Trends in Deep Learning for AI

Joe Marino

AI, the Brain, and Deep Learning

Intelligence

- reality is highly structured
  - when we measure the world, it is not independent/random

Intelligence

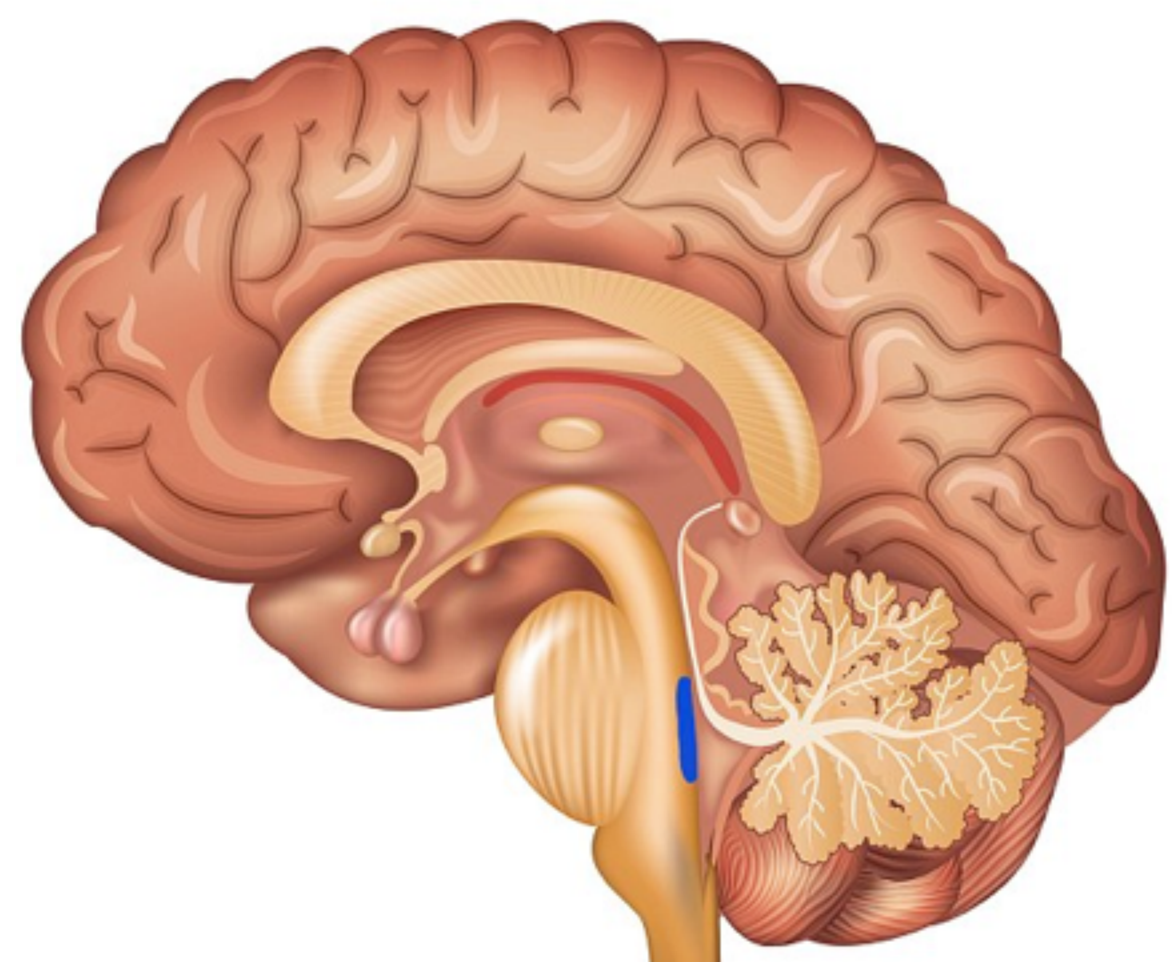
- reality is highly structured
  - when we measure the world, it is not independent/random
- an intelligent system can exploit reality's structure (to assist in survival)
- intelligence is a continuum

Biological Intelligence

- two main forms of intelligence
  - species level (evolution)
  - organism level (learning)

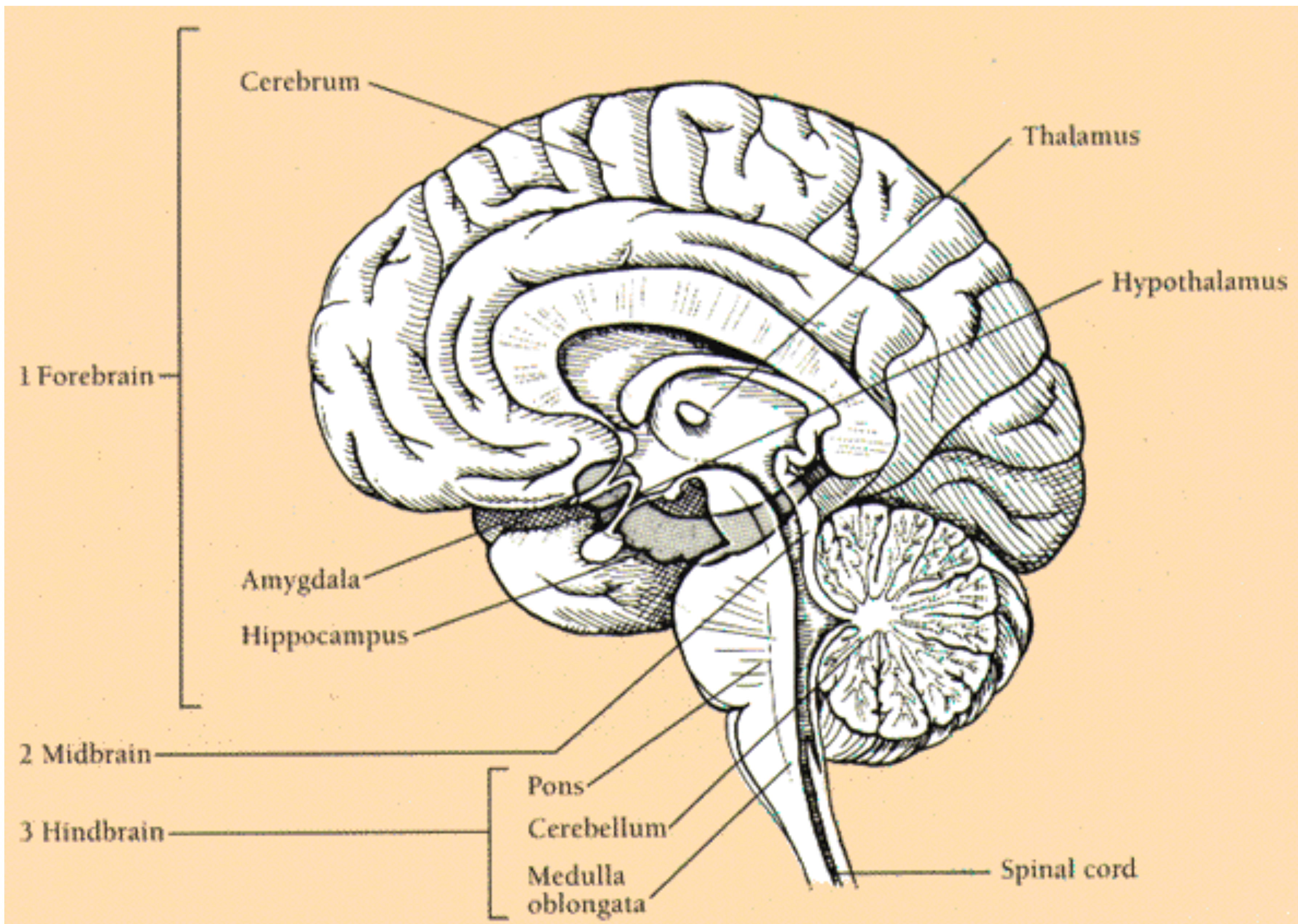
Human Intelligence

- intelligence resides in the nervous system
  - learning
  - memory
  - reasoning
  - planning
  - action
  - regulating body



Human Intelligence

- brain overview

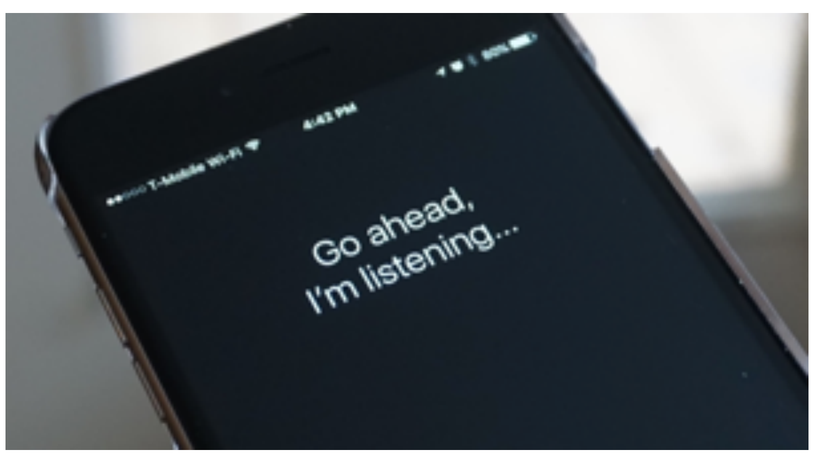


Human Intelligence

- we want to understand the learning mechanisms of the brain
  - structure
  - learning principles

Artificial (Machine) Intelligence

- why do we want to build intelligent systems?

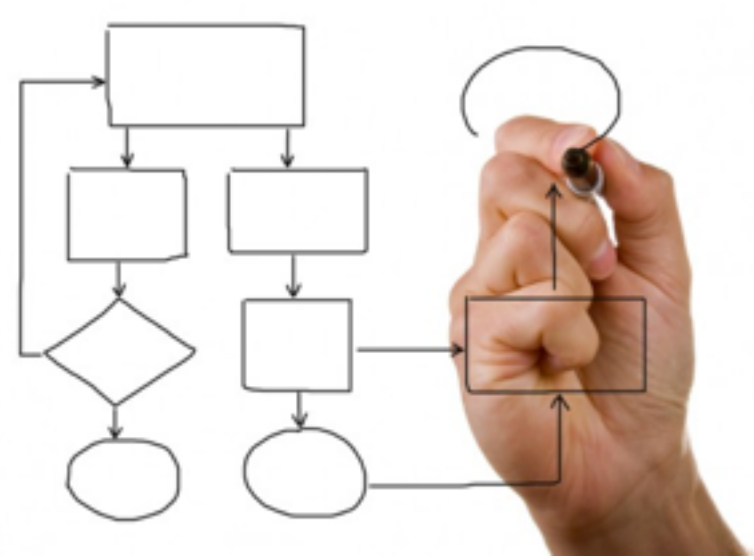


Artificial (Machine) Intelligence

- what capabilities should these systems have?
    - take in inputs from their environment
    - process these inputs
    - interact with their environment

Artificial (Machine) Intelligence

- what capabilities should these systems have?
  - take in inputs from their environment
  - process these inputs in some form
  - interact with their environment



