CRCNS Conference Workgroup #2 From Artificial Intelligence to Neuroscience, and back

Report

Main questions

This working group will put together views from the Artificial Intelligence (AI) and robotics community on the one hand, and from neuroscience (computational and experimental) on the other hand to address the following questions (and others):

- Is state-of-the-art AI (still) in position of fruitfully inseminating neuroscience? Have deep learning methods the potential to lead to new success stories? How could these and other current methods be used to analyse data, and to formalise information processing principles in neural circuits?

- Conversely, are discoveries in computational neuroscience in position of feeding AI and robotics with new ideas? In these domains, is neuro-inspiration complementary to the traditional engineering approach? What might neuroscience bring to the deep learning community?

Invited speakers & format

Time	AI & Neuroscience workgroup
10:00	Introduction and goals - Mehdi Khamassi, Boris Gutkin
10:10	**Position statement by Samuel Gershman** Questions & discussion
10:35	**Position statement by Haim Sopomlinsky** Questions & discussion
11:00	Coffee break
11:30	**Position statement by Sophie Deneve** Questions & discussion
11:55	**Position statement by Ron Meir** Questions & discussion
12:20	**General discussion & conclusions**
13:00	End of working group session

Historical

- **Long history of exchanges.**
  - Computer as human mind parable
  - Neural networks
  - Symbolic reasoning and language processing
  - Reinforcement Learning & action selection
  - Optimal feedback control theory & motor functions
  - Cognitive architectures
- **Nevertheless, alternance between convergence & separation because at first glance AI and Neuroscience have different goals:**
  - Understanding brain/thought.
  - Doing things that work on particular problems.

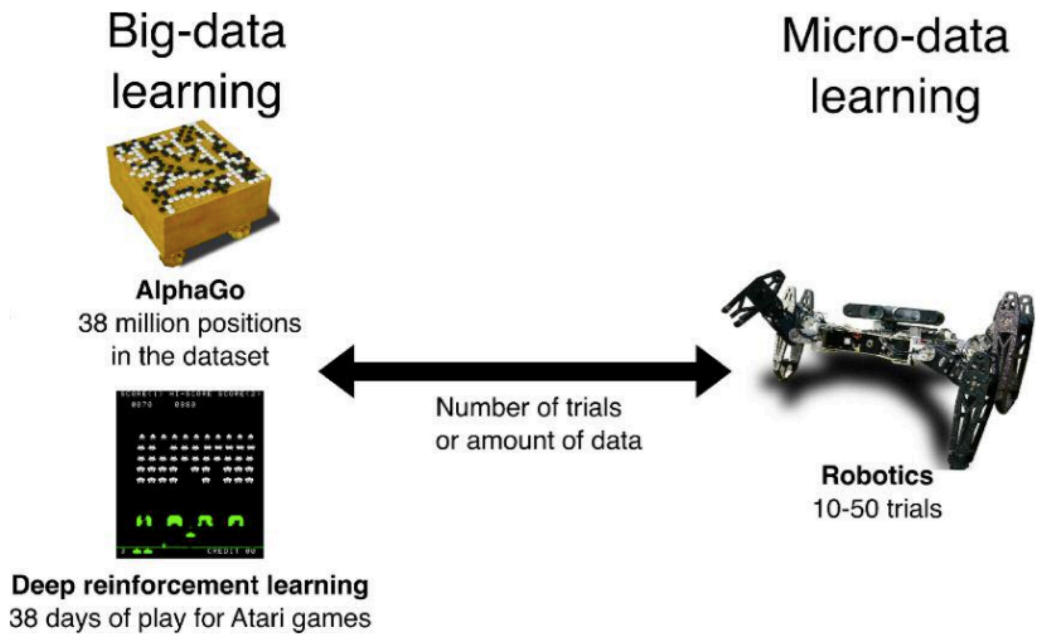
Possible cross-fertilizations

Importantly:

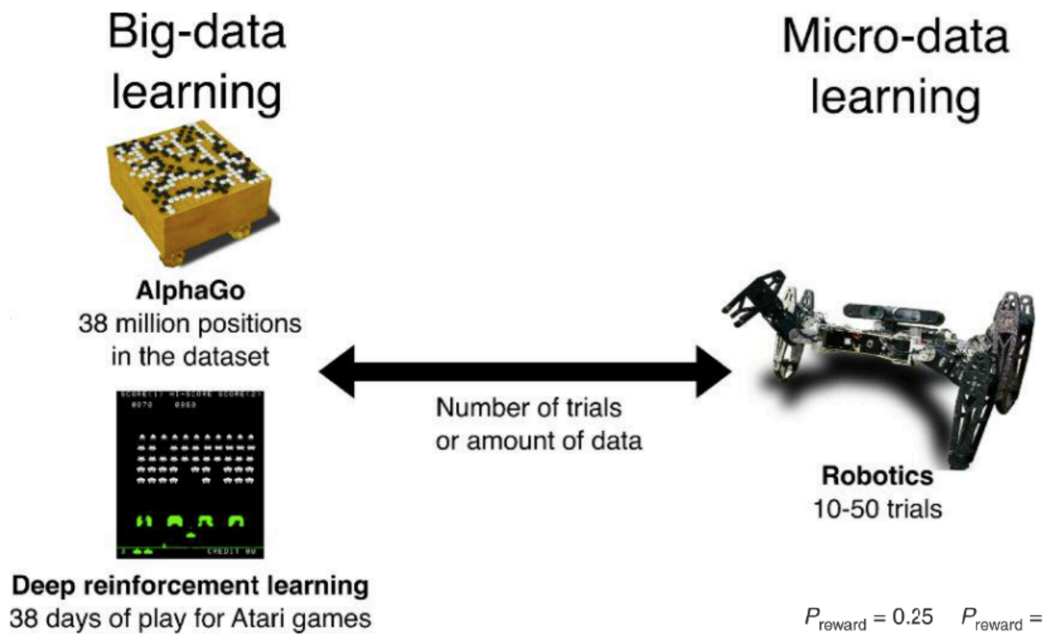
- **Common AI/NeuroComp affinity:** architecture, representation, most importantly reliance on learning. No intelligent design
- **Common challenges:** moving beyond statistical pattern recognition/distributed vector representations to models of cognitive systems. Moving beyond batch learning to enable online learning, generalization and life-long learning (learning-to-learn / meta-learning).

What about Deep learning?

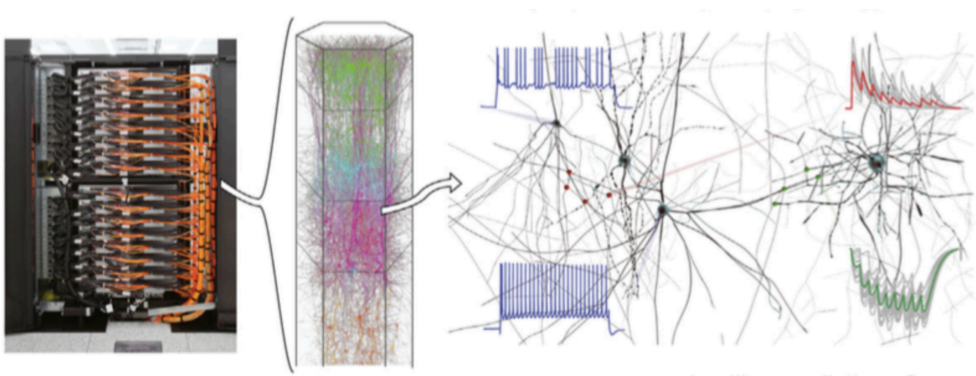
- Recent growing interests in deep neural networks, also interesting the industry (Google, Facebook, ..)
- Have deep learning methods the potential to lead to new success stories and contribute to better understanding brain functions?



Mouret 2016
ERCIM NEWS,
Special Issue on Machine Learning

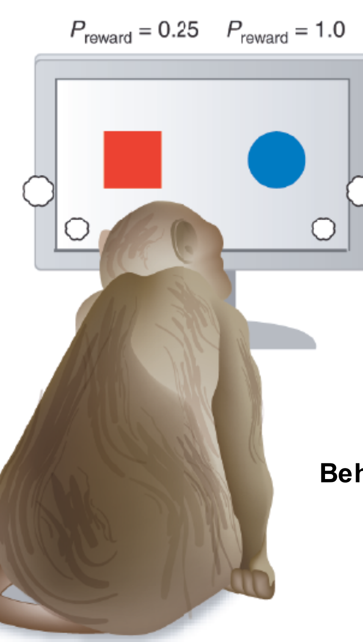


Mouret 2016
ERCIM NEWS,
Special Issue on Machine Learning



Markram et al. 2015

Reconstruction of neocortical circuitry
~31,000 neurons ~8 million connections
~37 million synapses



Niv et al. 2006

Behavioral & Cognitive Neuroscience
100-500 trials

Workgroup #2: From Artificial Intelligence to Neuroscience, and back

Messages from the talks

- **Sam Gershman:** not only care about biological plausibility but also psychological plausibility; take inspiration from human's cognitive abilities currently absent of deep neural nets: intuitive physics / intuitive psychology / learning-to-learn ability (meta-learning / life-long learning) / causality / compositionality.
- **Haim Sompolinsky:** Neuroscience provides AI with potential mechanisms for computation (new possibilities) / biological constraints (restrictions).

