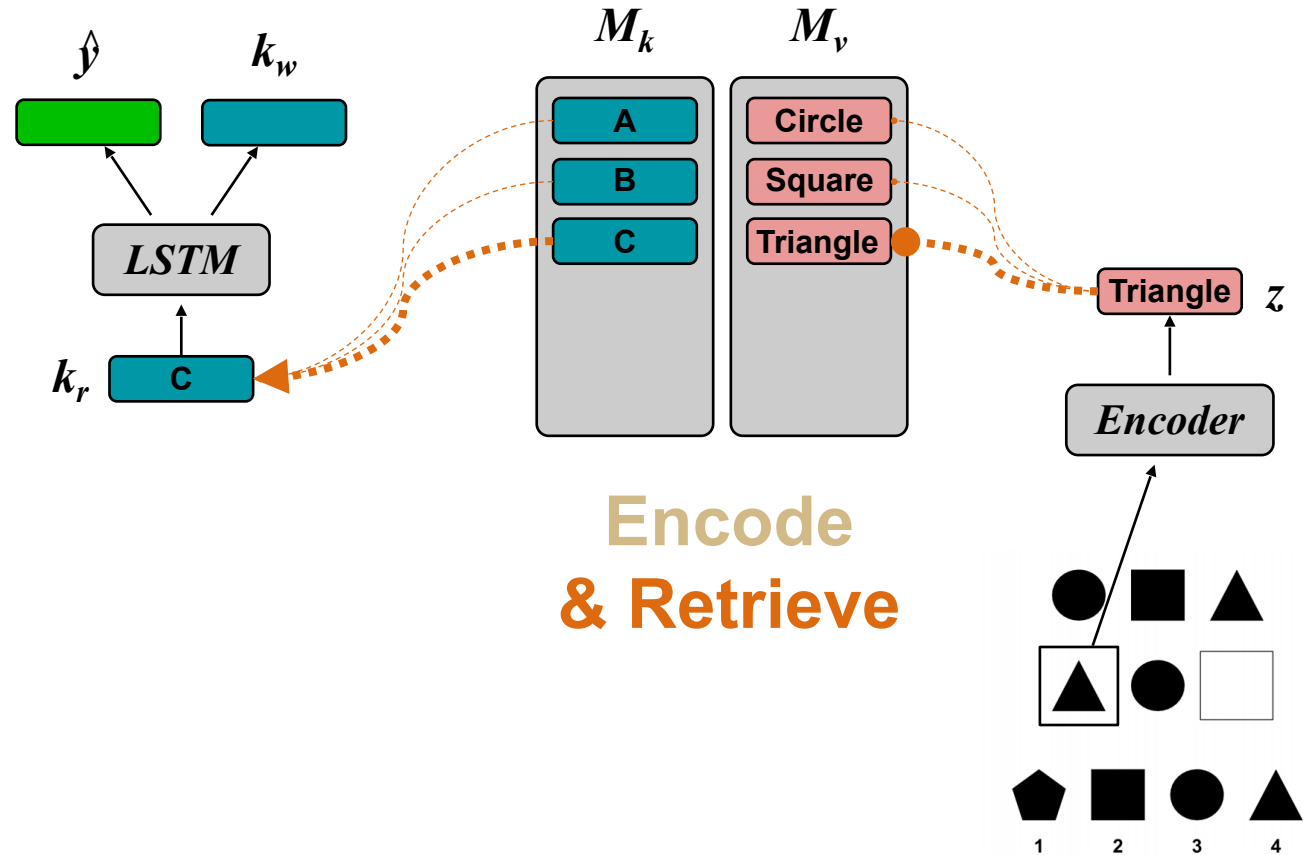
# Example Trial: Distribution of Three



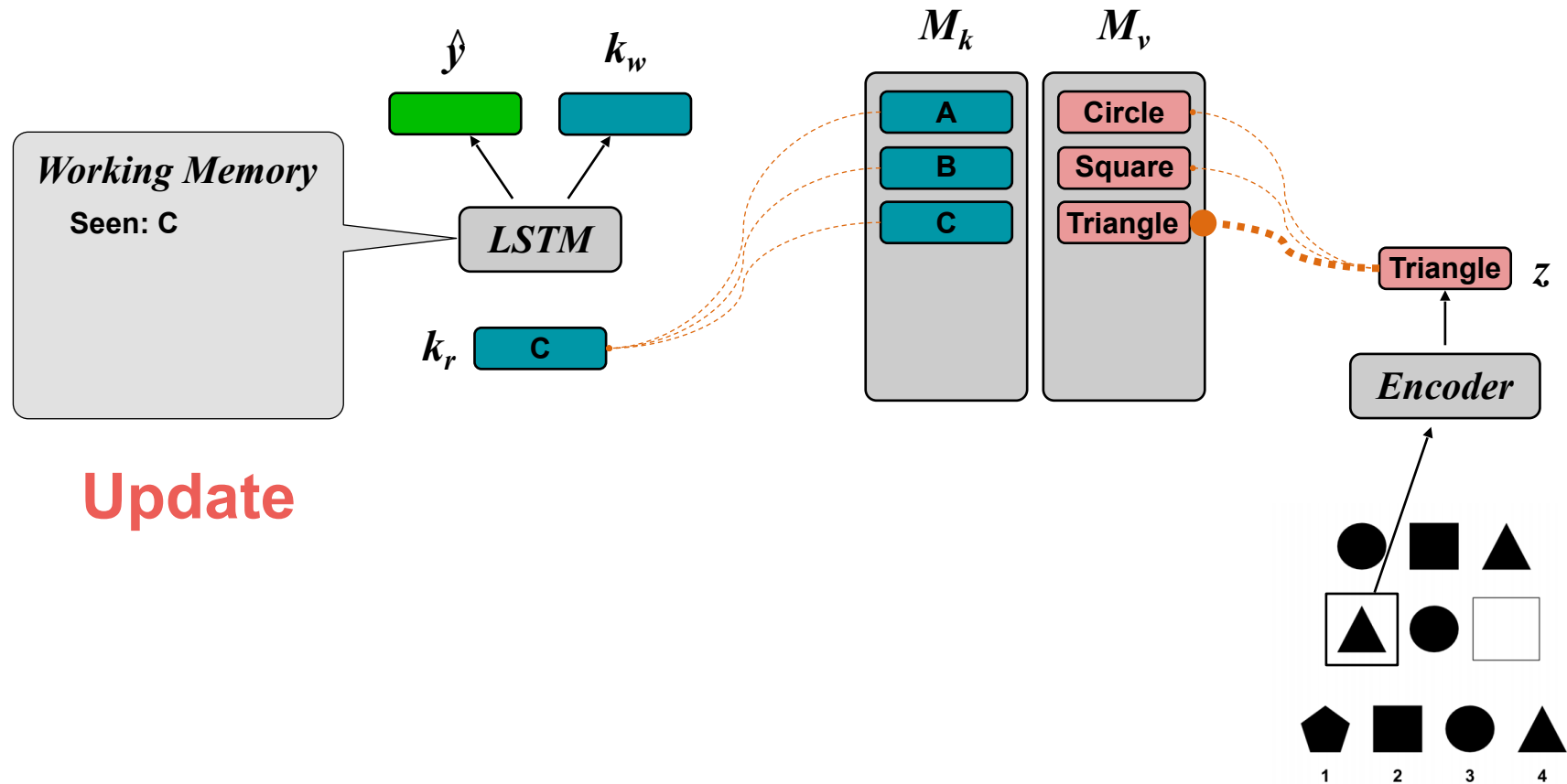
Encode  
& Retrieve



# Example Trial: Distribution of Three

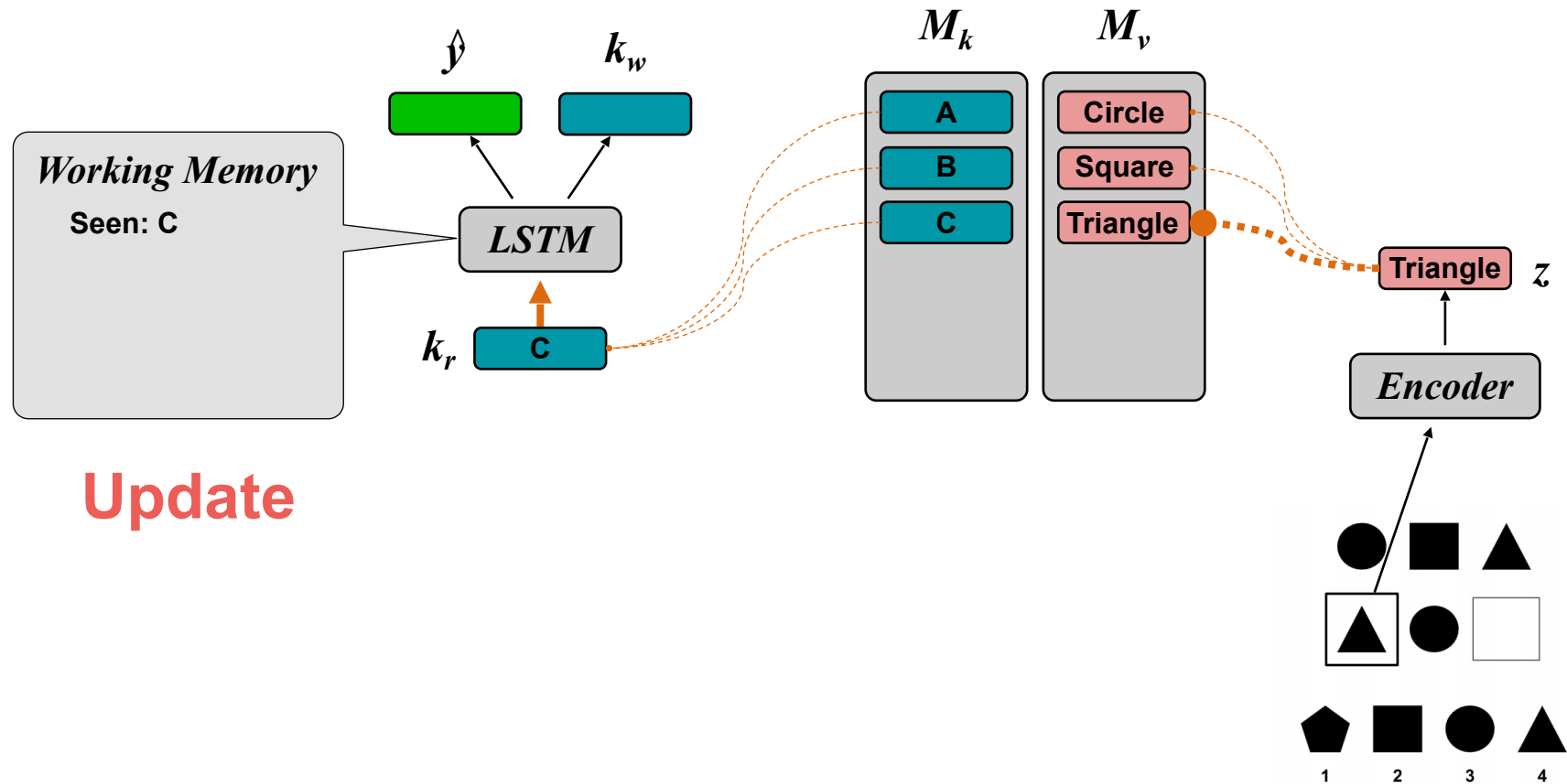


# Example Trial: Distribution of Three

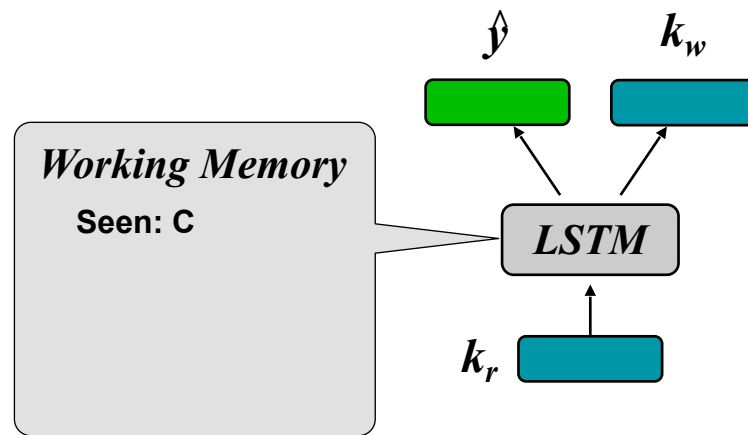




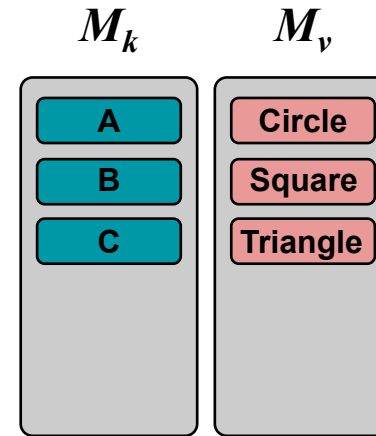
# Example Trial: Distribution of Three



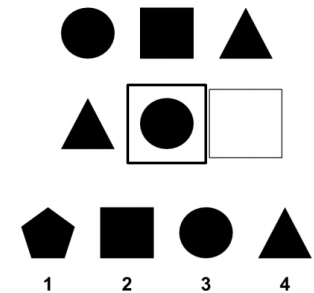
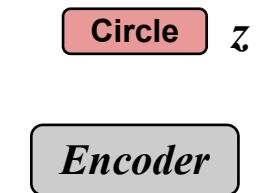
# Example Trial: Distribution of Three