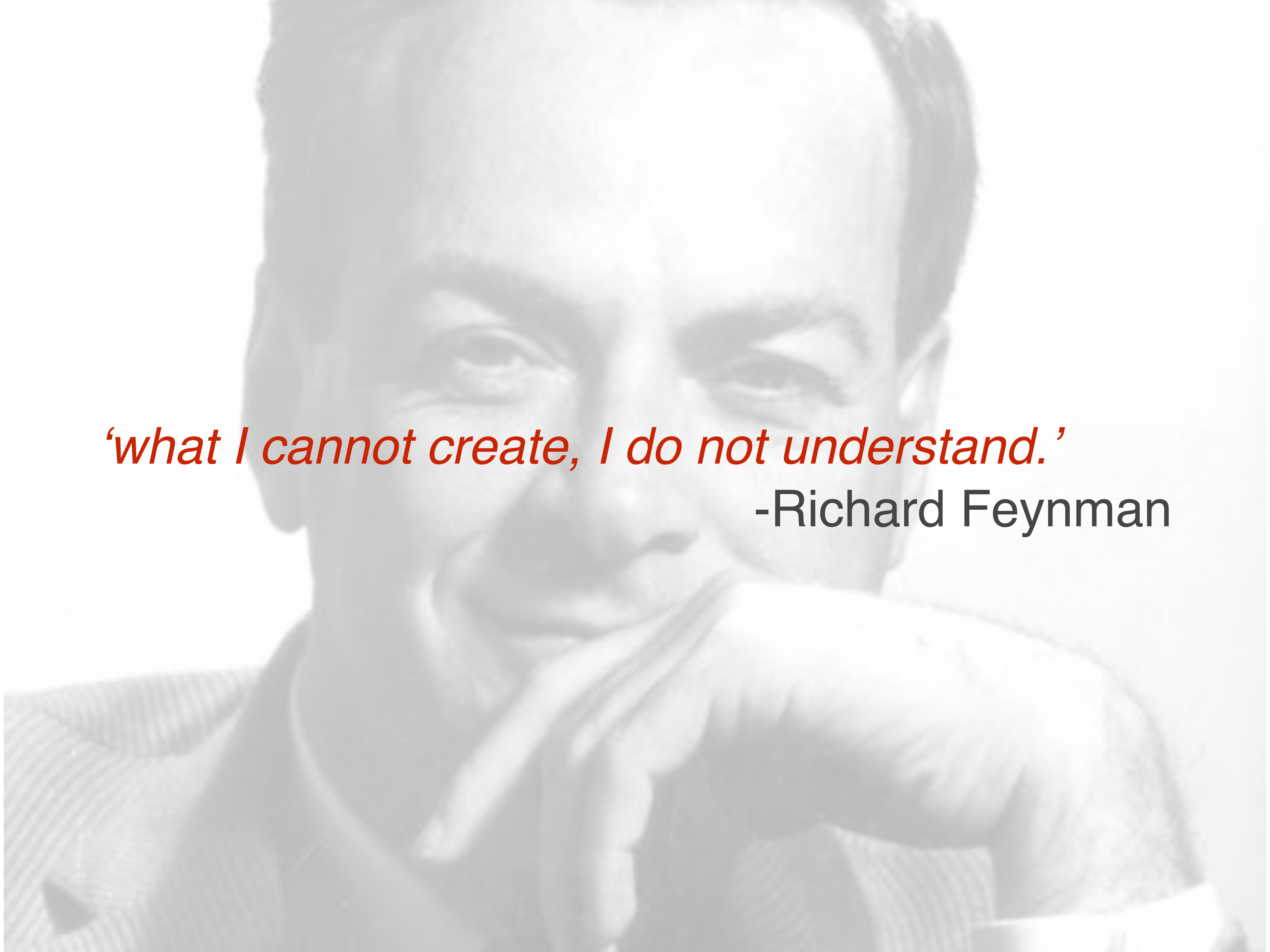
Artificial (Machine) Intelligence

- do we want to create an machine version of the human brain?
  - **no**
- intelligent machine systems will likely use some of the same underlying learning principles as biological intelligence, but they will be drastically different entities

Deep Learning

- decompose a signal into many layers
- works well for vision, audio, some other domains
- is it the answer to artificial intelligence?

Generative Models



‘what I cannot create, I do not understand.’

-Richard Feynman

Generative Models

- generative models *generate* data, using some form of density estimation
- by modeling how the data is generated, we attempt to understand the underlying mechanisms
- three broad categories
  - fully-observed models
  - transformation models
  - latent variable models

Generative Models

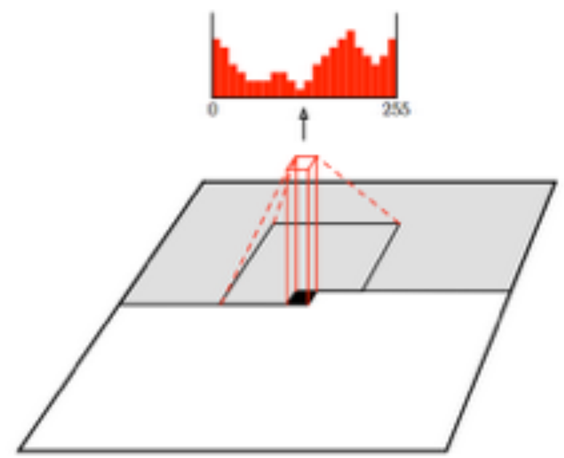
- **why** generative models?
  - move beyond associating inputs to outputs
  - understand and imagine how the world evolves
  - recognize objects in the world and their factors of variation
  - detect surprising events in the world
  - establish concepts as useful for reasoning and decision making
  - imagine and generate rich plans for the future

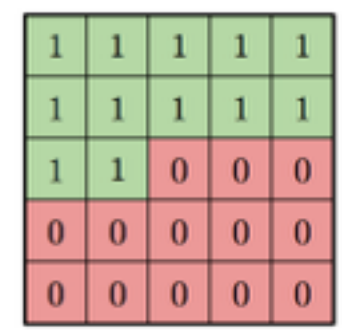
Generative Model Applications

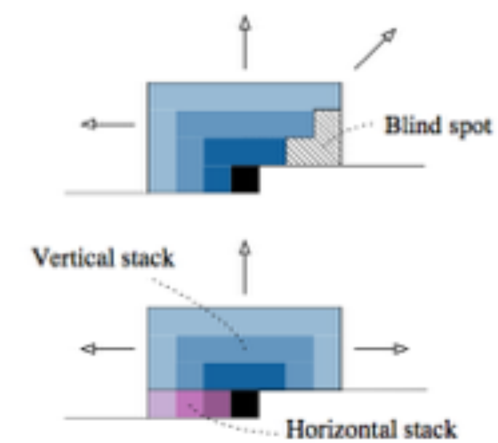
- compression & communication
- environment simulation
- one-shot generalization
- macro-actions and planning
- density-based exploration
- missing data imputation
- scene understanding
- representation learning
- visual concept learning
- semi-supervised classification
- 3D scene generation

Fully-Observed Models

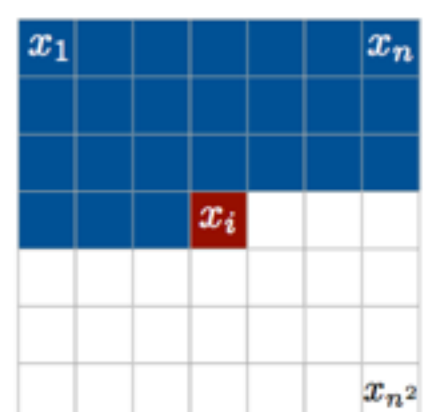
- Pixel RNN/CNN
  - auto-regressive models
  - generate an image pixel by pixel by conditioning on previous pixels