Messages from the talks

- **Sophie Deneve:** Lessons for biological recurrent neural networks: Efficient coding / excitation-inhibition balance / global function with local learning.
- **Ron Meir:** Can recent trends in AI suggest a link between biology and theory, through the implementation of engineering directed functionality by biologically inspired methods?

Suggestions of actions (I)

During the workshop, it has been recognized that recent progress in AI (especially in artificial neural networks) offers a unique opportunity for highly fruitful collaboration between computational neuroscience, AI/machine Learning (especially in academic research), and cognitive science.

Here are a series of listed areas of common interests and suggestions of actions:

1. **drawing ideas from AI:** new learning rules, architectures, and theoretical analyses of deep and other Machine Learning (ML) networks can provide insights for neuroscience models.

Suggestions of actions (II)

2. **drawing data from AI systems:** data both on natural signals and on neural-like representations of sensory signals in deep networks can provide useful testbed for computational neuroscience models of sensory hierarchies.

Suggestions of actions (III)

3. **contributing ideas to AI:** Ideas from computational neuroscience may be useful for AI, in particular the potential role of dynamics in AI networks is a fruitful common area of research interfacing with many computational neuroscience models (e.g., working memory models, excitation/inhibition balance, reservoir computing). What kinds of biological properties are useful for building intelligent machines that haven't yet been incorporated into AI, e.g., computing with spikes? Memory / decision / learning systems interaction / coordination principles within cognitive architectures?

Suggestions of actions (V)

4. **interaction with cognitive science:** Cognitive science can help both computational neuroscience and AI in characterizing human-like intelligence, such as intuitive physics, intuitive psychology, causal reasoning, compositionality, meta-learning (learning to learn), cognitive control. Building neural models that exhibit such features are crucial not only for the understanding of human cognition but also have the potential for generalization of knowledge far beyond the specialized domain in which current deep networks operate.

Suggestions of actions (VI)

5. **interaction with evolutionary psychology / neuroscience:** combining morphological, neurophysiological, anatomical and behavioral results in different species (with an emphasis on primate studies) can help better understand how the brain architecture was iteratively built. It can also help better understand the role of embodiment in cognition. This can both impact human/brain cognition and AI researches.

Suggestions of actions (VII)

6. **inclusion of sensorimotor control:** this issue covers the middle ground between currently popular pattern recognition styles successes and cognitive neuroscience. Moreover, it offers aspects which go beyond the current applications of deep networks, specifically embodiment and exploration. It is both critical to understand how human cognitive functions are anchored on sensori-motor information and to enable robust/efficient applications of AI to robotics.

Suggestions of actions (IV)

7. **psychiatry/neurology:** how can neural models and large AI networks help understand and treat brain disorders at the system level? The recent field of computational psychiatry, although still young, opens new avenues for collaboration between AI and neuroscience. A dedicated funding initiative could be directed towards this. Studying diseased AI could both contribute to understand human mental diseases and to understand how not to build mentally diseased AI.