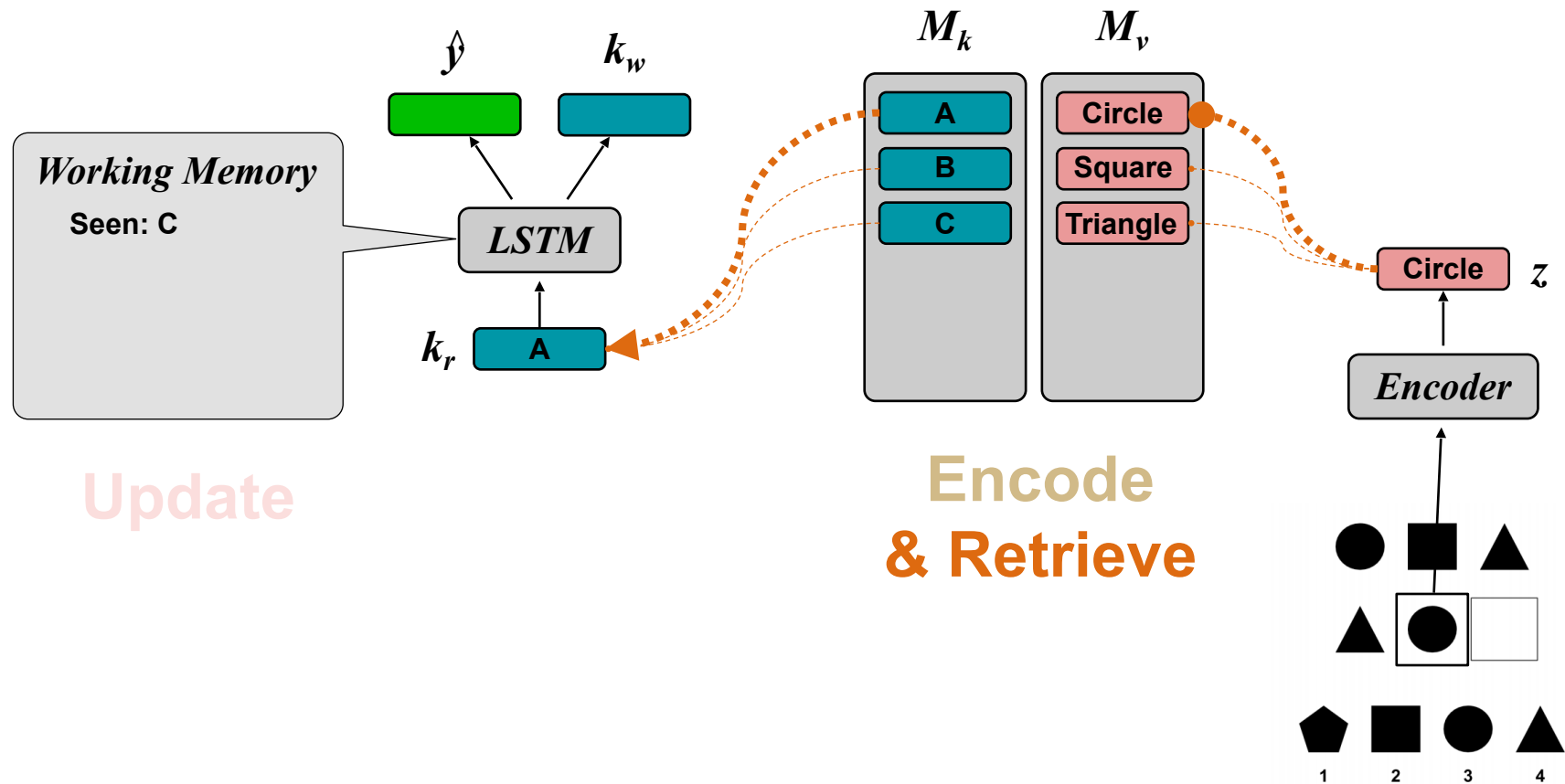
Update



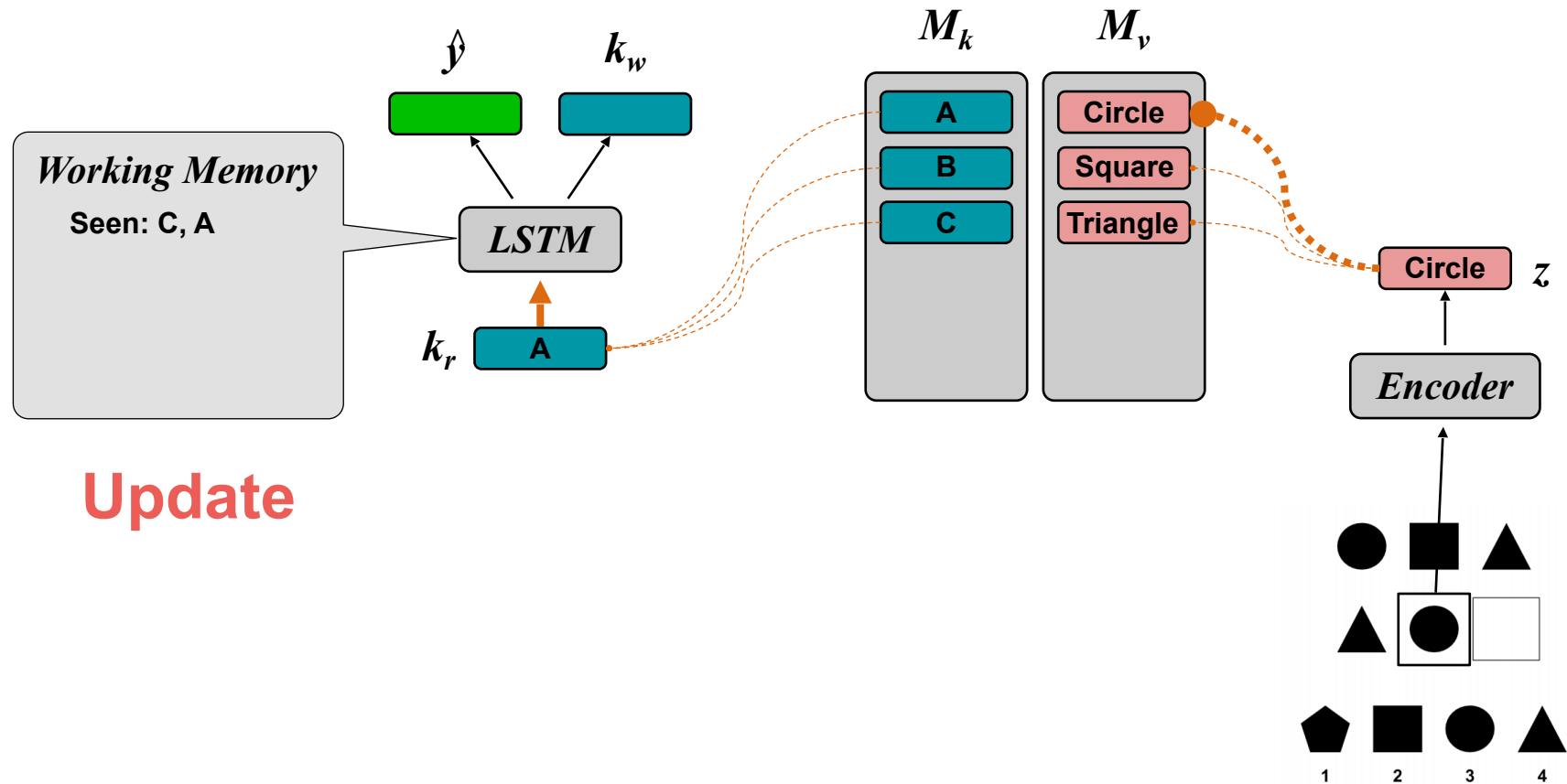
Encode  
& Retrieve



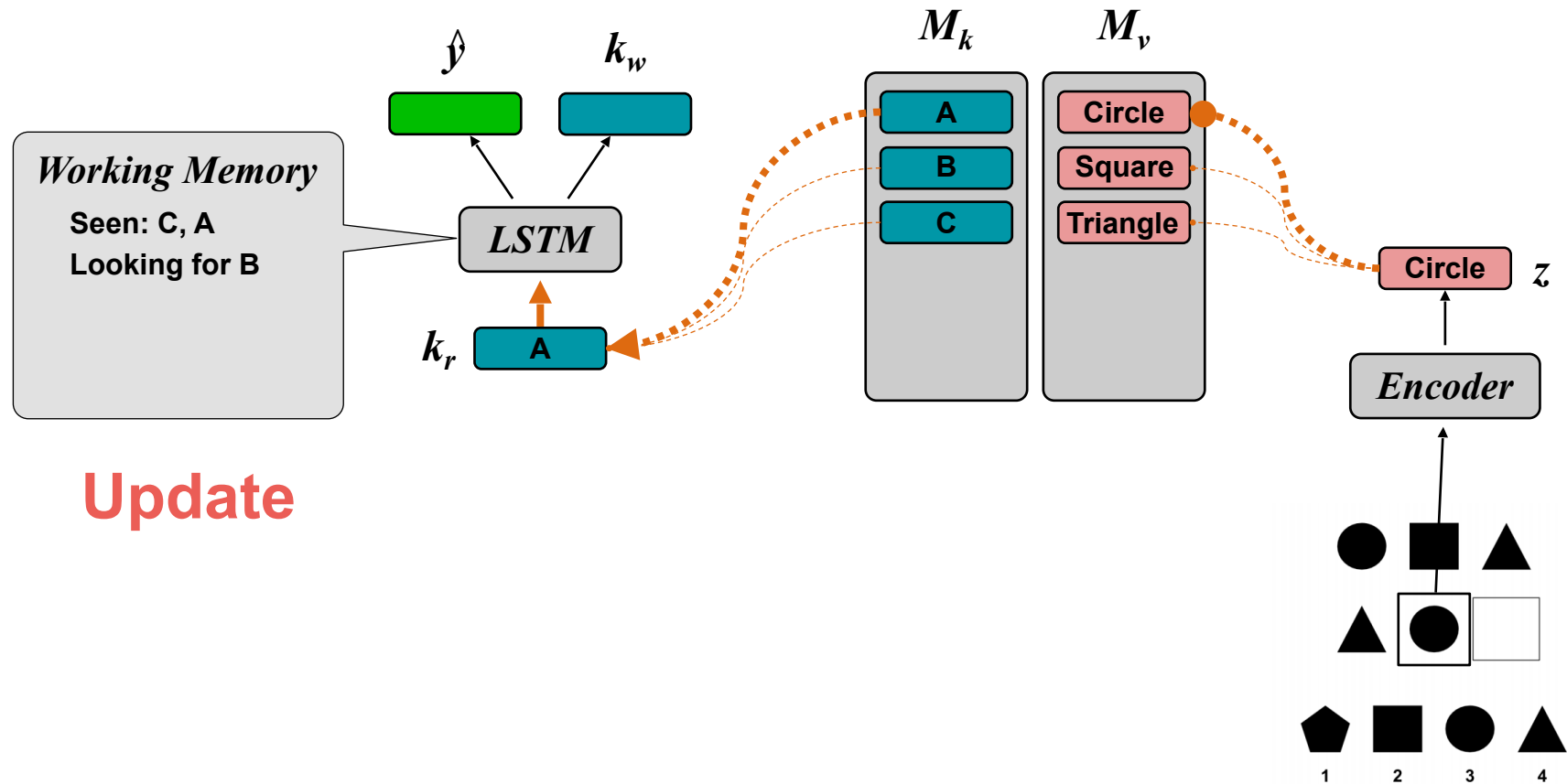
# Example Trial: Distribution of Three



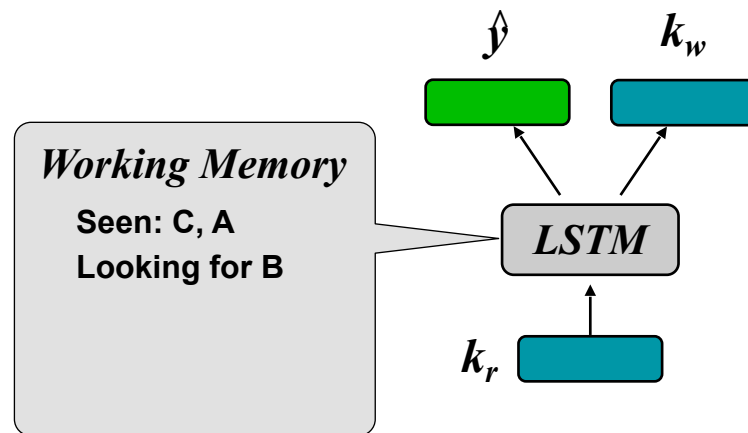
# Example Trial: Distribution of Three



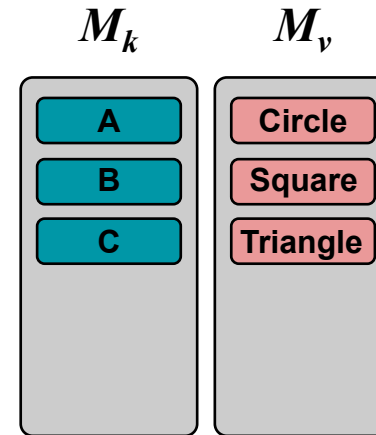
# Example Trial: Distribution of Three



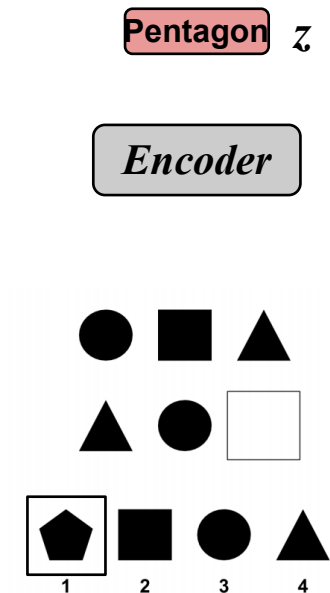