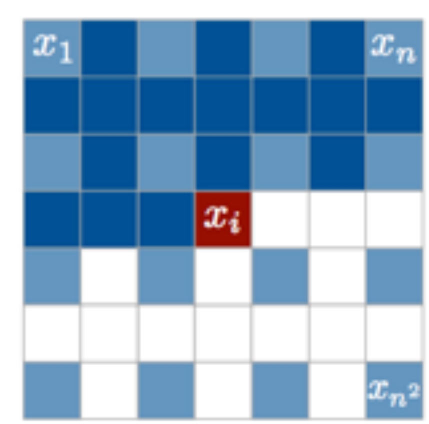




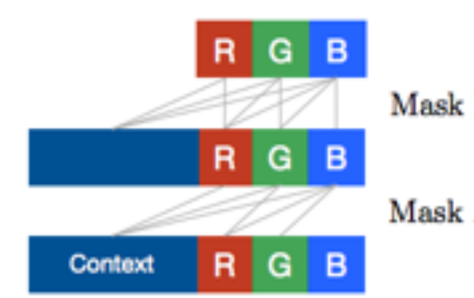
van den Oord et al., 2016



Context



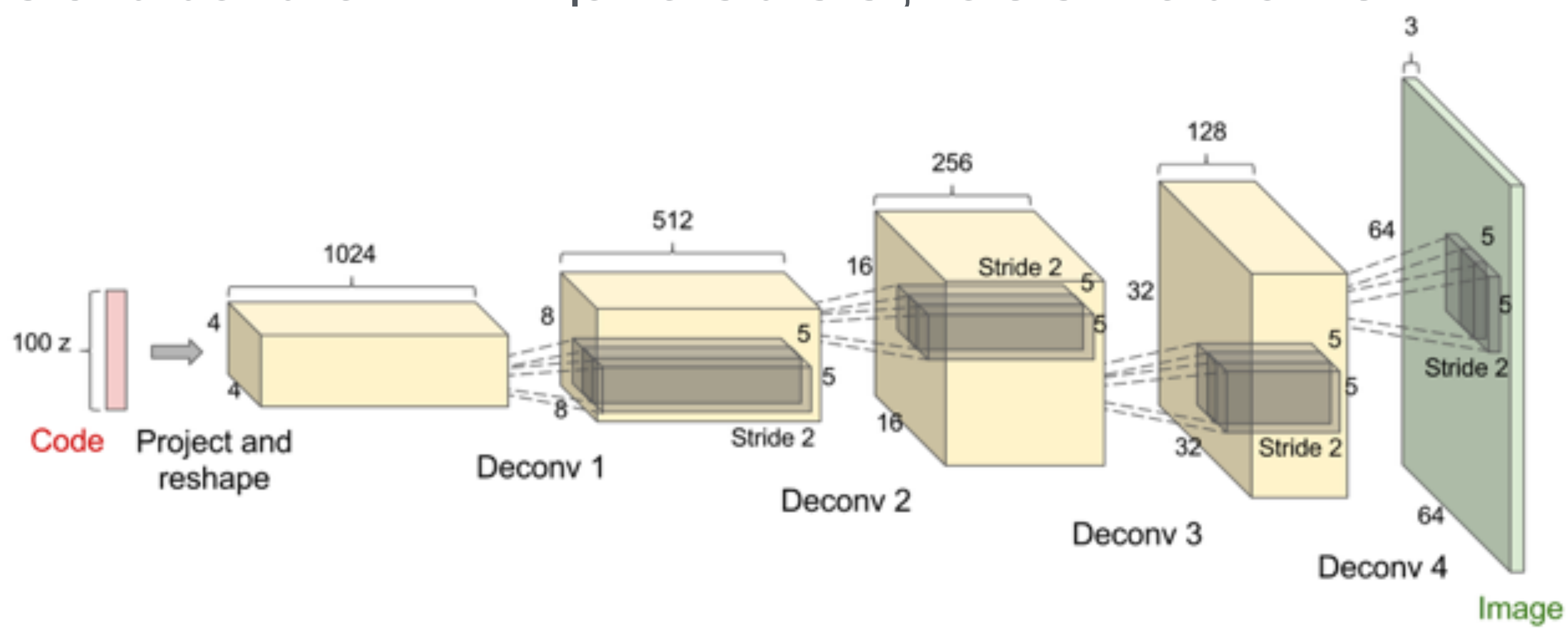
Multi-scale context



van den Oord et al., 2016

Transformation Models

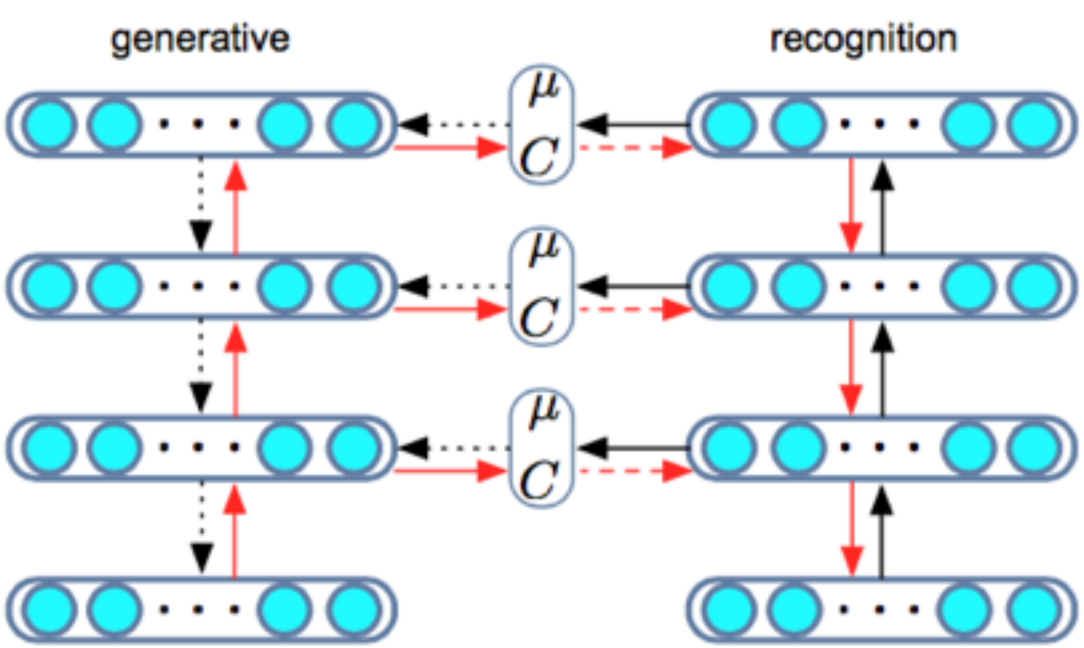
- generative adversarial networks (GANs)
  - train two competing models, one that generates images and another that determines whether they are fake or real
  - difficult to train in practice, oscillations

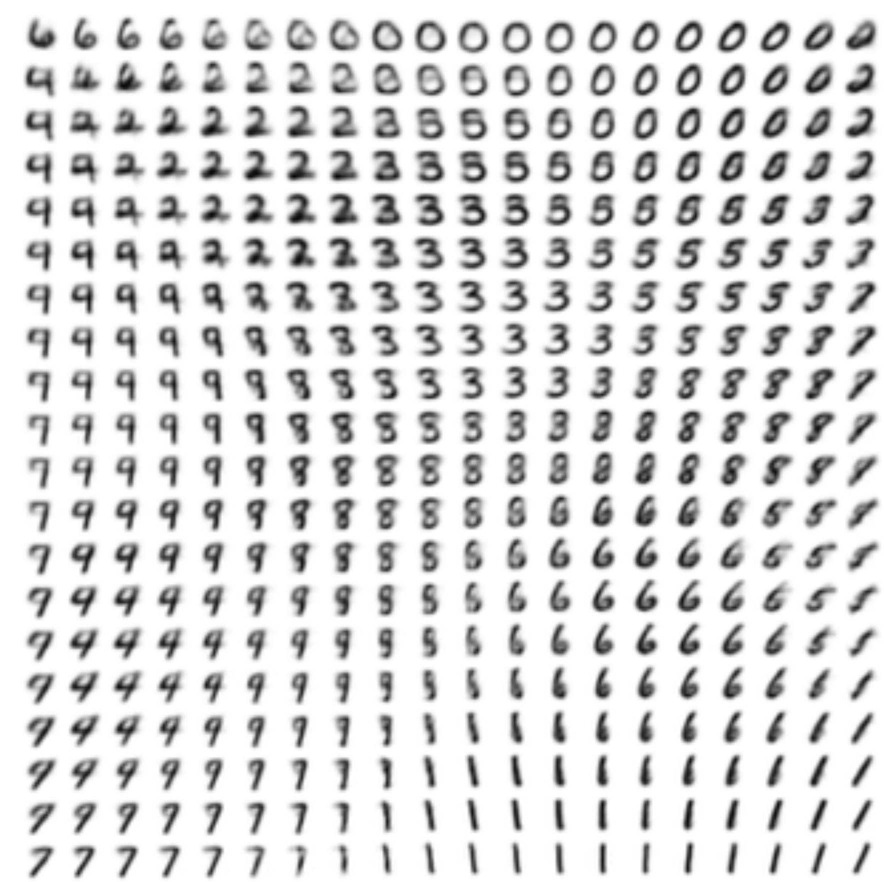


Radford et al., 2016

Latent Variable Models

- variational auto-encoders (VAEs)
    - introduce unobserved local random variables that represent hidden causes



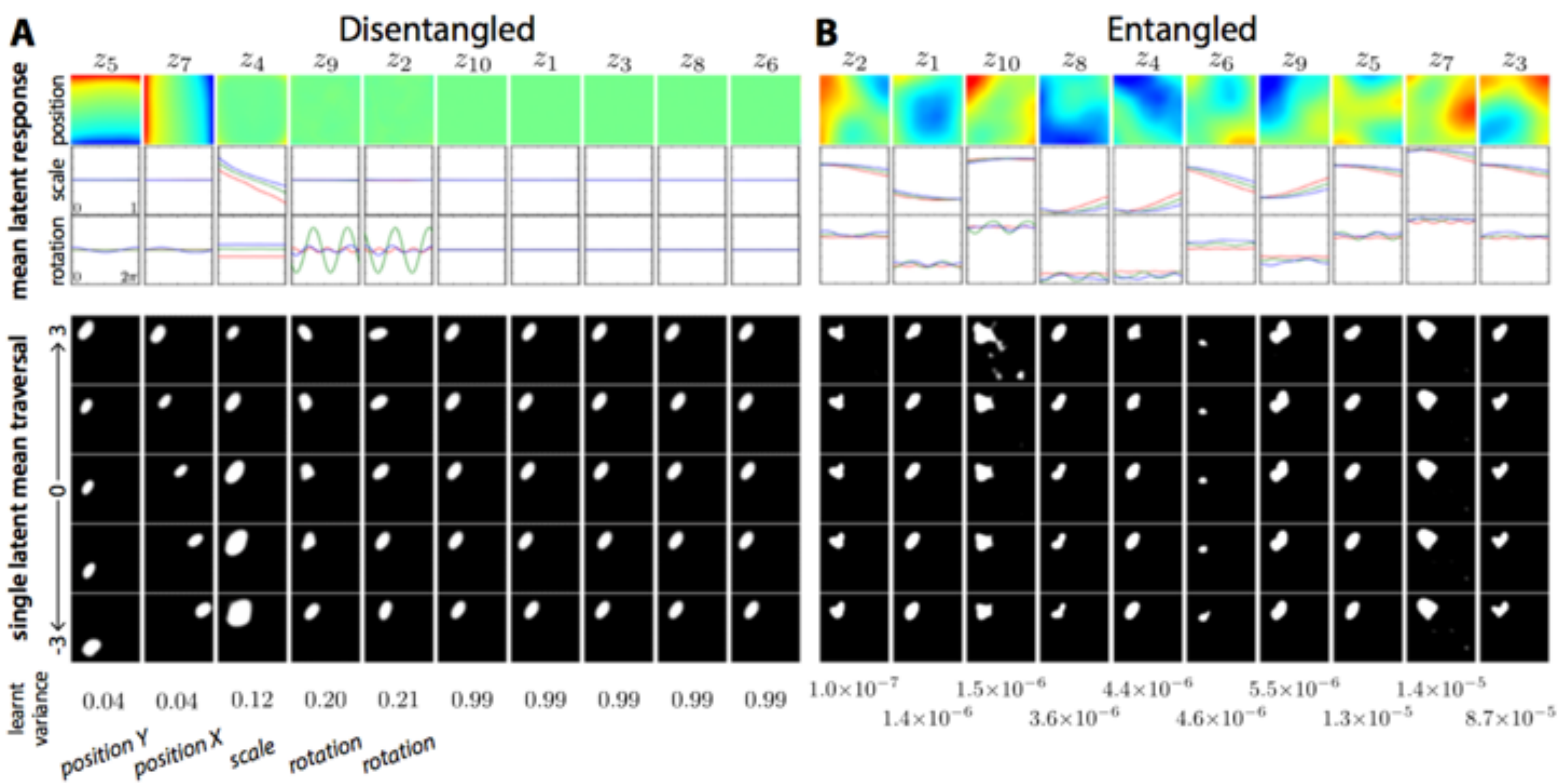


Rezende et al., 2014

Kingma et al., 2014

Latent Variable Models

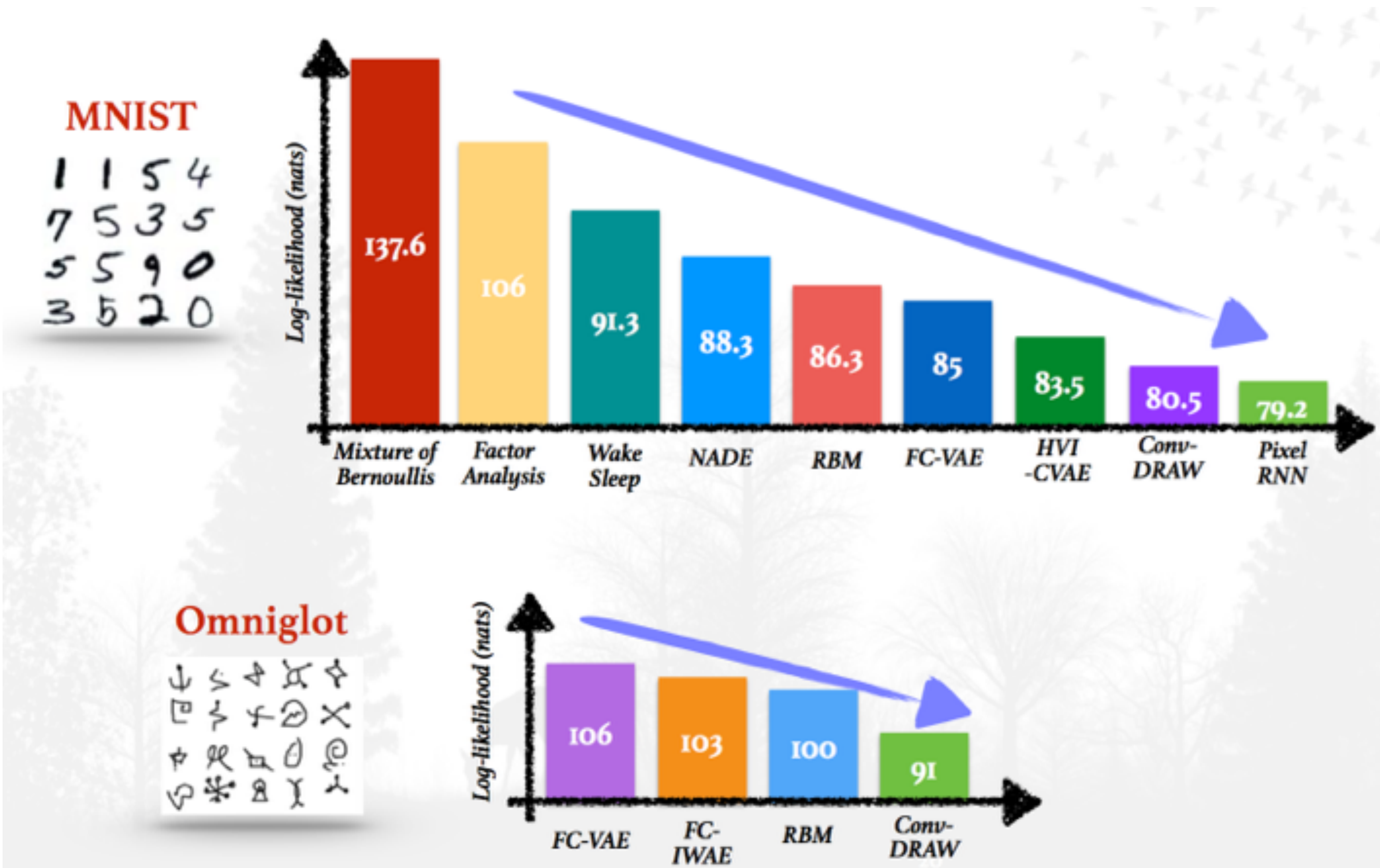
- disentangling factors of variation
  - regularization weight encourages disentangled latent representation