# Example Trial: Distribution of Three



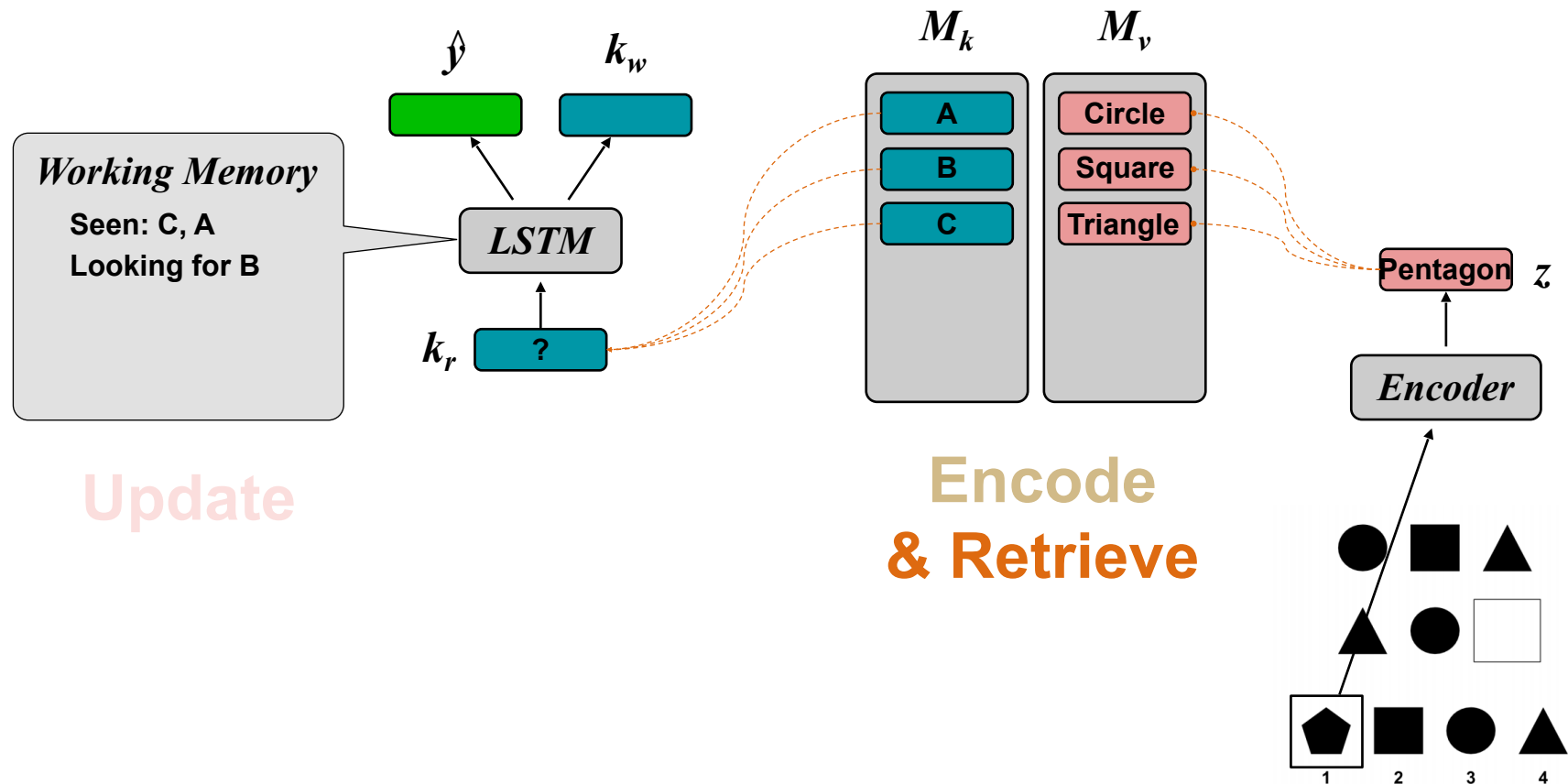
Update



Encode  
& Retrieve

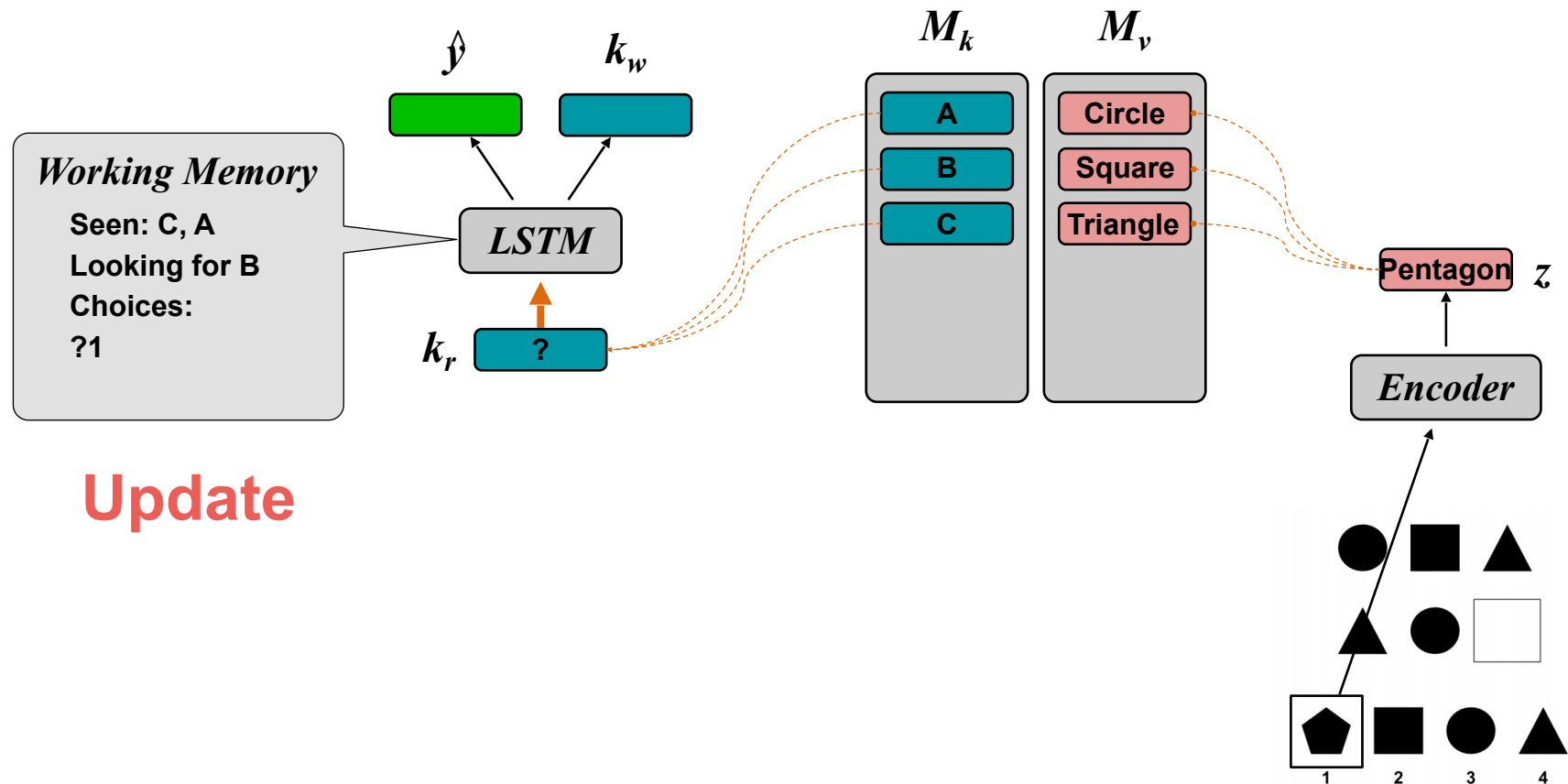


# Example Trial: Distribution of Three

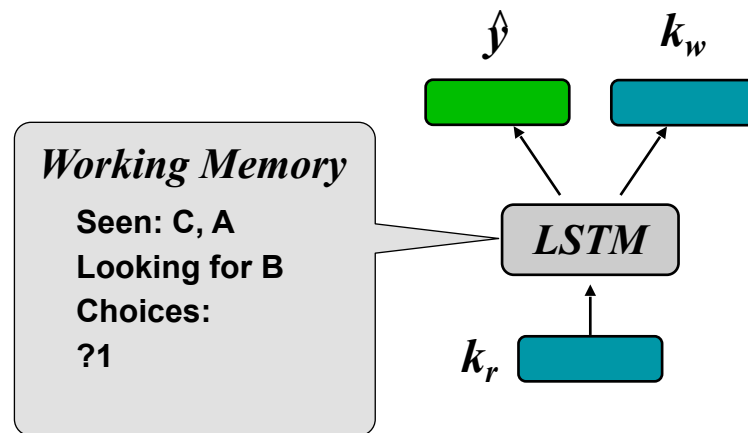




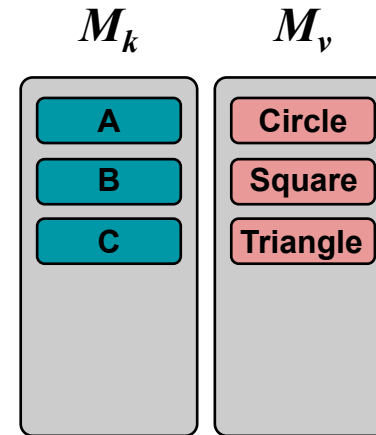
# Example Trial: Distribution of Three



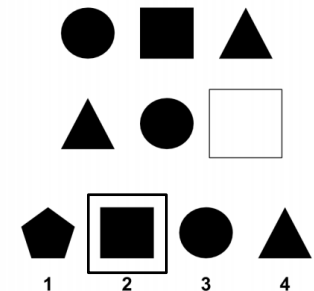
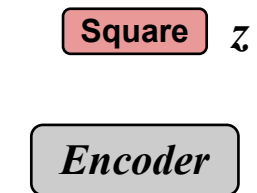
# Example Trial: Distribution of Three



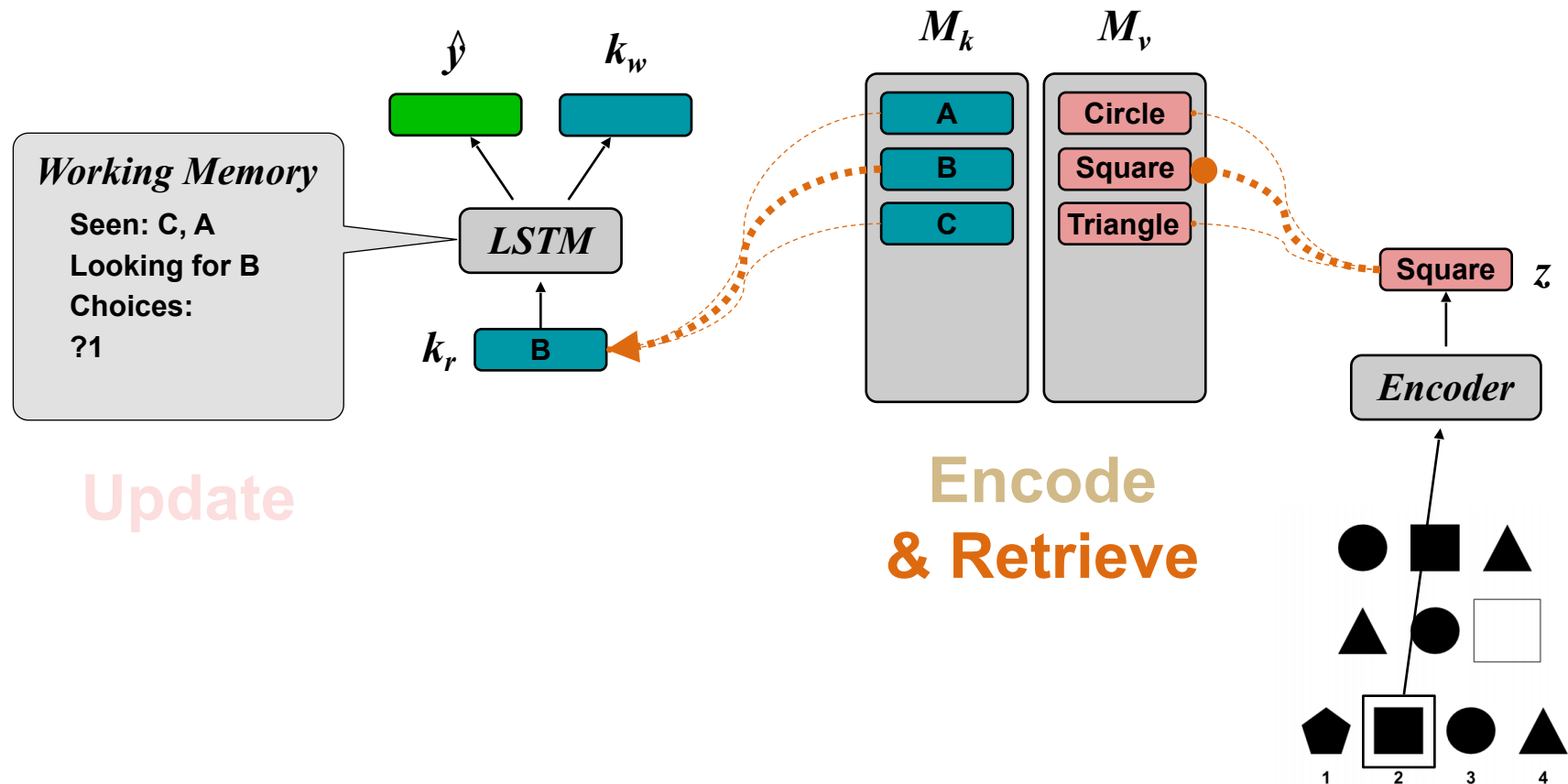
Update



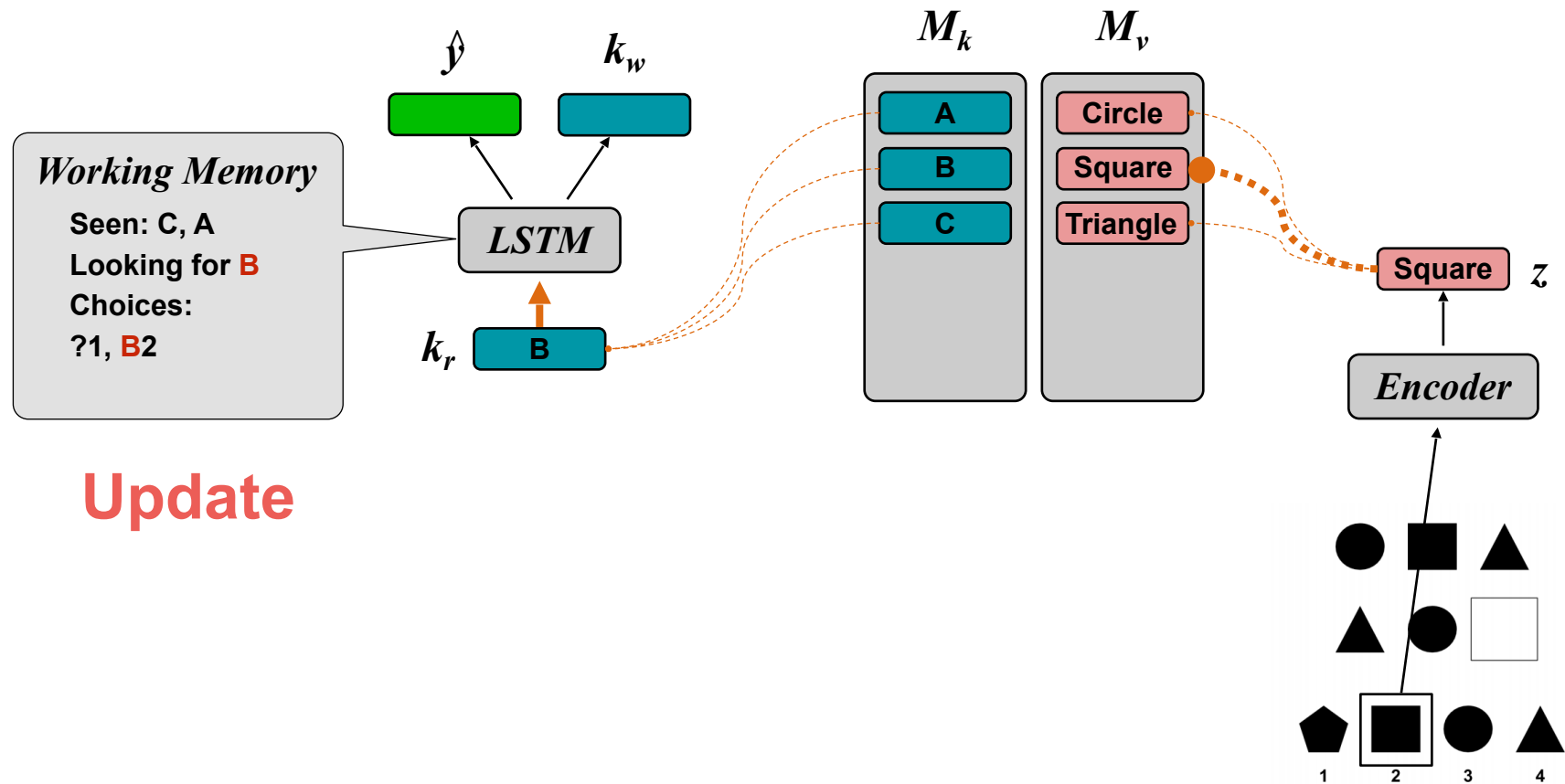
Encode  
& Retrieve



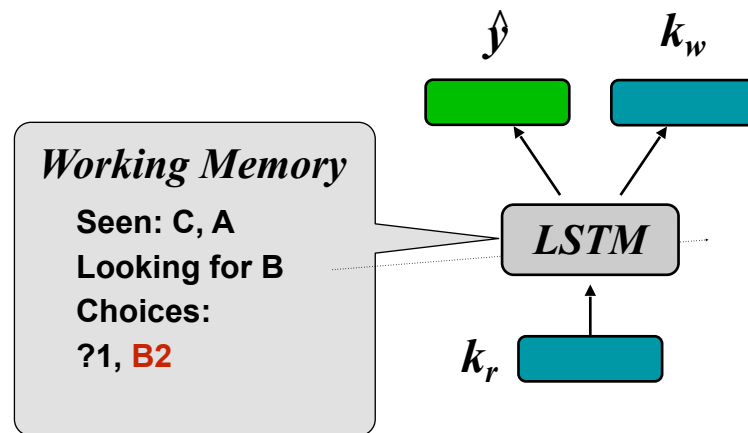
# Example Trial: Distribution of Three



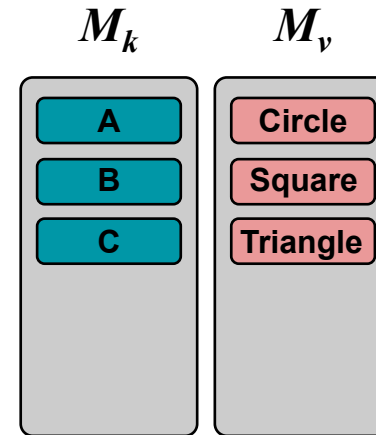
# Example Trial: Distribution of Three



# Example Trial: Distribution of Three



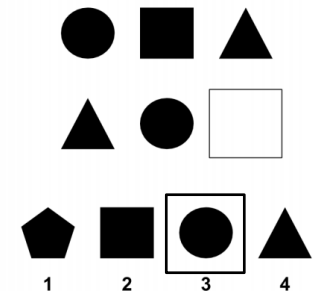
Update



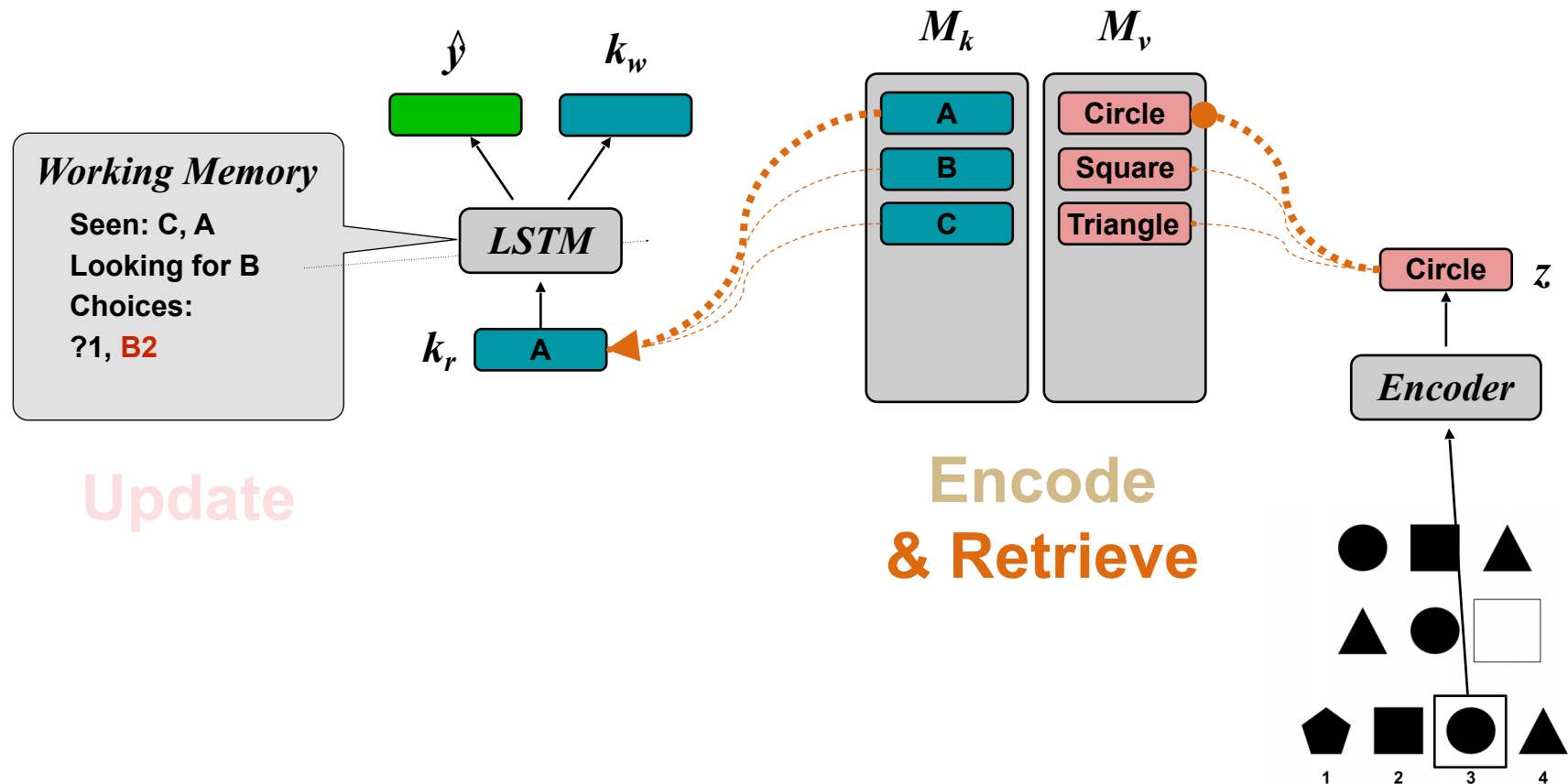
Encode  
& Retrieve

Circle  $z$

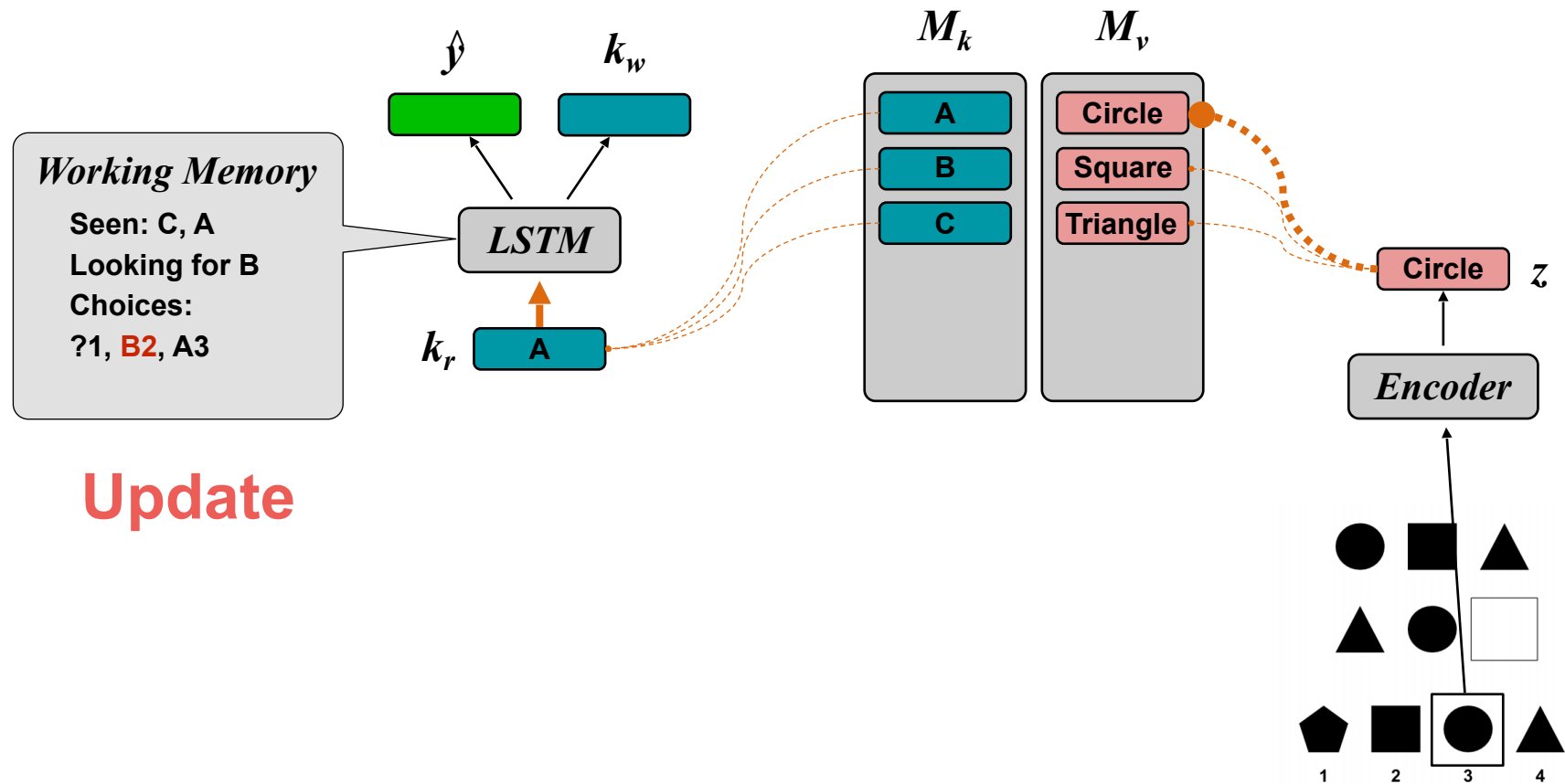
*Encoder*



# Example Trial: Distribution of Three

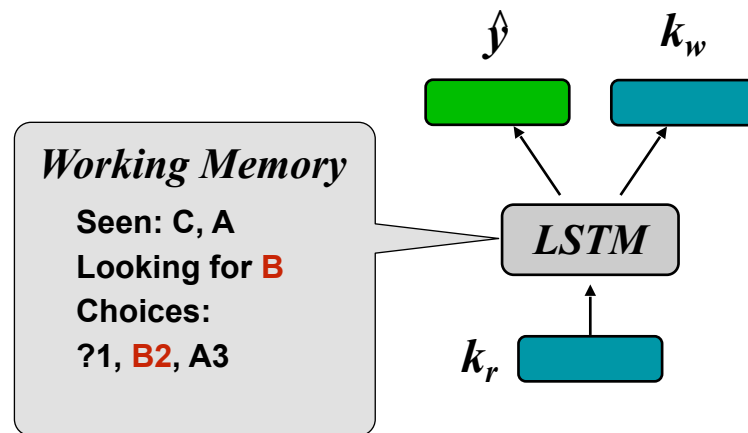


# Example Trial: Distribution of Three

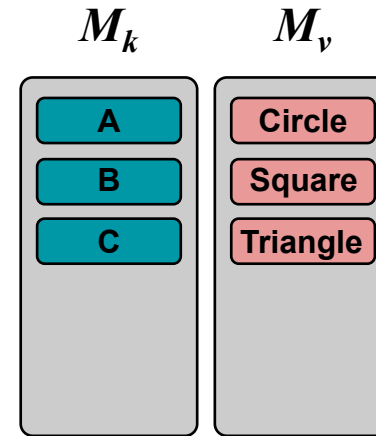




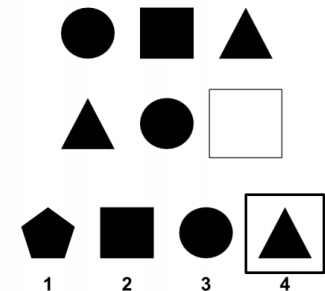
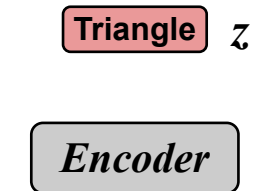
# Example Trial: Distribution of Three