Higgins et al., 2016

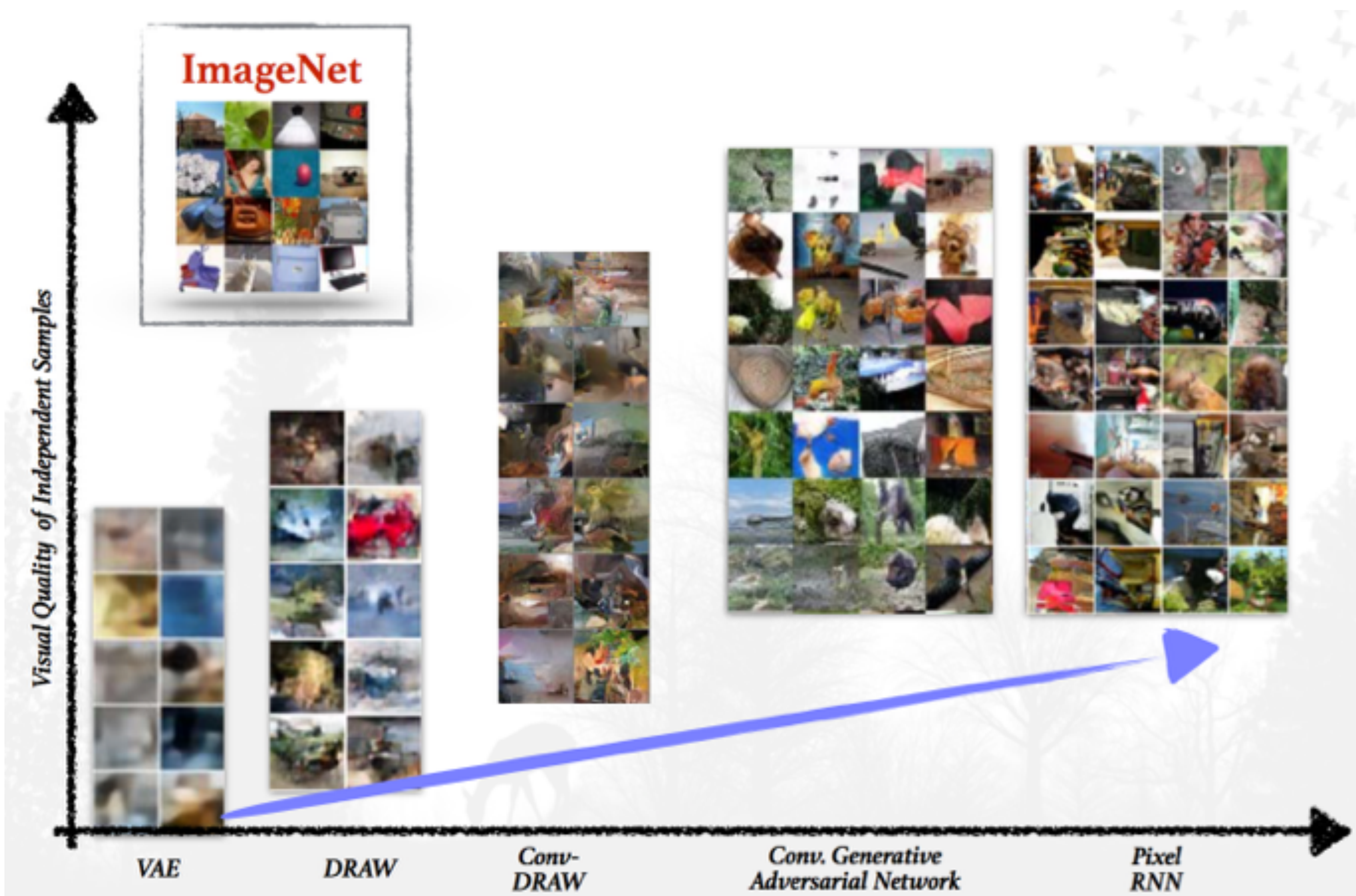
Performance

- generative models are typically evaluated by log likelihood and quality of samples



Performance

- generative models are typically evaluated by log likelihood and quality of samples



Performance

- generative models are typically evaluated by log likelihood and quality of samples
- in high dimensional data, these two performance measures are largely independent
  - a model can generate great samples but have poor LL and vice versa

Theis et al., 2016

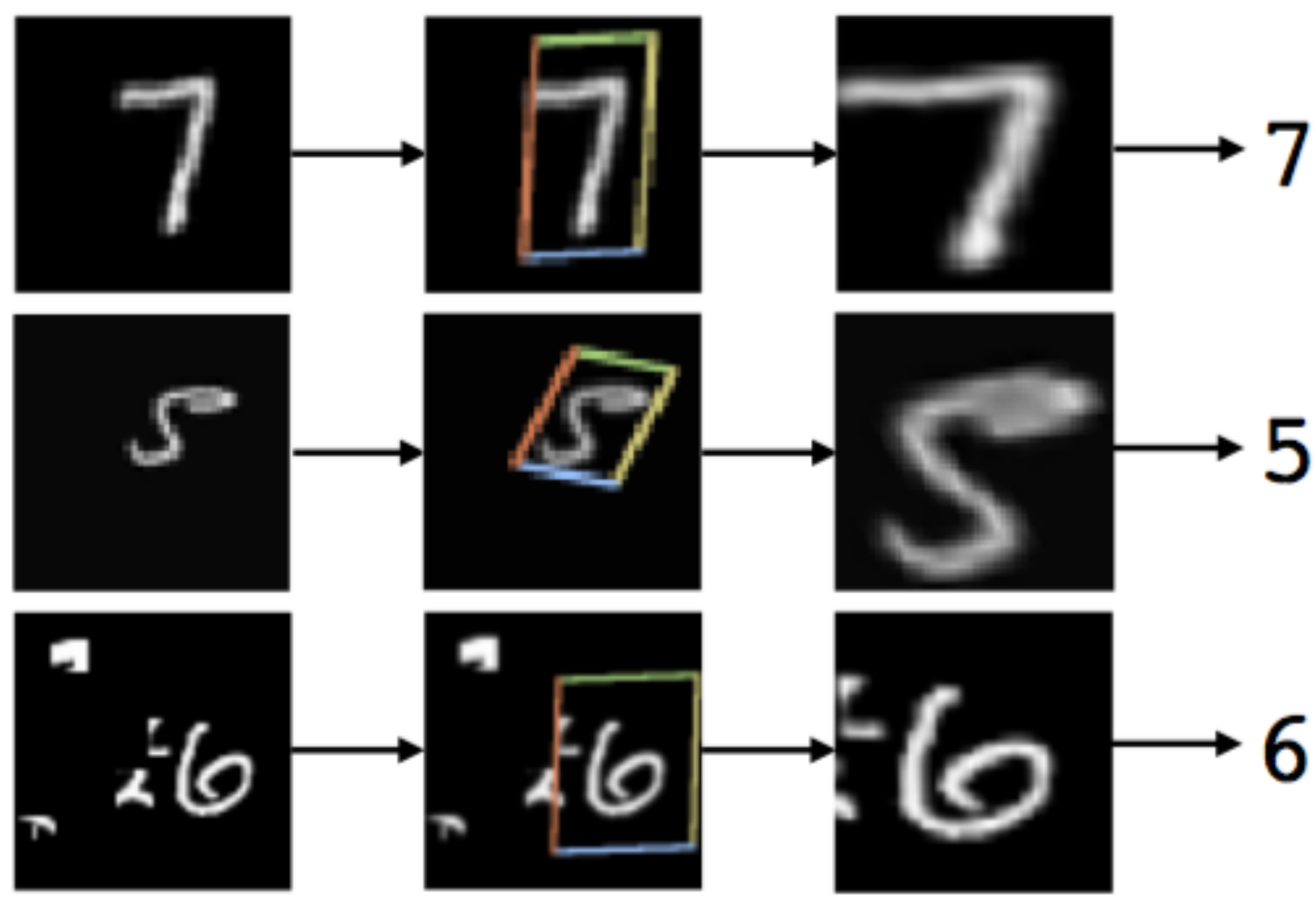
Attention, Memory, Reasoning, and Planning

Attention

- focus on relevant parts of signal, memory to assist in some task

Attention

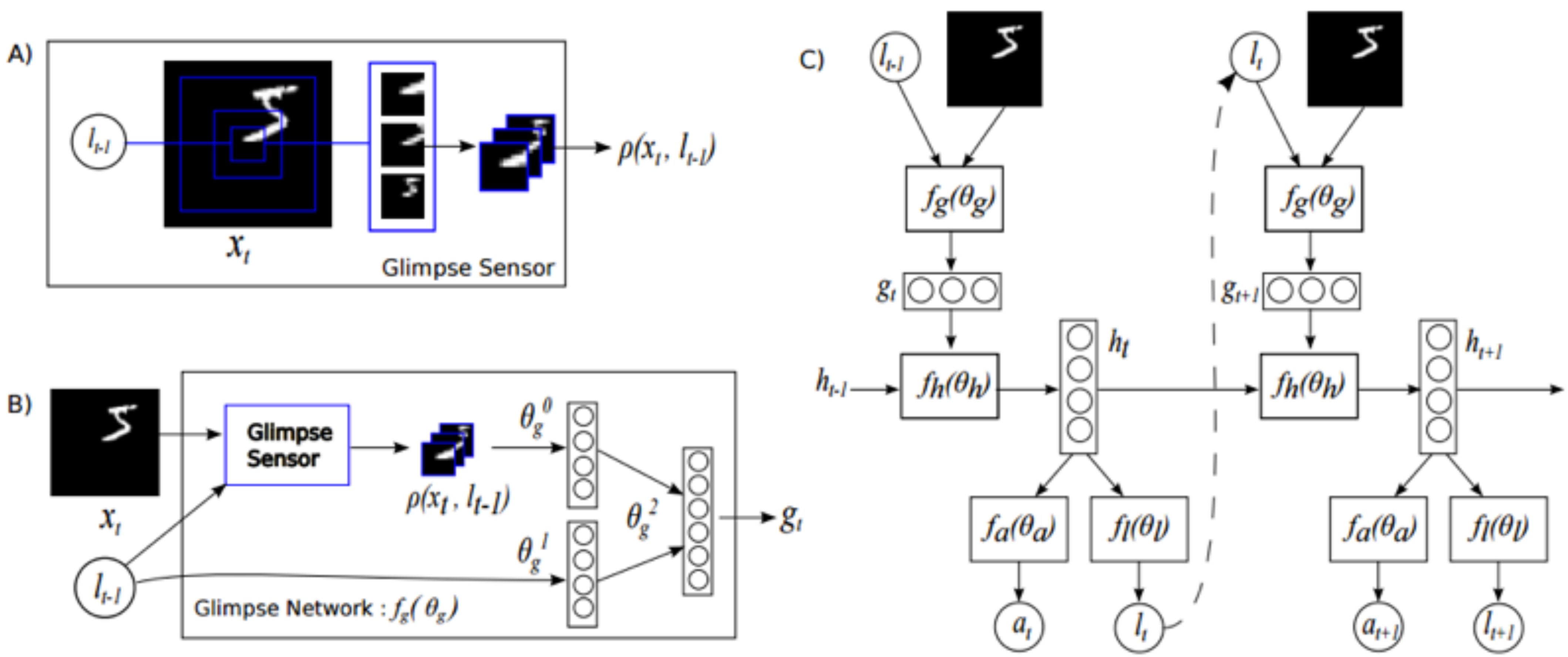
- visual attention
  - spatial transformer