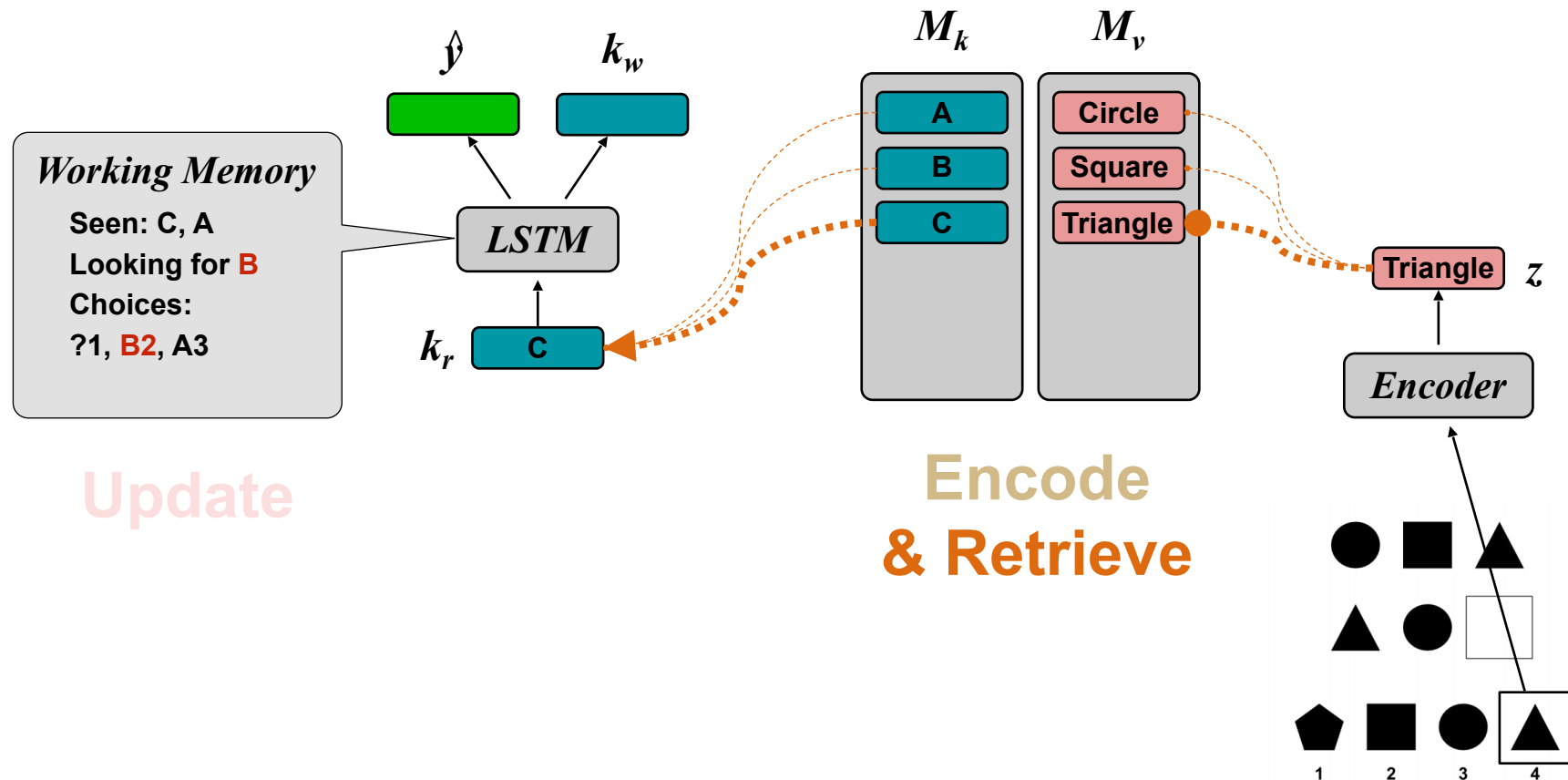
Update



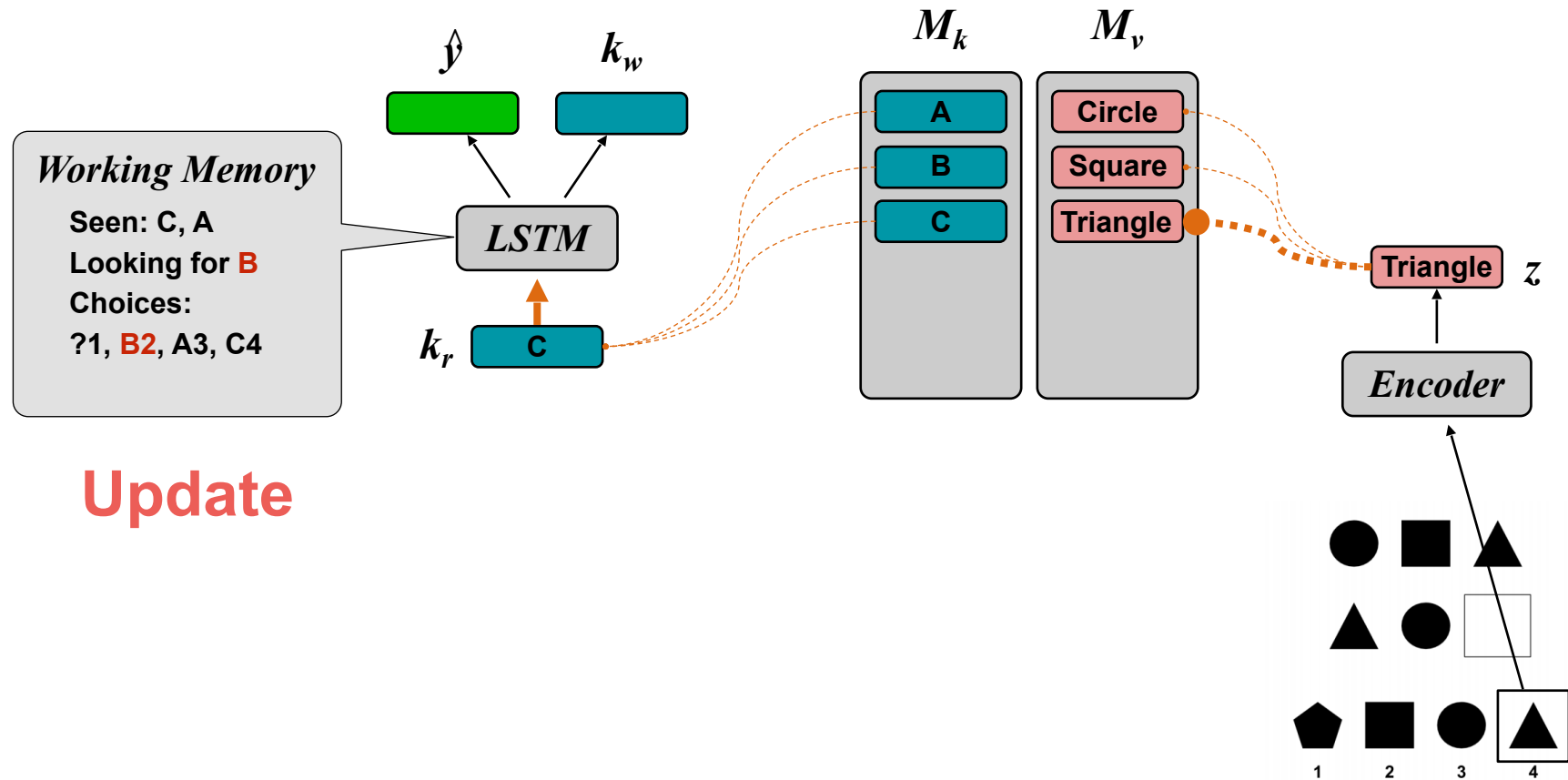
Encode  
& Retrieve



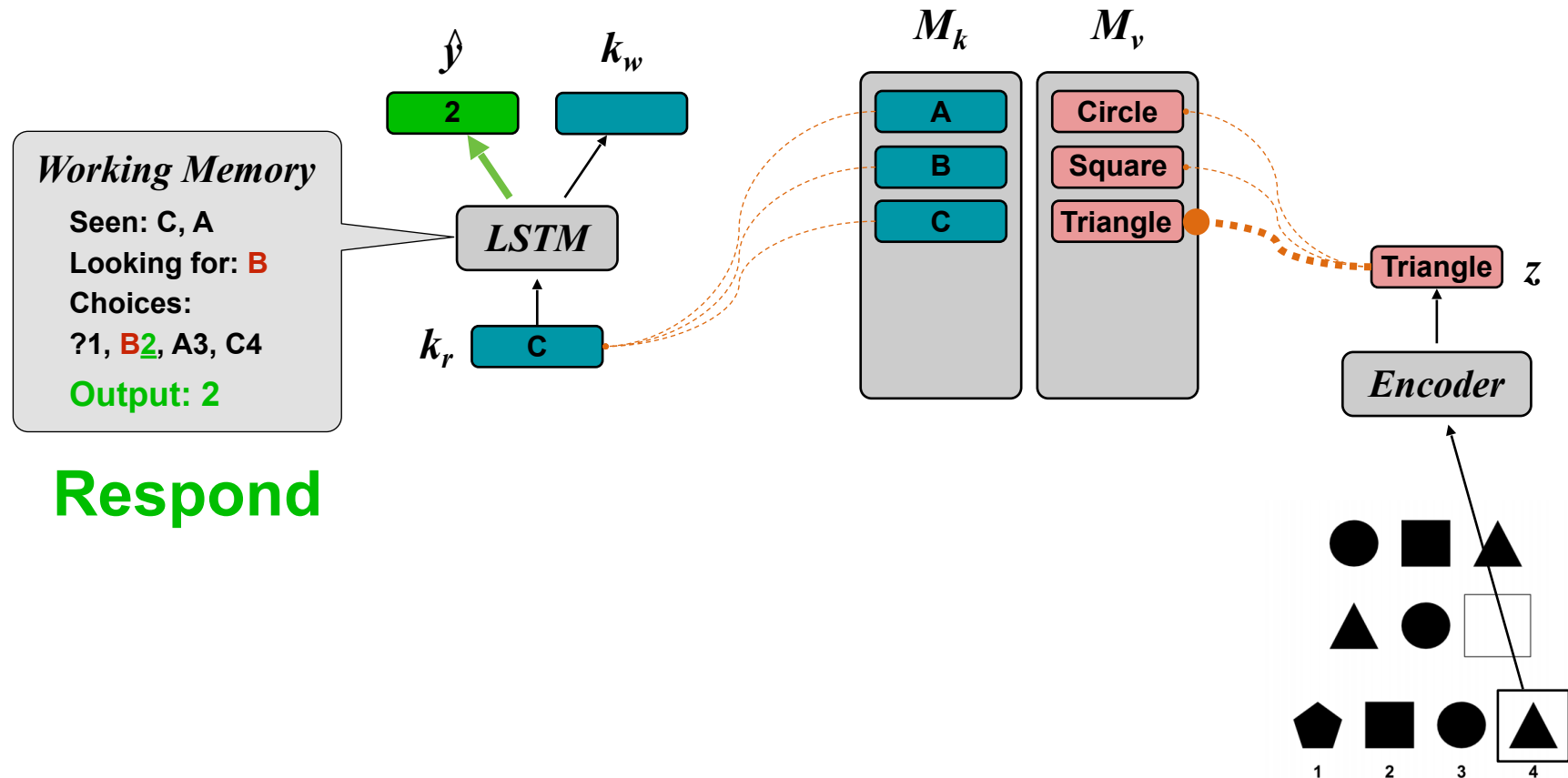
# Example Trial: Distribution of Three



# Example Trial: Distribution of Three

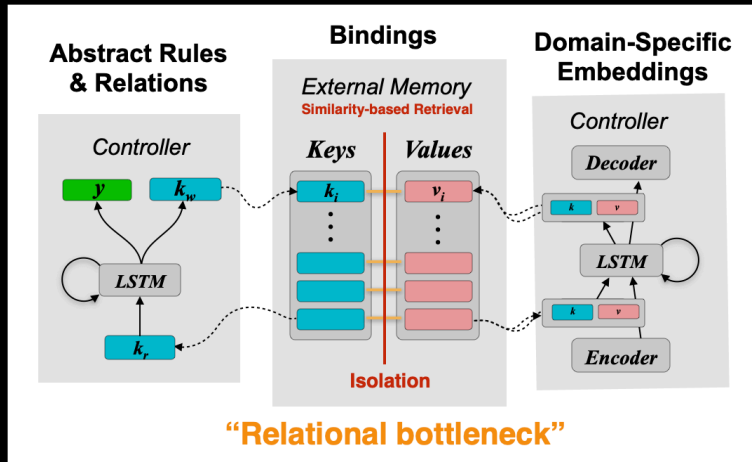


# Example Trial: Distribution of Three



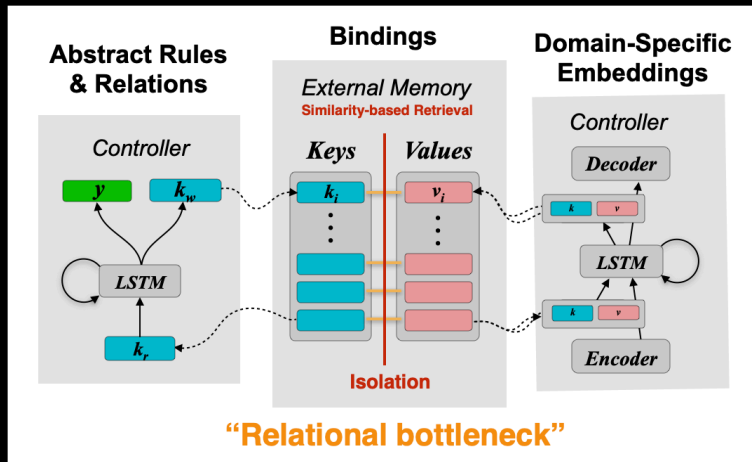
# Relational Bottleneck

---



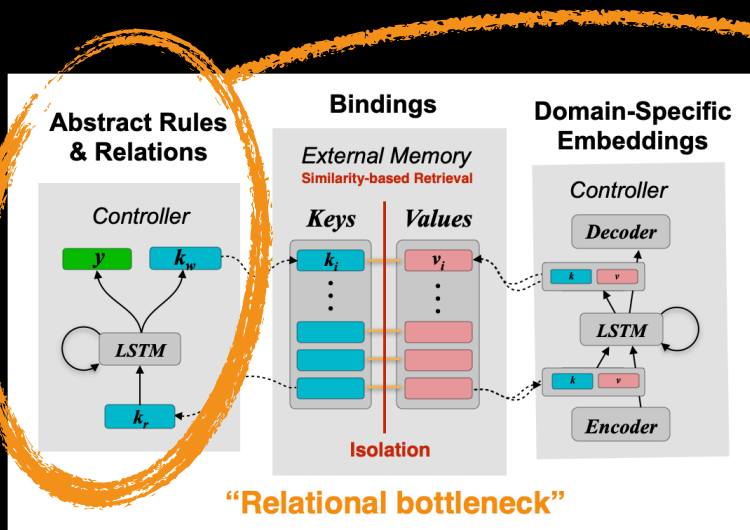
# Relational Bottleneck

External memory for isolation + similarity-based retrieval  
⇒ relational bottleneck



# Relational Bottleneck

External memory for isolation + similarity-based retrieval  
⇒ relational bottleneck



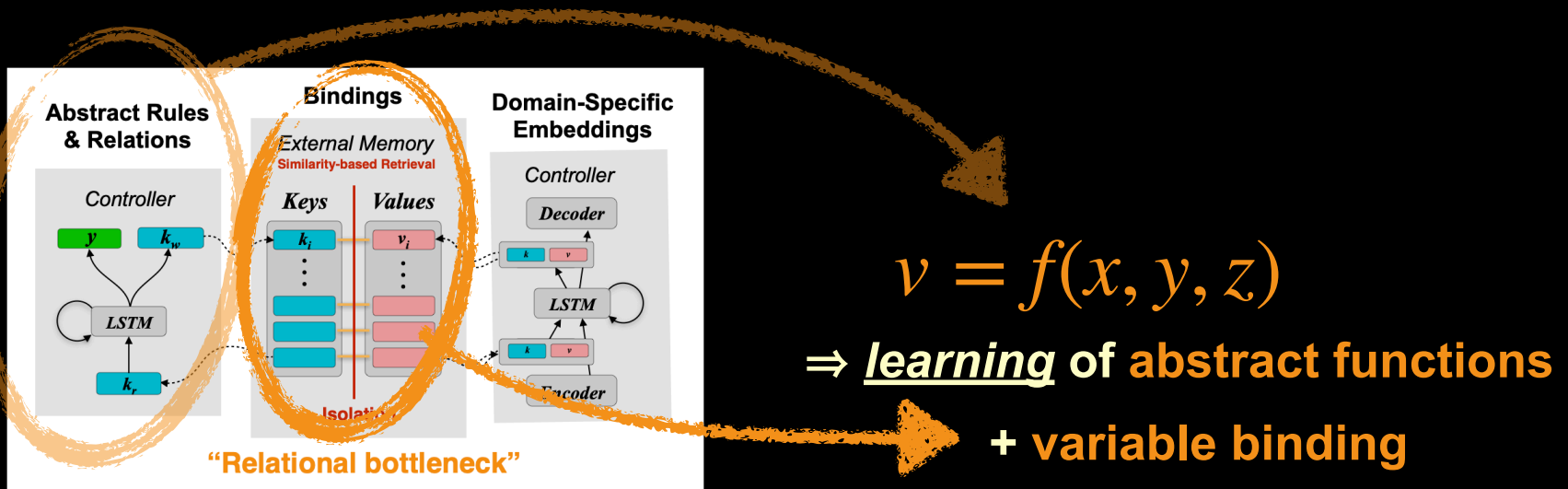
$$v = f(x, y, z)$$

⇒ learning of abstract functions

# Relational Bottleneck

External memory for isolation + similarity-based retrieval

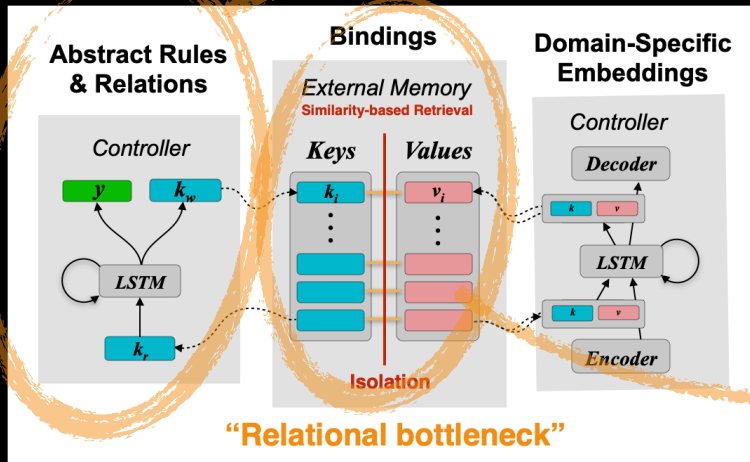
⇒ relational bottleneck





# Relational Bottleneck

External memory for isolation + similarity-based retrieval  
⇒ relational bottleneck



$$v = f(x, y, z)$$

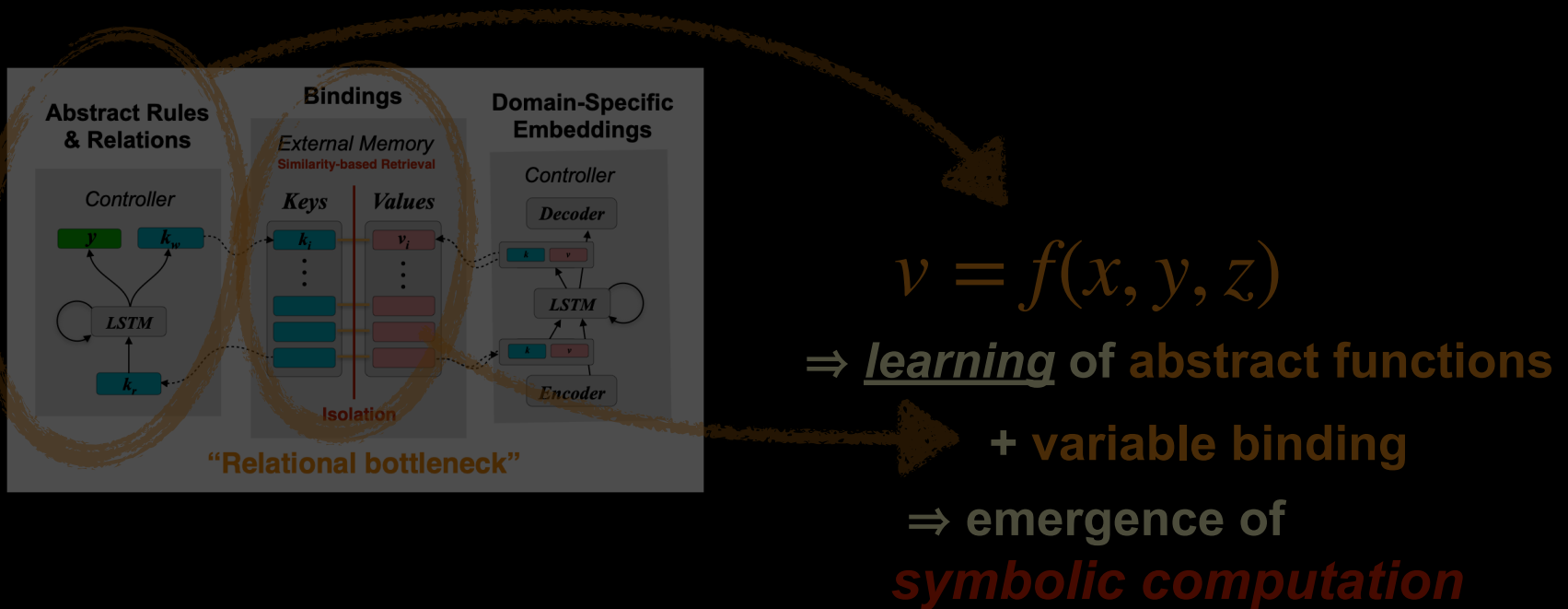
⇒ learning of abstract functions

+ variable binding

⇒ emergence of  
*symbolic computation*

# Relational Bottleneck

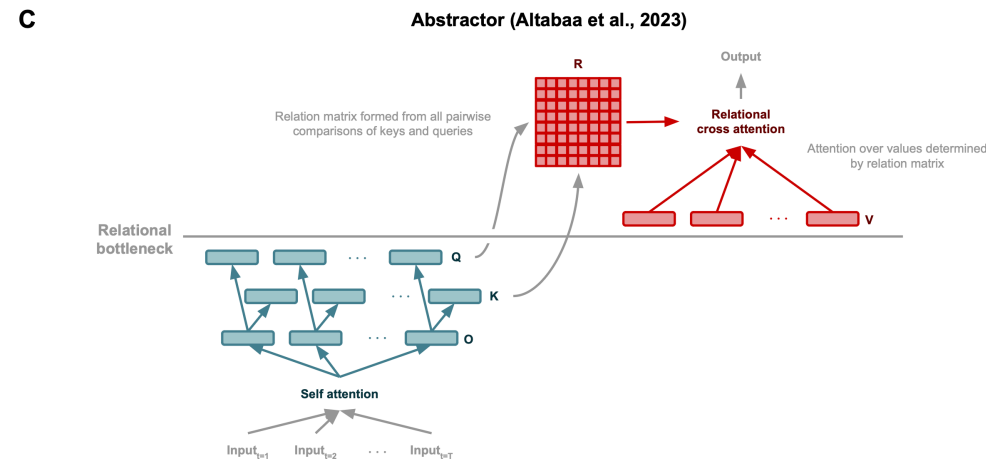
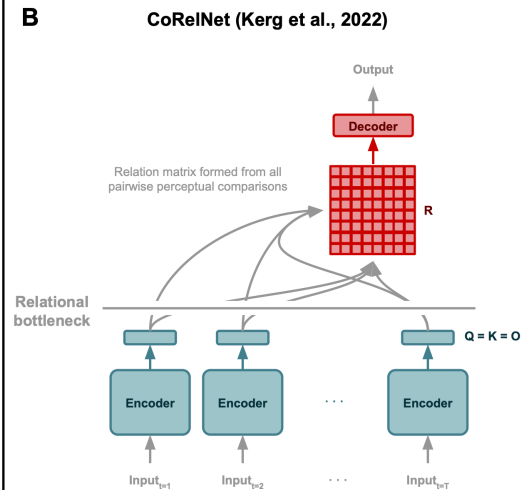
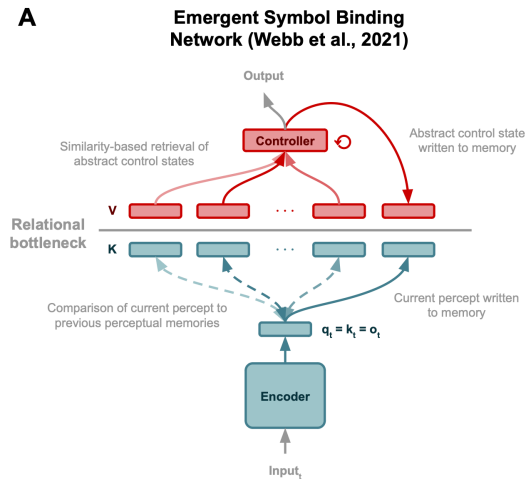
External memory for isolation + similarity-based retrieval  
⇒ relational bottleneck



Broad applicability...

# Principle Applies Across Architectures...

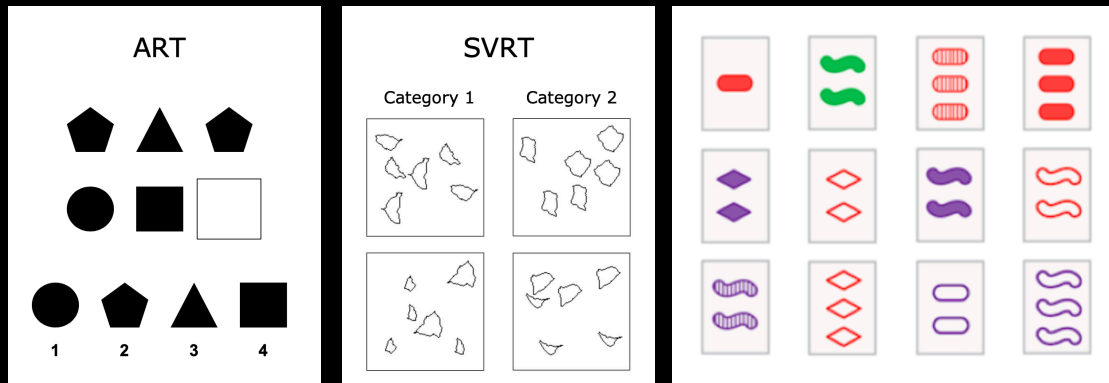
(Webb et al., 2023)



# ...and Task Domains

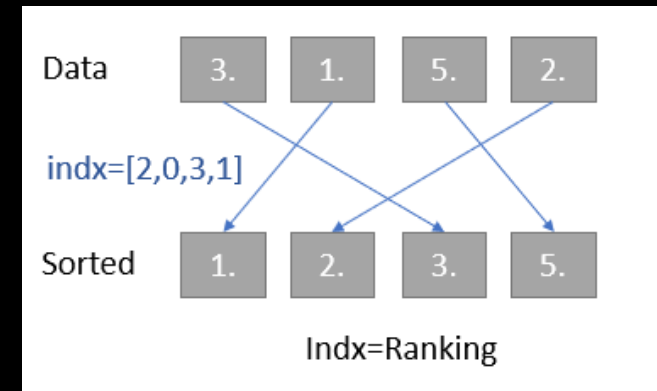
## Visual Relational Reasoning

(Mondal, Cohen & Webb, ICML 2023)



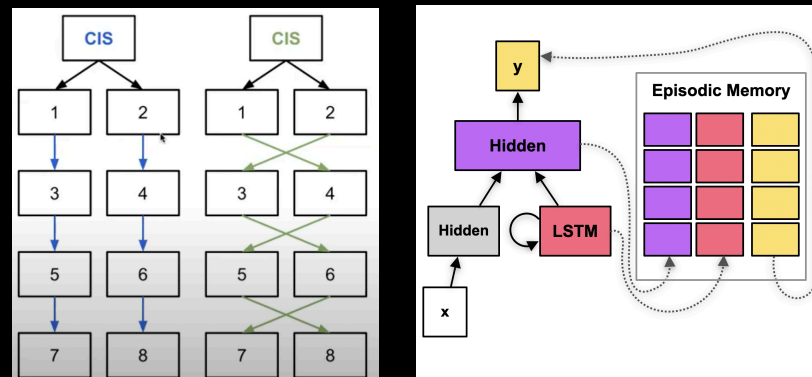
## Sequential Ordering

(Altabaa, Webb, Cohen & Lafferty, 2023)