Jaderberg et al., 2015

Attention

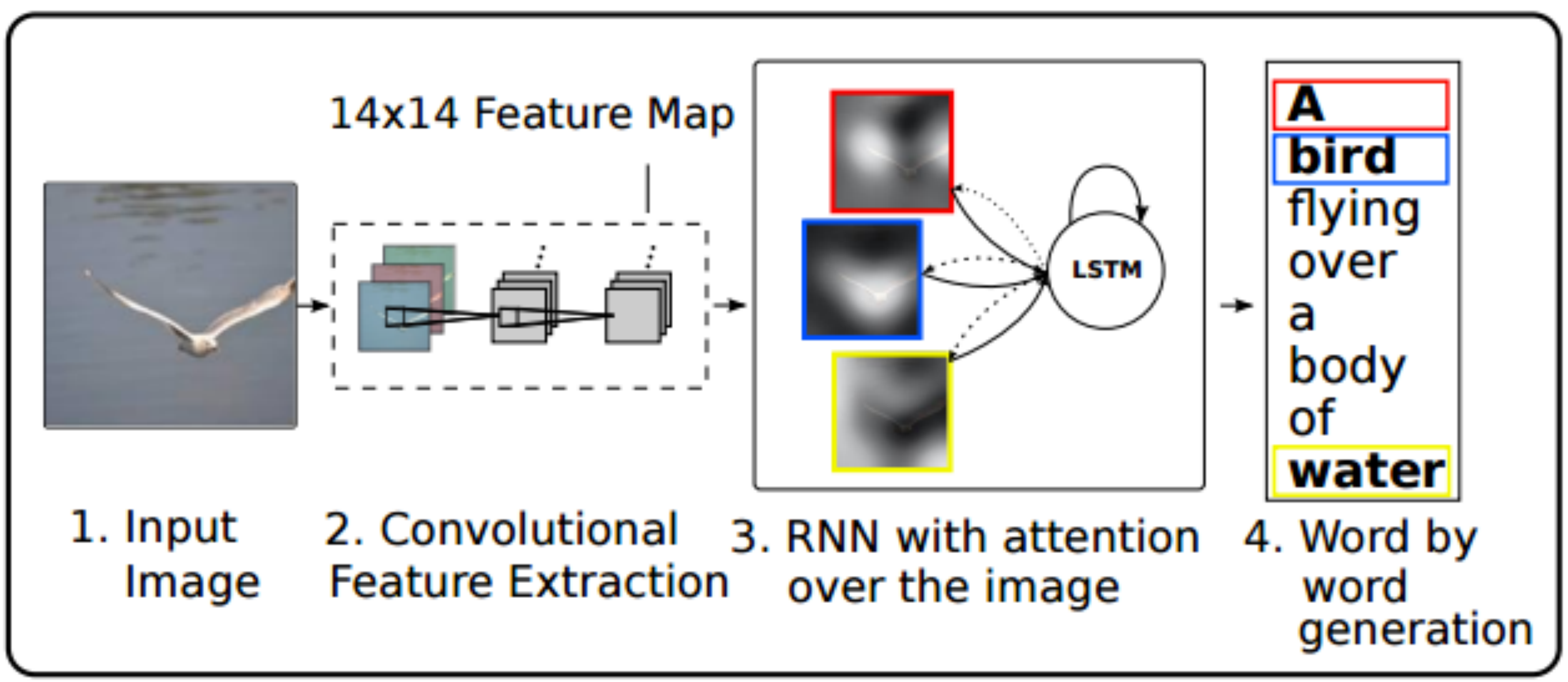
- visual attention
  - recurrent models



Mnih et al., 2015

Attention

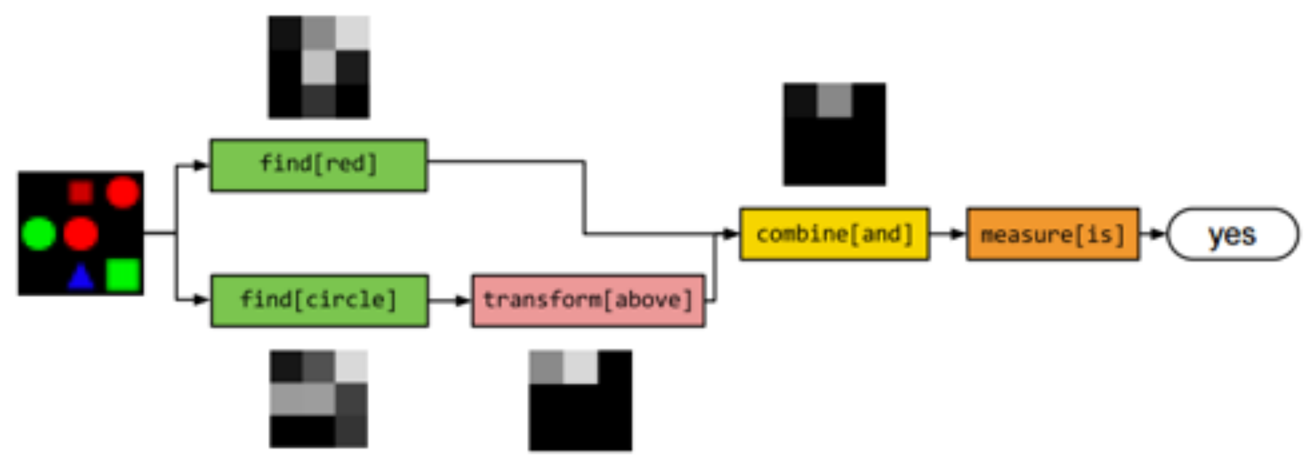
- visual attention
  - recurrent models



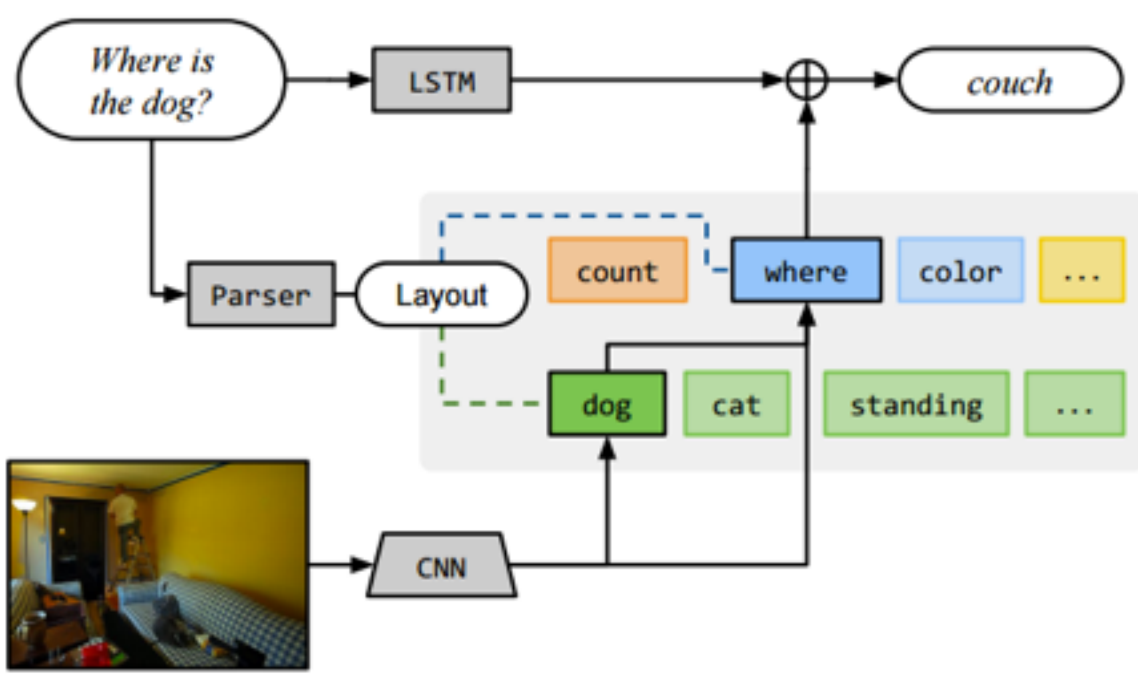
Xu et al., 2015

Attention, Reasoning

- visual attention,reasoning
  - neural module networks



is there a red shape above a circle?



Andreas et al., 2016

Memory

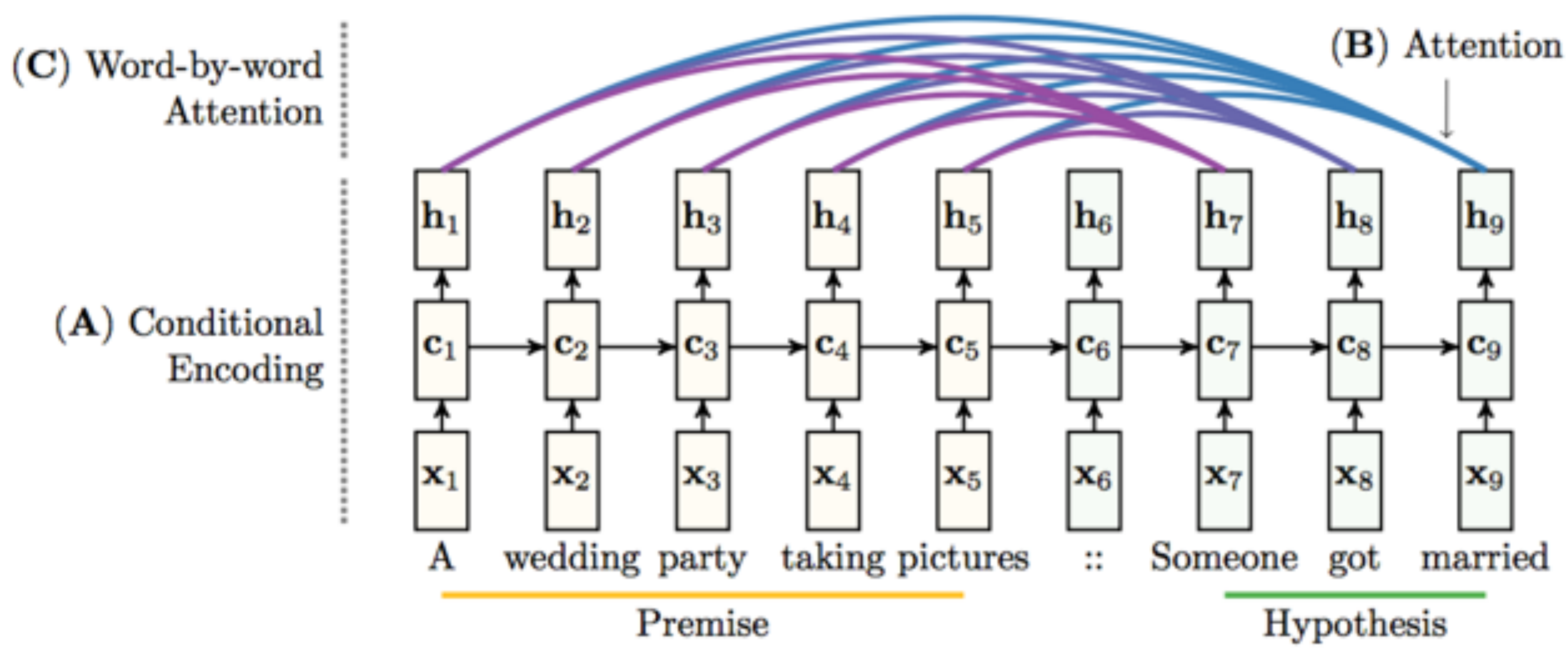
- need some method of storing and accessing information

Memory

- recurrent models
  - LSTMs, GRUs prevent gradients from vanishing/exploding, allowing for longer dependencies
  - good for learning sequences