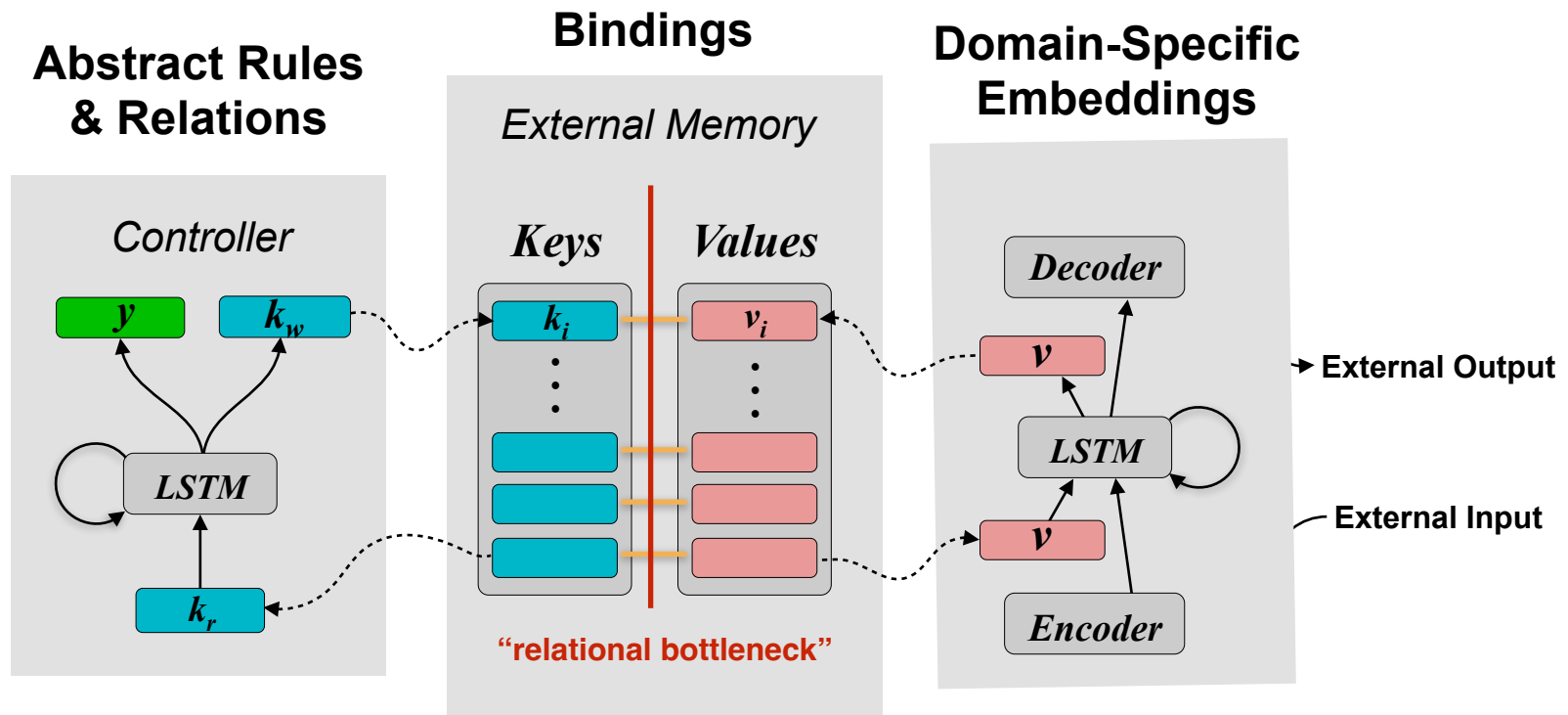


## Prediction and Planning

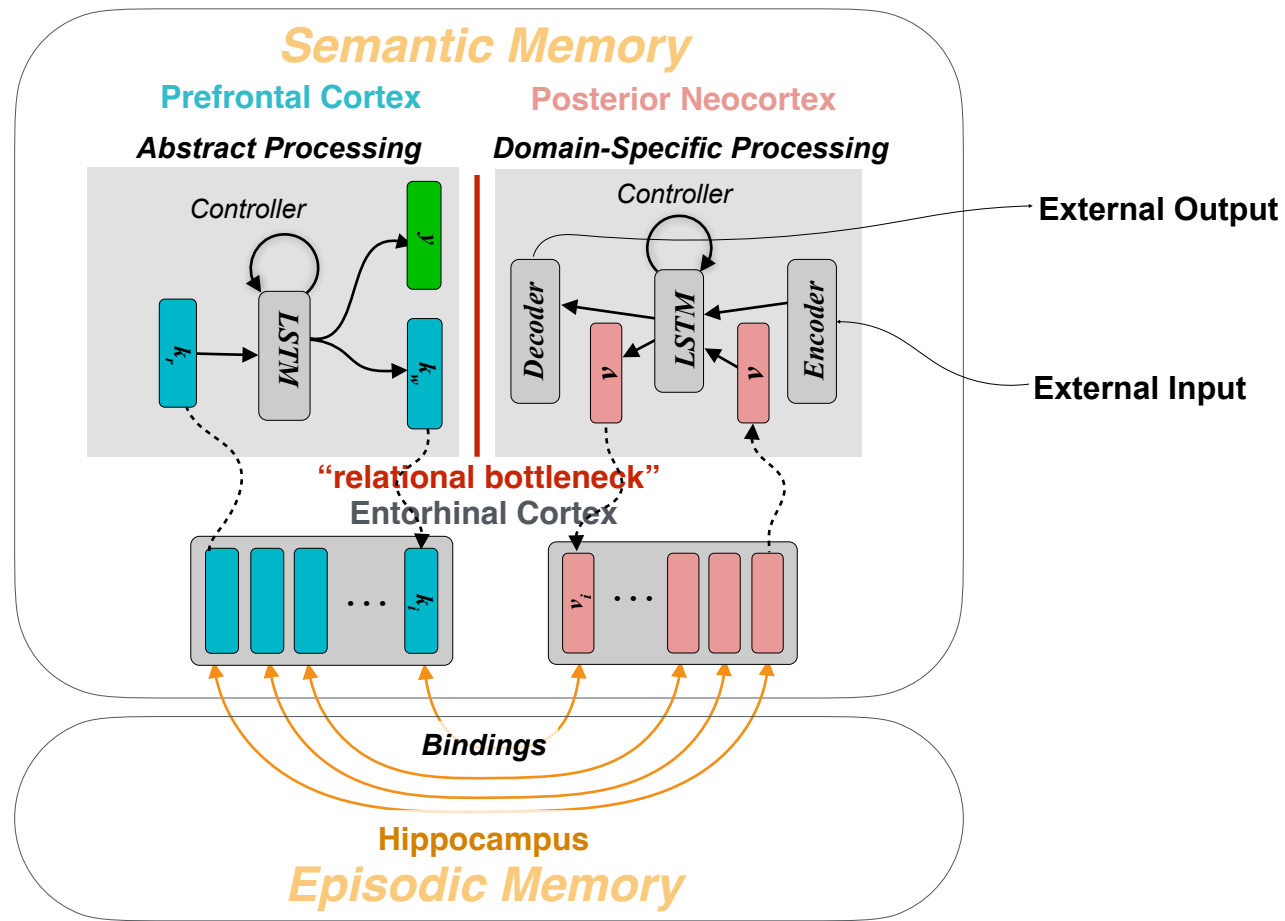
(Giallanza, Campbell & Cohen, 2023)



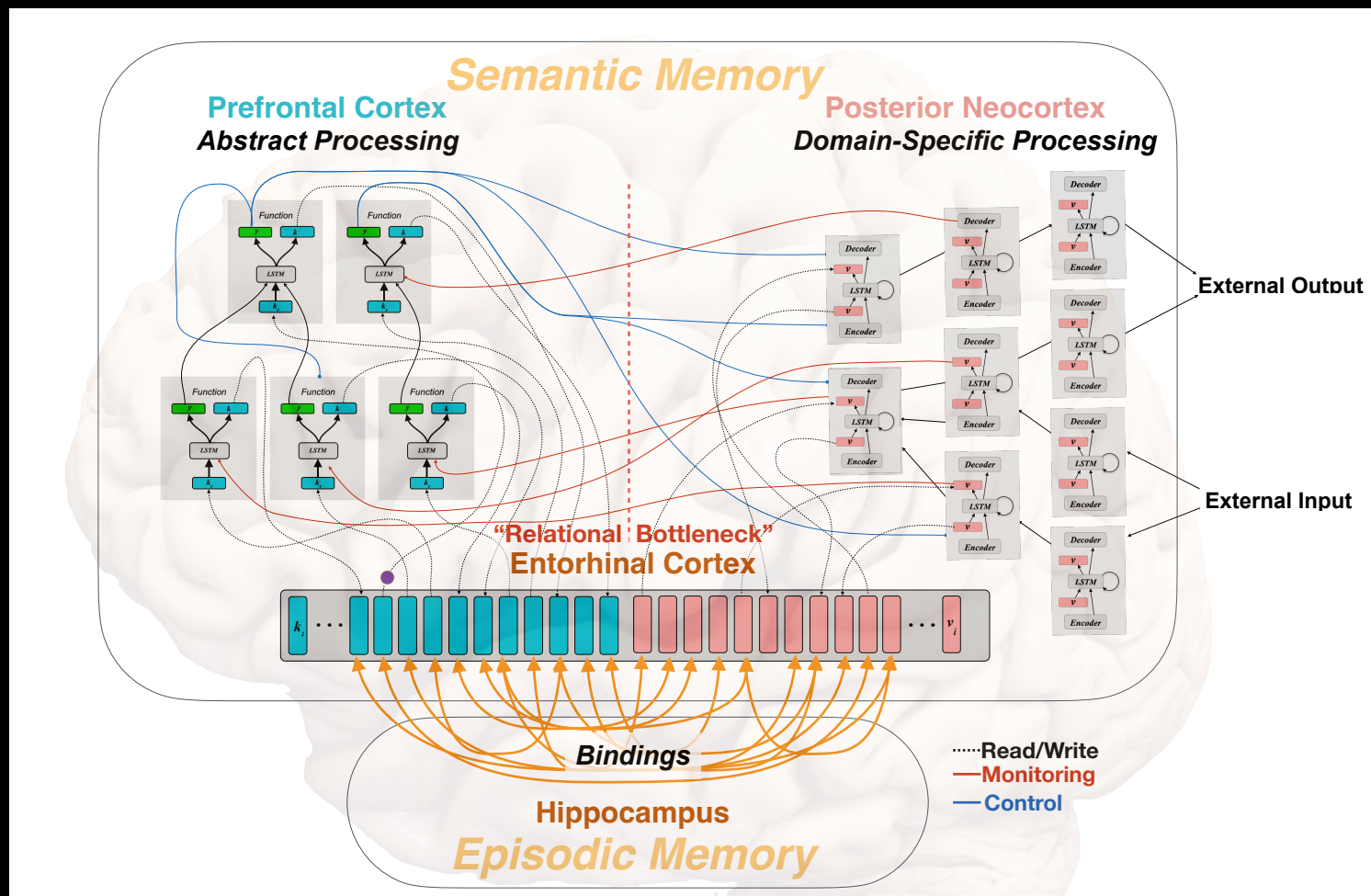
# Comports with Architecture of the Brain



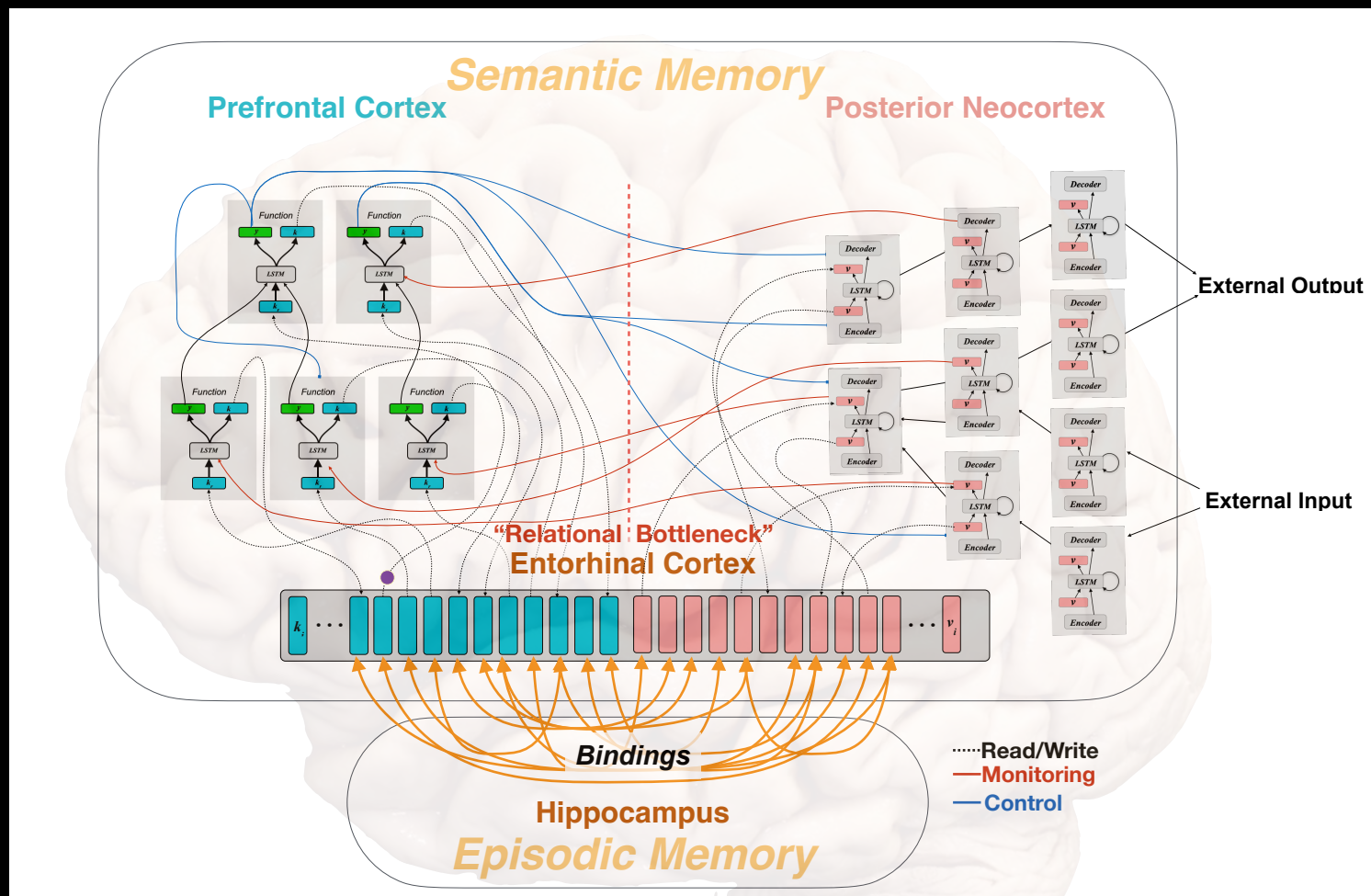
# Relational Bottleneck in the Brain



# Relational Bottleneck in the Brain



# Relational Bottleneck in the Brain



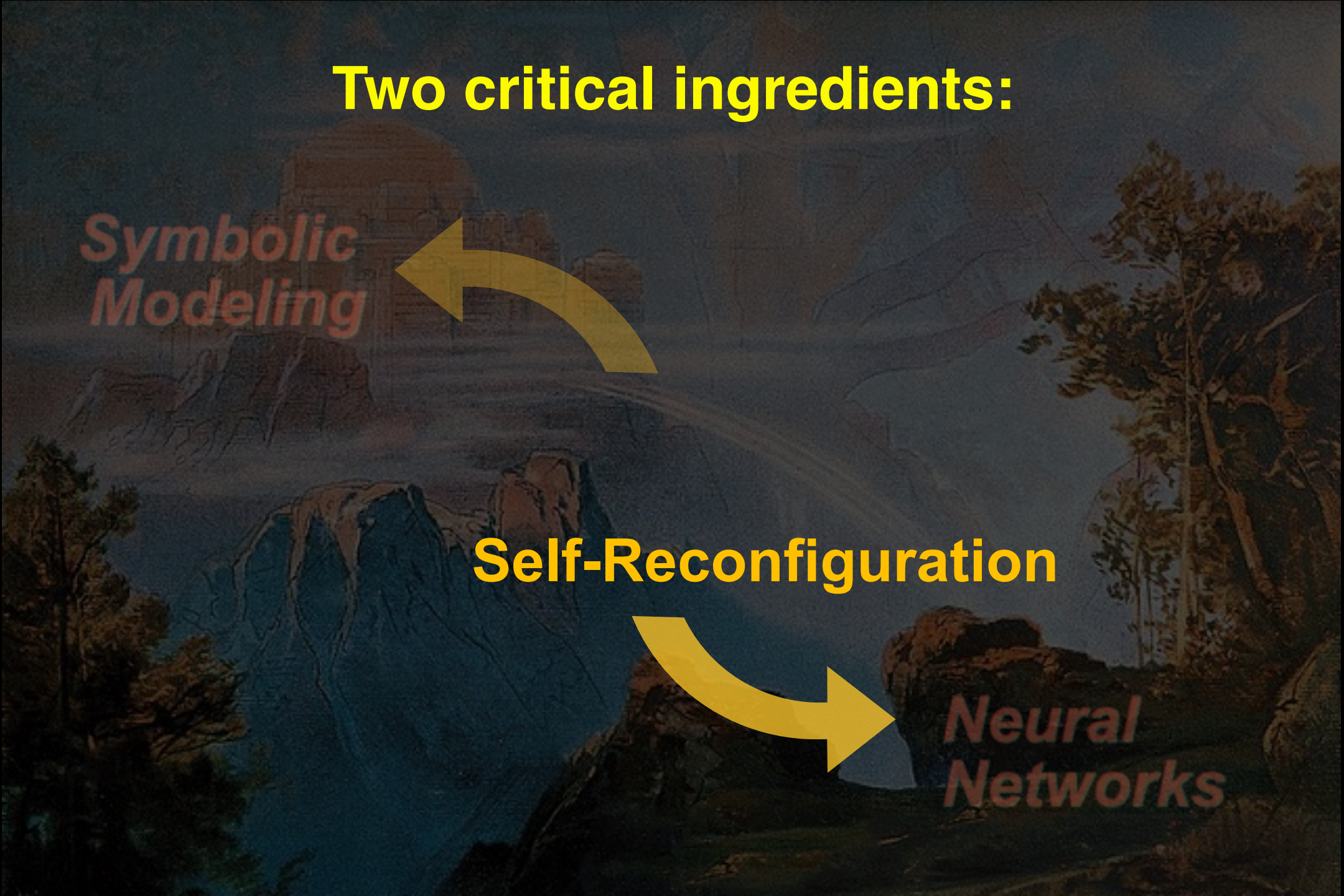


# Two critical ingredients:

*Symbolic  
Modeling*

**Self-Reconfiguration**

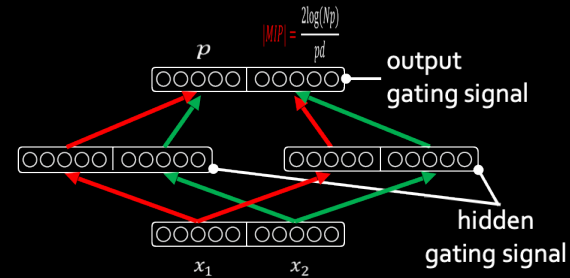
*Neural  
Networks*



# Rational Self-Reconfiguration

- **Formal analysis of learning speed vs. processing efficiency**  
(Musslick, Saxe et al., 2017)

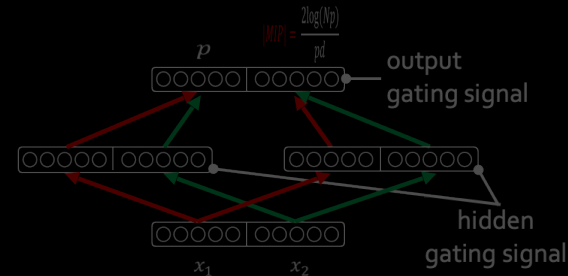
$$(\text{training time})^2 \propto \frac{(\text{multitasking capacity})}{(\% \text{shared representations})}$$



# Rational Self-Reconfiguration

- **Formal analysis of learning speed vs. processing efficiency**  
(Musslick, Saxe et al., 2017)

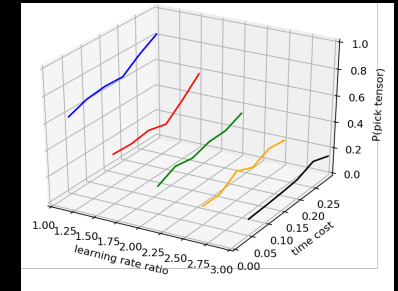
$$(\text{training time})^2 \propto \frac{(\text{multitasking capacity})}{(\% \text{shared representations})}$$



- **Bayesian optimal process model**  
(Sagiv, Musslick & Cohen, 2018)

$$\mathbb{E}_B[R|t] = \sum_{i=1}^{\min\{N,K\}} \mathbb{P}(\alpha = i) \sum_{j=0}^{i-1} \mathbb{P}_B(\text{success on task } j)(1 - jC)$$

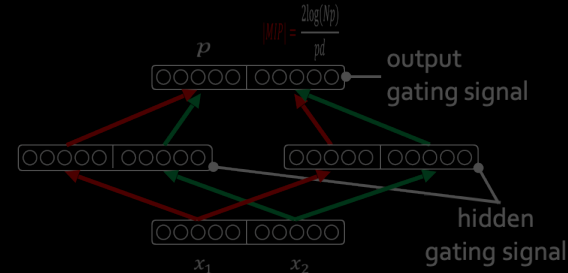
$$\mathbb{E}_T[R|t] = \sum_{i=1}^{\min\{N,K\}} \mathbb{P}(\alpha = i) \sum_{j=0}^{i-1} \mathbb{P}_T(\text{success on task } j)(1 - jC)$$



# Rational Self-Reconfiguration

- **Formal analysis of learning speed vs. processing efficiency**  
(Musslick, Saxe et al., 2017)

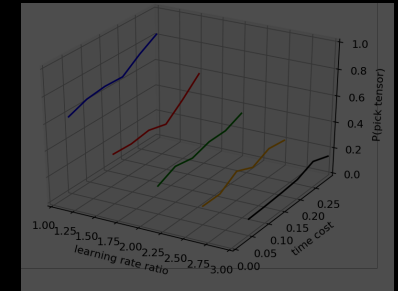
$$(\text{training time})^2 \propto \frac{(\text{multitasking capacity})}{(\% \text{shared representations})}$$



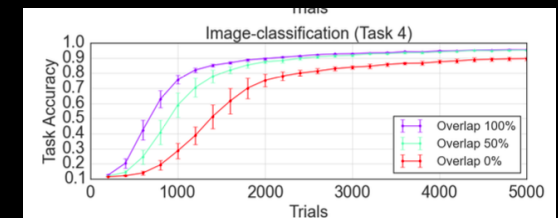
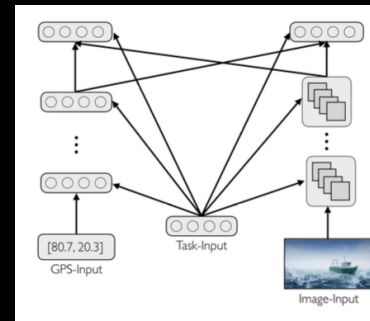
- **Bayesian optimal process model**  
(Sagiv, Musslick & Cohen, 2018)

$$\mathbb{E}_B[R|i] = \sum_{i=1}^{\min\{N,K\}} \mathbb{P}(\alpha = i) \sum_{j=0}^{i-1} \mathbb{P}_B(\text{success on task } j)(1 - jC)$$

$$\mathbb{E}_T[R|i] = \sum_{i=1}^{\min\{N,K\}} \mathbb{P}(\alpha = i) \sum_{j=0}^{i-1} \mathbb{P}_T(\text{success on task } j)(1 - jC)$$



- **Deep learning applications**  
(Ravi, Musslick & Cohen, under review)

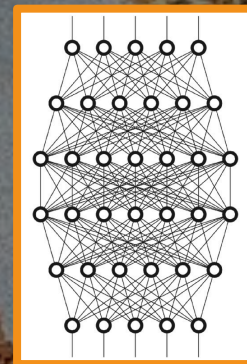
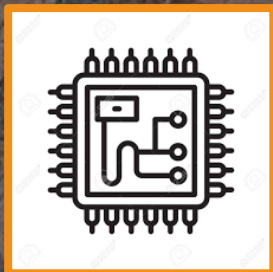


**Big Picture...**



# Natural Intelligence

***Symbolic  
Computing***

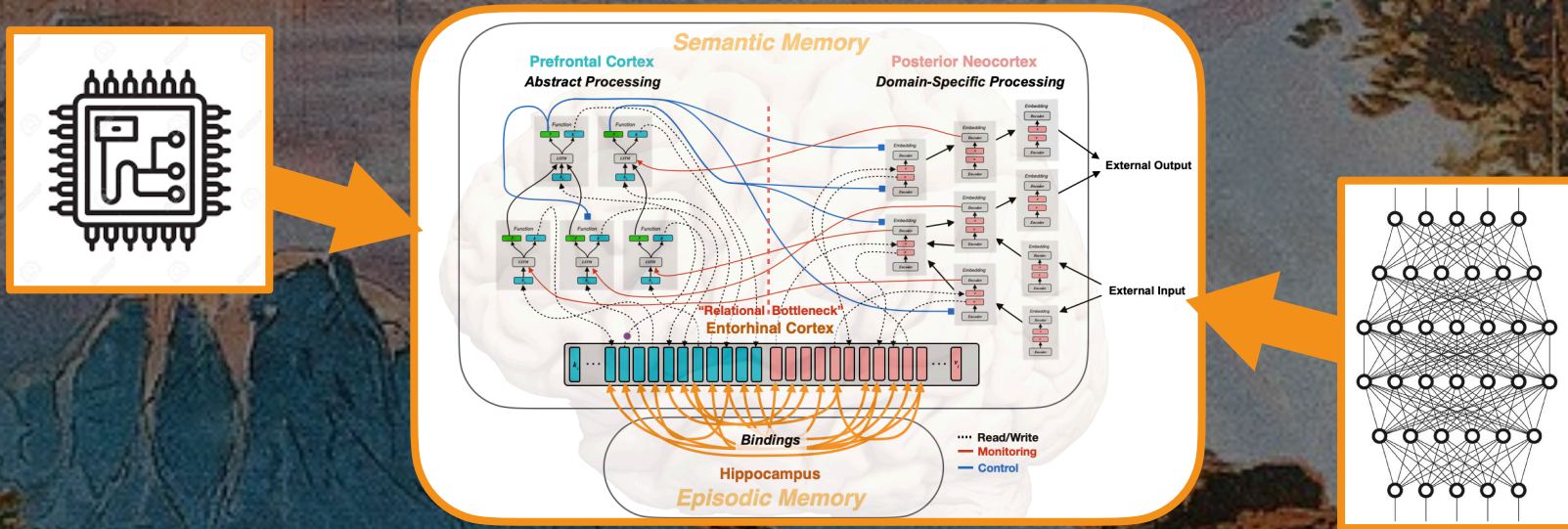


***Neural  
Networks***



# Natural Intelligence

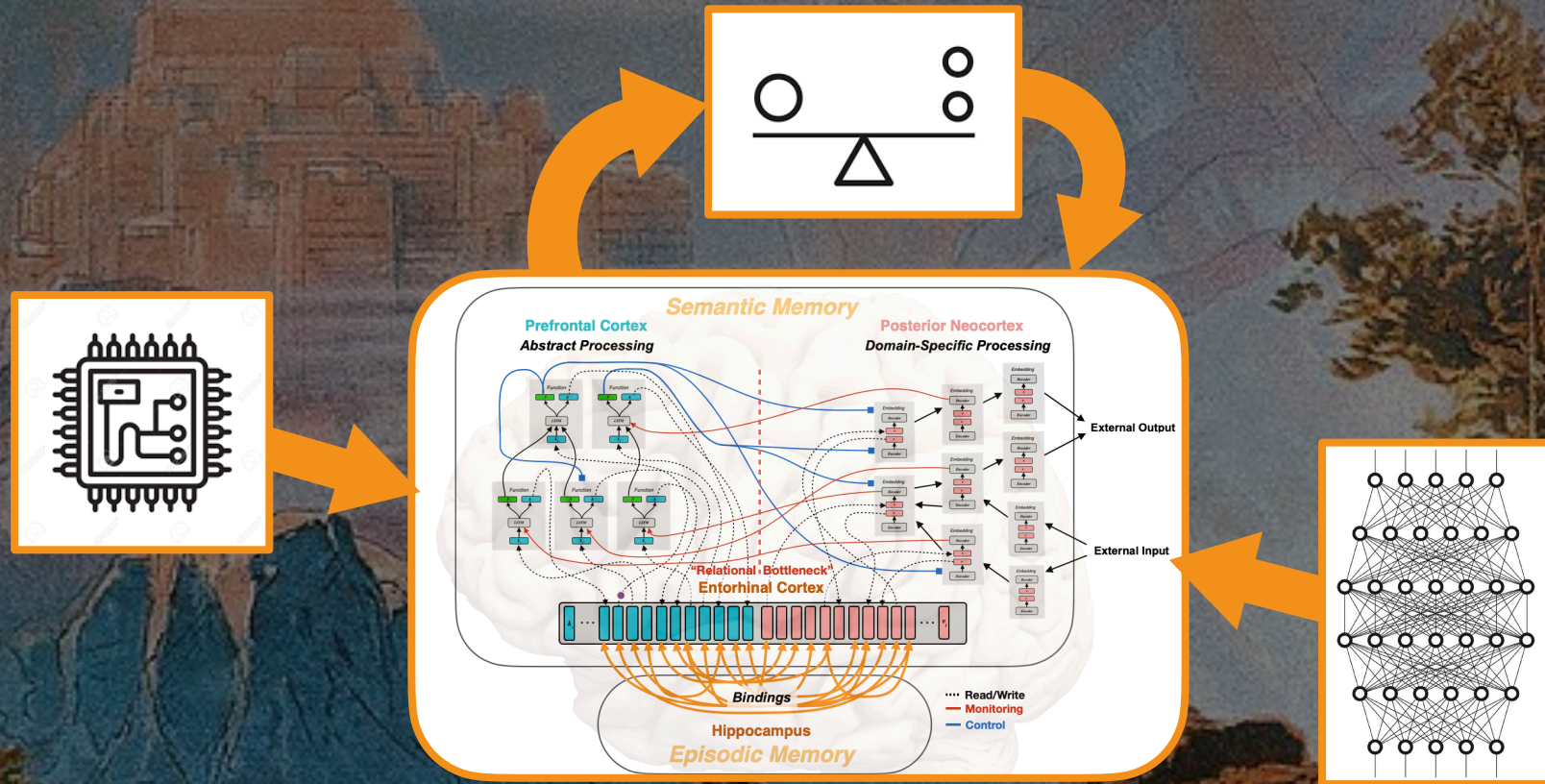
*Symbolic  
Computing*



*Neural  
Networks*

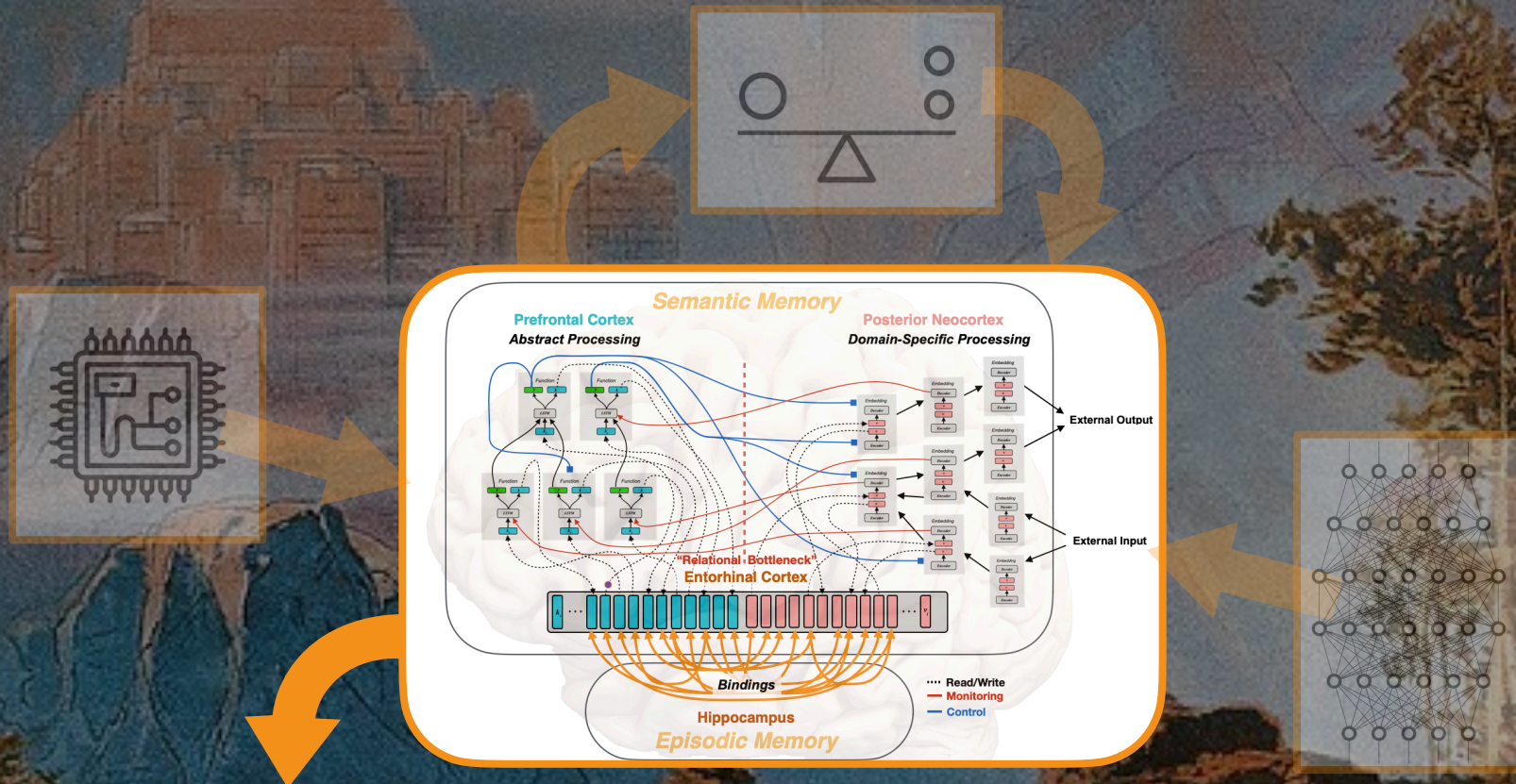


# Natural Intelligence





# Natural Intelligence



- **Computational neuropsychiatry**
- **Cognitively informed agent-based models**
- **[Explainable AI]**
- **Autonomous artificial agents**
- **More humane-machine interactions...**