Attention, Memory

- use two LSTMs, one for premise, one for hypothesis



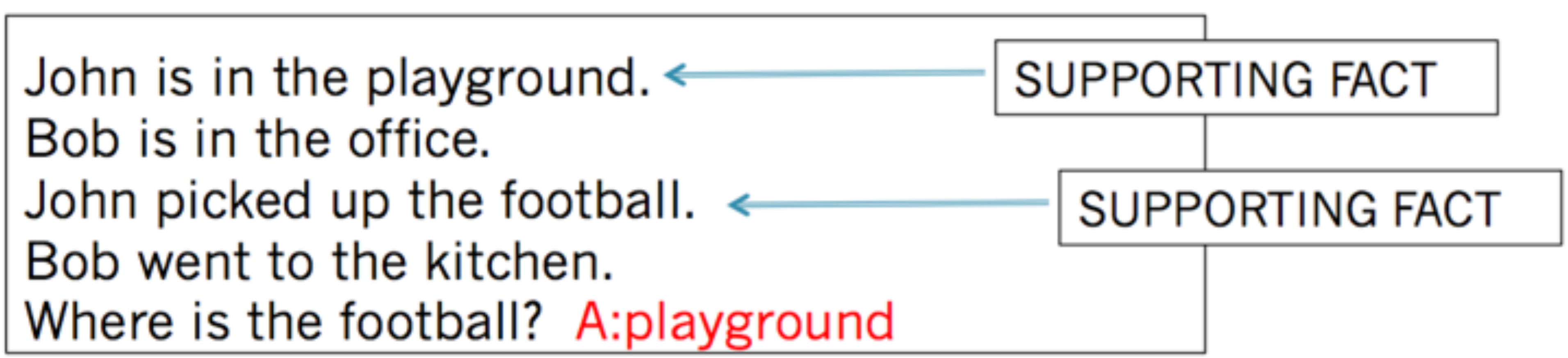
Rocktäschel et al., 2015

Memory

- recurrent models
  - LSTMs, GRUs prevent gradients from vanishing/exploding, allowing for longer dependencies
  - good for learning sequences
  - not highly scalable
  - short-term memory, working memory

Attention, Memory, Reasoning

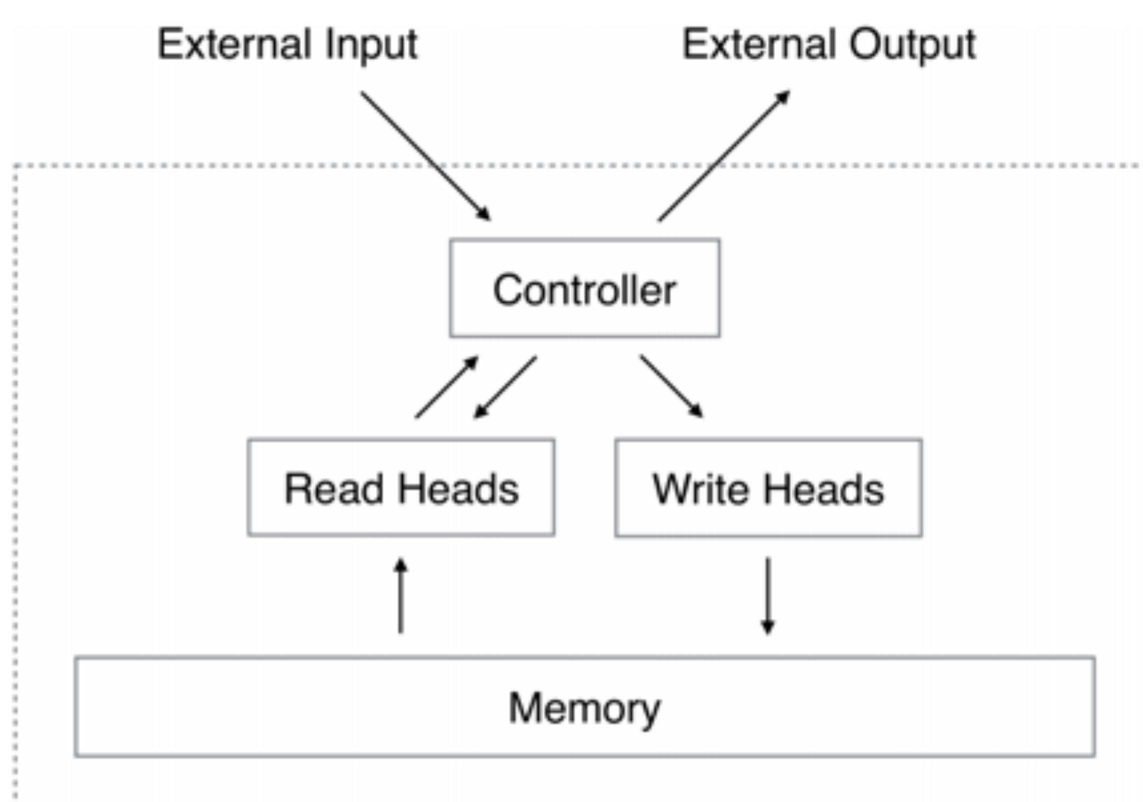
- bAbI dataset
  - 20 tasks
  - language skills: conjunction, negation, etc.
  - reasoning skills: counting, path finding, etc.



Weston et al., 2015

Attention, Memory, Reasoning

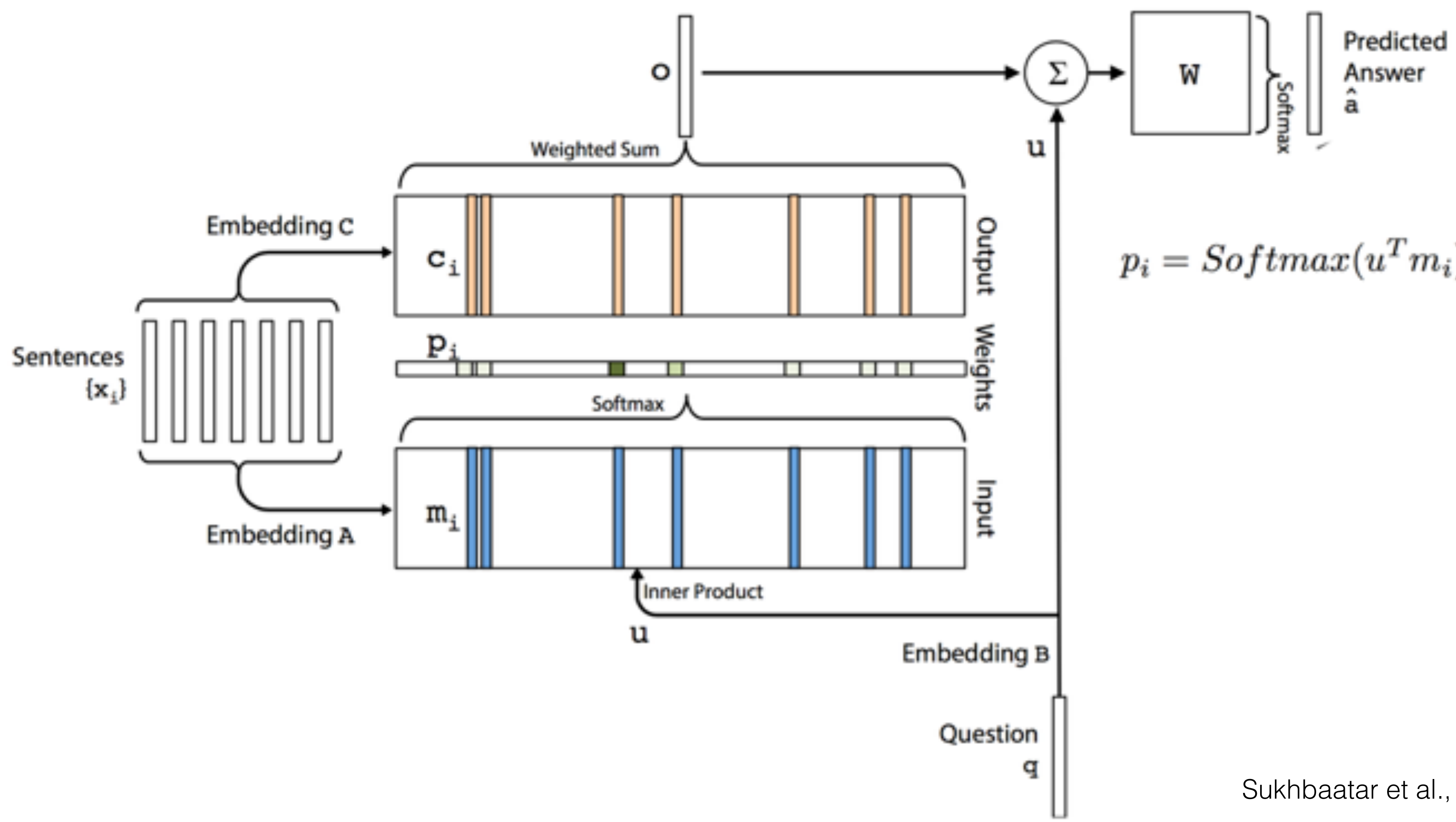
- memory bank/controller models
  - controller takes in input, reads/writes from memory bank, produces output
- neural turing machine



Graves et al., 2014

Attention, Memory, Reasoning

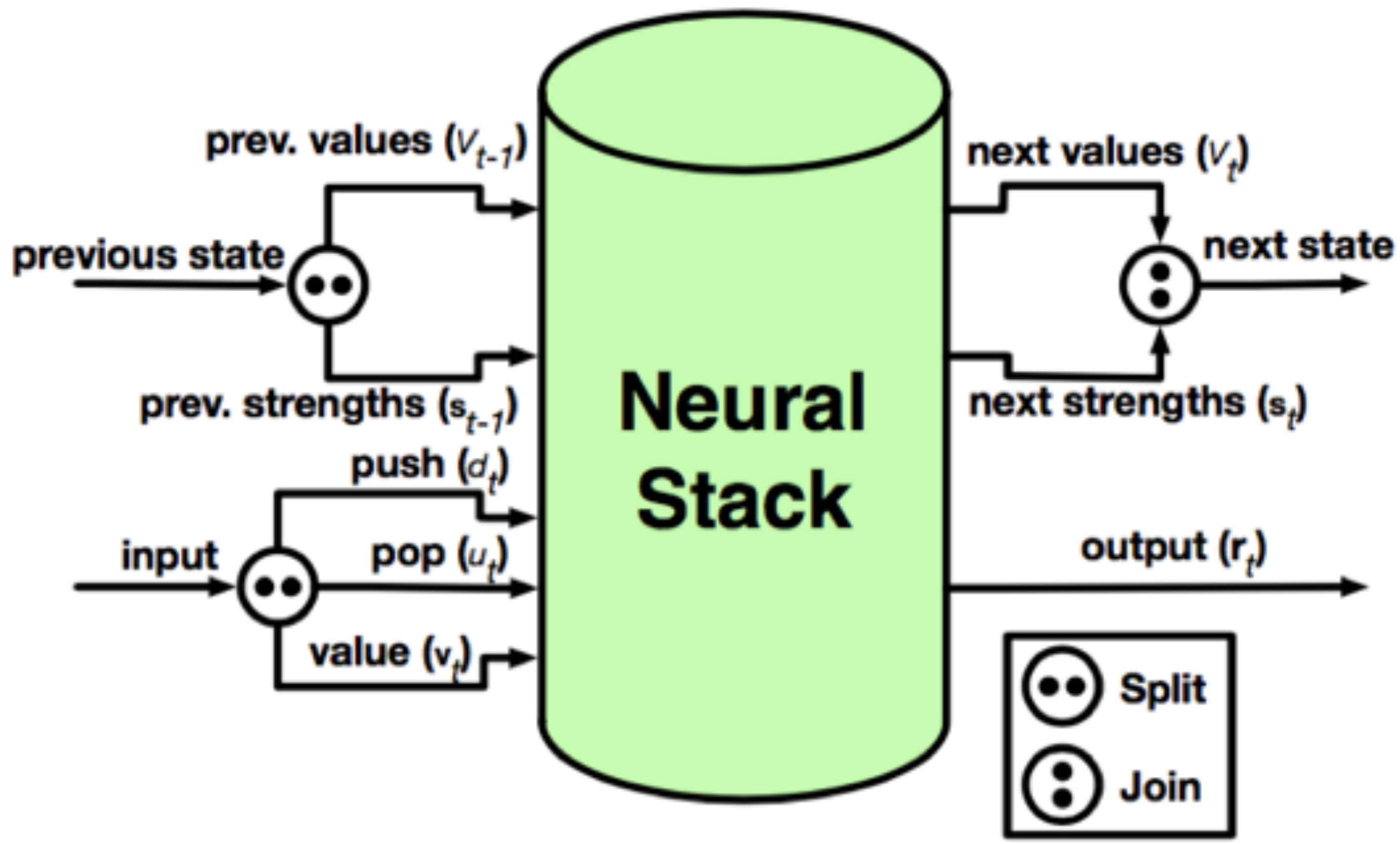
- memory bank/controller models
    - controller takes in input, reads/writes from memory bank, produces output
- MemNN



Sukhbaatar et al., 2015

Attention, Memory, Reasoning

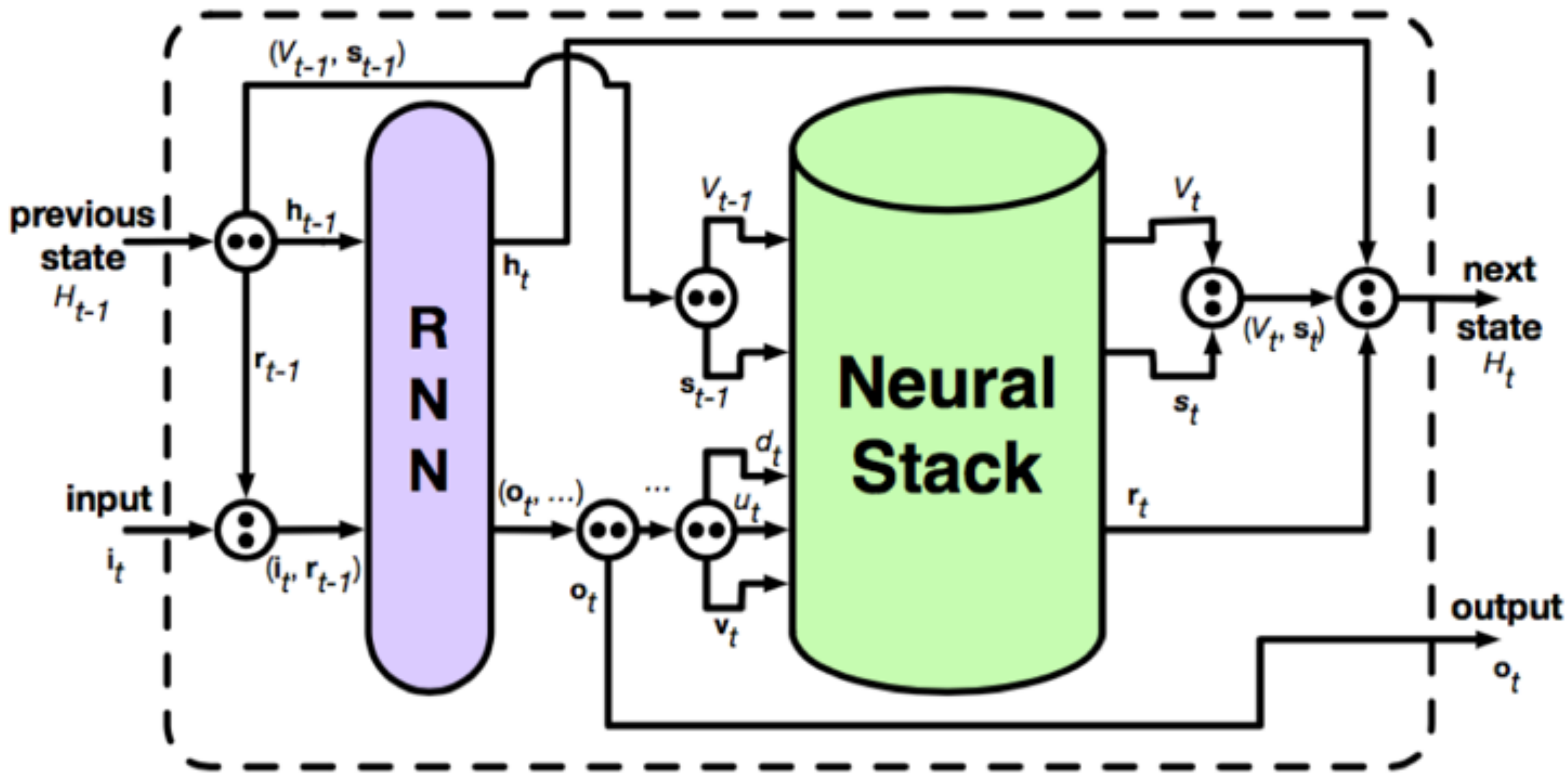
- augmented RNNs
  - neural stack



Grefenstette et al., 2015

Attention, Memory, Reasoning

- augmented RNNs
  - neural stack with RNN controller



Grefenstette et al., 2015

Planning, Acting

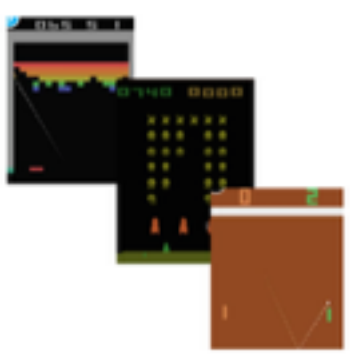
- want to plan, carry out actions in environment
  - reinforcement learning

Kohl and Stone, 2004

Ng et al, 2004

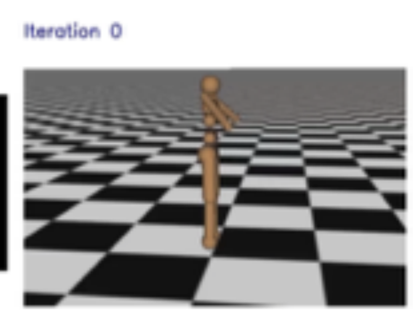
Tedrake et al, 2005

Kober and Peters, 2009

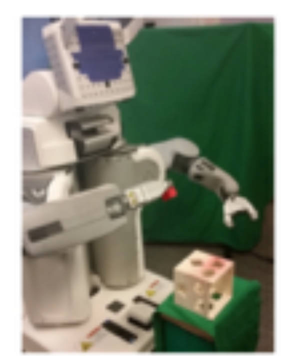


Schulman et al, 2015 (TRPO)
Mnih et al, 2015 (A3C)

Silver et al, 2014 (DPG)
Lillicrap et al, 2015 (DDPG)



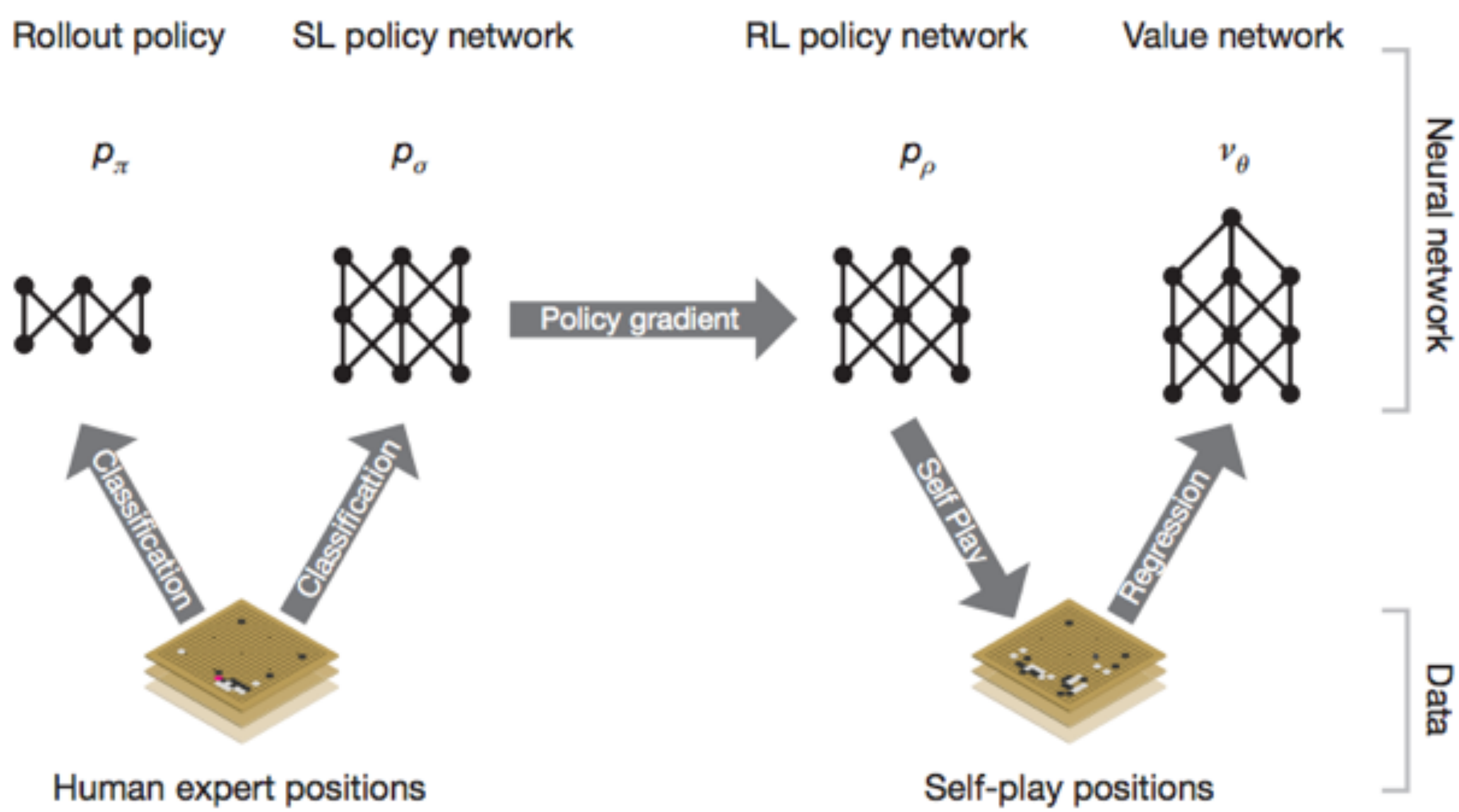
Schulman et al, 2016 (TRPO + GAE)



Levine*, Finn*, et al, 2016 (GPS)

Planning, Acting

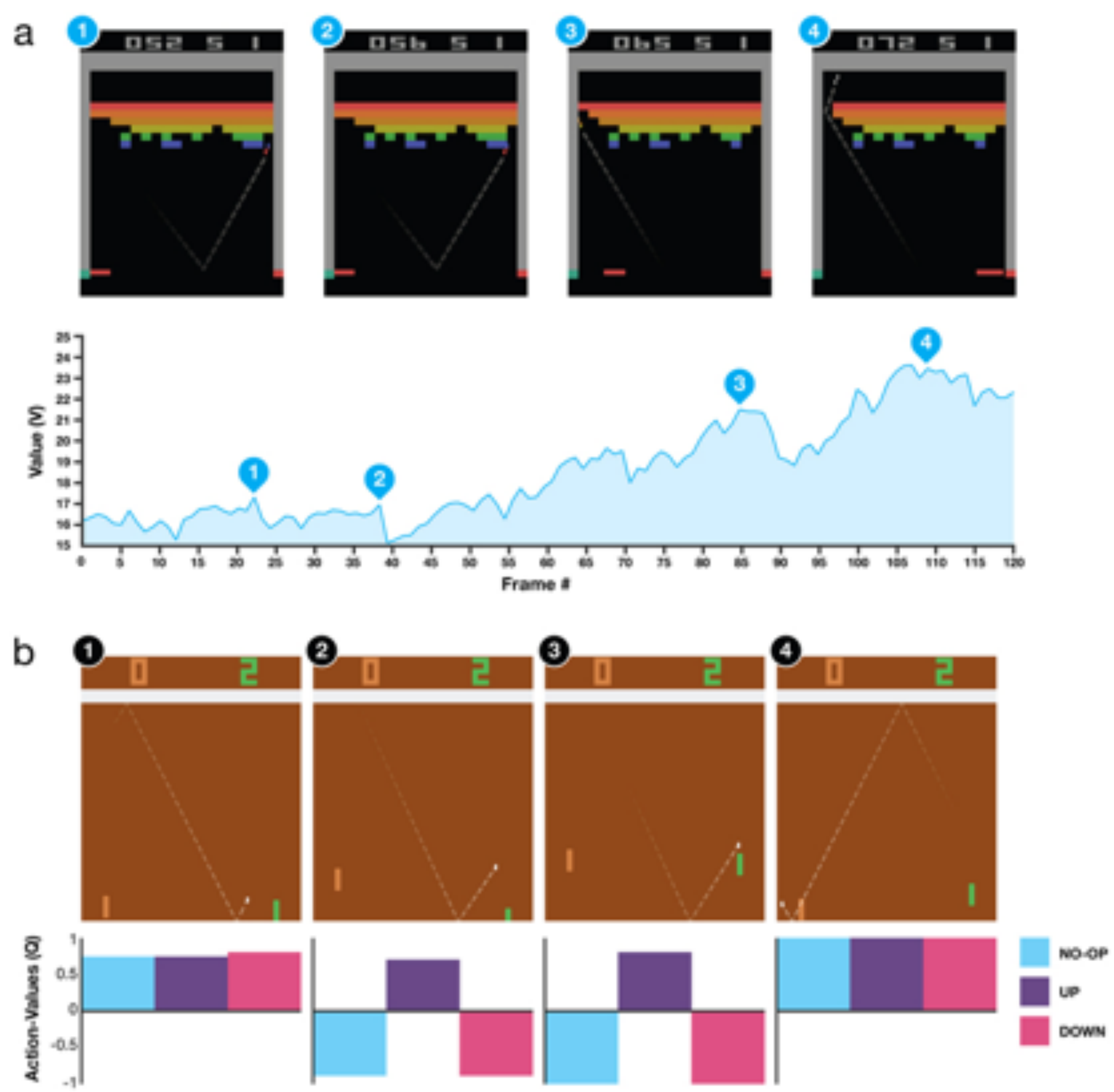
- want to plan, carry out actions in environment
  - reinforcement learning
  - AlphaGo



Silver et al., 2016

Planning, Acting

- want to plan, carry out actions in environment
  - reinforcement learning
  - atari



Mnih et al., 2015

How does it all fit together?

Artificial (Machine) Intelligence

- take in and process inputs
    - cameras, microphones, etc.

- **internal processing**
    - pattern recognition, memory, reasoning, planning, etc.

- interact with the environment
    - robots, speakers, screens, etc.

Artificial (Machine) Intelligence

- if we want to combine all of the internal processing functions,
  - what will this system look like?
  - how will it learn?

The Brain

- there are different mechanisms for constructing neural circuits
  - genetic
  - random
  - **learning**
- need some method of performing *credit assignment* for learning
  - simple Hebbian learning rules are not sufficient

Marblestone et al., 2016

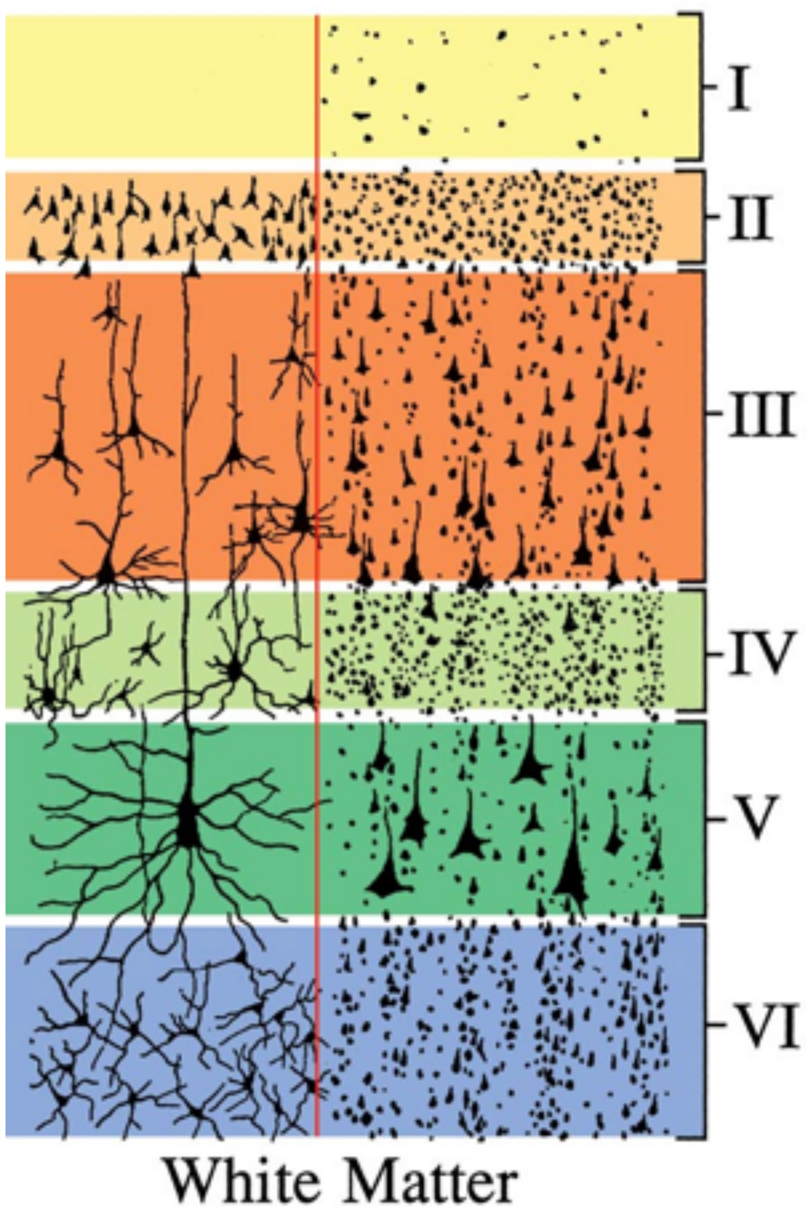
The Brain

- the brain contains many different structures that handle different functions
  - optimizing different objective functions

Marblestone et al., 2016

Sensory Cortex

- layered architecture
- hierarchical structure
- feedforward and feedback connections

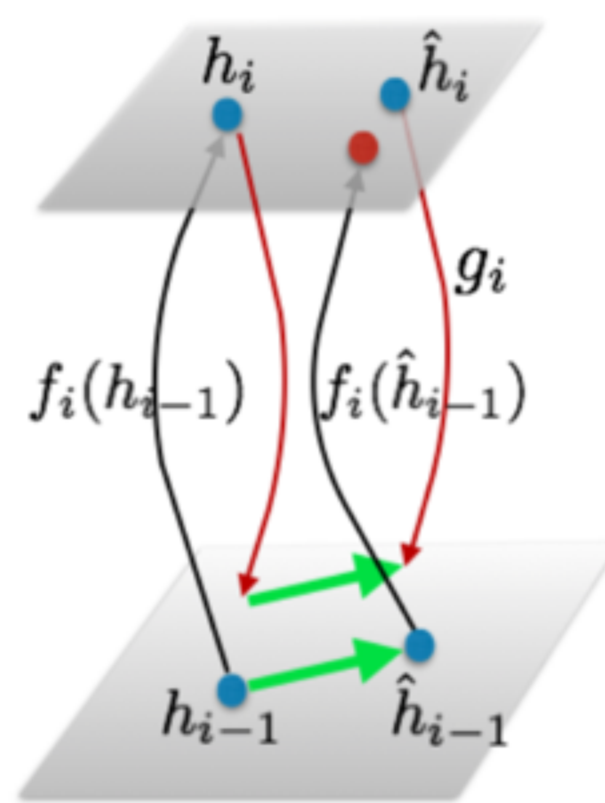


Sensory Cortex

- hierarchical processing
- builds a generative model of the world
- constantly making predictions

Target Propagation

- auto-encoder
- feedforward network processes sensory stimuli
- feedback network generates predictions (targets)
- learning occurs through comparing feedforward and feedback activations



Bengio, 2014

Resources

- Deep Learning Summer School
    - slides: https://sites.google.com/site/deeplearningsummerschool2016/home
    - video: http://videolectures.net/deeplearning2016_montreal/
- OpenAI Blog: https://openai.com/blog/
- Google Research Blog: https://research.googleblog.com/
- Facebook Research Blog: https://research.facebook.com/blog/

References

Andreas, Jacob, et al. "Neural module networks." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016.

Weston, Jason, et al. "Towards ai-complete question answering: A set of prerequisite toy tasks." arXiv preprint arXiv:1502.05698 (2015).

Sukhbaatar, Sainbayar, Jason Weston, and Rob Fergus. "End-to-end memory networks." Advances in neural information processing systems. 2015.

Graves, Alex, Greg Wayne, and Ivo Danihelka. "Neural turing machines." arXiv preprint arXiv:1410.5401 (2014).

Xu, Kelvin, et al. "Show, attend and tell: Neural image caption generation with visual attention." arXiv preprint arXiv:1502.03044 2.3 (2015): 5.

van den Oord, Aaron, Nal Kalchbrenner, and Koray Kavukcuoglu. "Pixel Recurrent Neural Networks." arXiv preprint arXiv:1601.06759 (2016).

Radford, Alec, Luke Metz, and Soumith Chintala. "Unsupervised representation learning with deep convolutional generative adversarial networks." arXiv preprint arXiv:1511.06434 (2015).

Lee, Dong-Hyun, et al. "Difference target propagation." Joint European Conference on Machine Learning and Knowledge Discovery in Databases. Springer International Publishing, 2015.

Theis, Lucas, Aäron van den Oord, and Matthias Bethge. "A note on the evaluation of generative models." arXiv preprint arXiv:1511.01844 (2015).

Higgins, Irina, et al. "Early Visual Concept Learning with Unsupervised Deep Learning." arXiv preprint arXiv:1606.05579 (2016).

Marblestone, Adam, Greg Wayne, and Konrad Kording. "Towards an integration of deep learning and neuroscience." arXiv preprint arXiv:1606.03813 (2016).

Grefenstette, Edward, et al. "Learning to transduce with unbounded memory." Advances in Neural Information Processing Systems. 2015.

Rocktäschel, Tim, et al. "Reasoning about entailment with neural attention." arXiv preprint arXiv:1509.06664 (2015).

Jaderberg, Max, Karen Simonyan, and Andrew Zisserman. "Spatial transformer networks." Advances in Neural Information Processing Systems. 2015.

Kingma, Diederik P., and Max Welling. "Auto-encoding variational bayes." arXiv preprint arXiv:1312.6114 (2013).

Rezende, Danilo Jimenez, Shakir Mohamed, and Daan Wierstra. "Stochastic backpropagation and approximate inference in deep generative models." arXiv preprint arXiv:1401.4082 (2014).

Mnih, Volodymyr, et al. "Human-level control through deep reinforcement learning." Nature 518.7540 (2015): 529-533.

Silver, David, et al. "Mastering the game of Go with deep neural networks and tree search." Nature 529.7587 (2016): 484-489.

Oord, Aaron van den, et al. "Conditional Image Generation with PixelCNN Decoders." arXiv preprint arXiv:1606.05328 (2016).