Building machines that learn and think like people

Sam Gershman
Harvard University

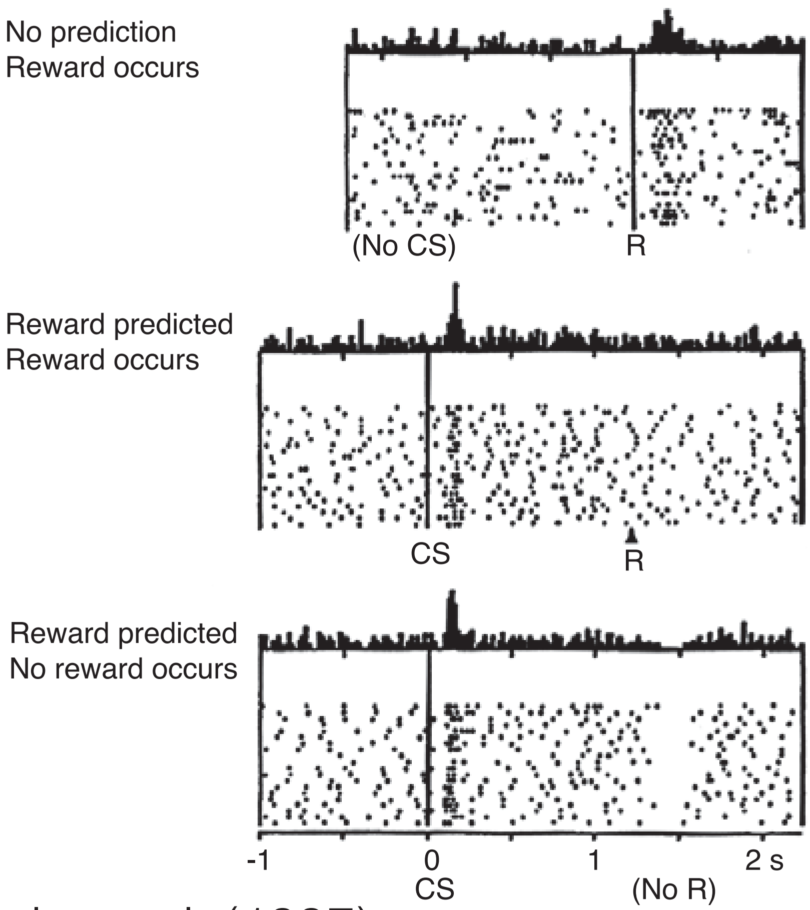
CRCNS 2016, Paris
Neuroscience and AI Workshop

Acknowledgments

- Brenden Lake (NYU)
- Josh Tenenbaum (MIT)
- Tomer Ullman (MIT)

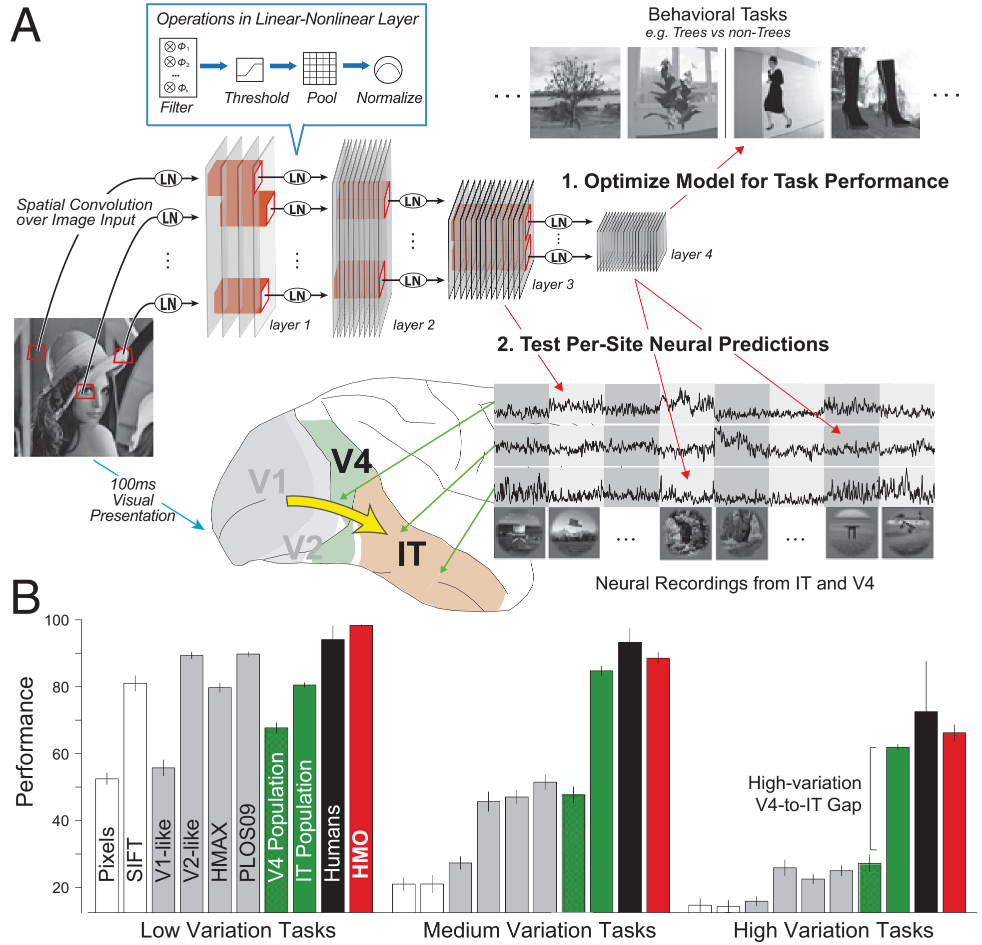
Computational neuroscience has learned a lot from AI

Reinforcement learning



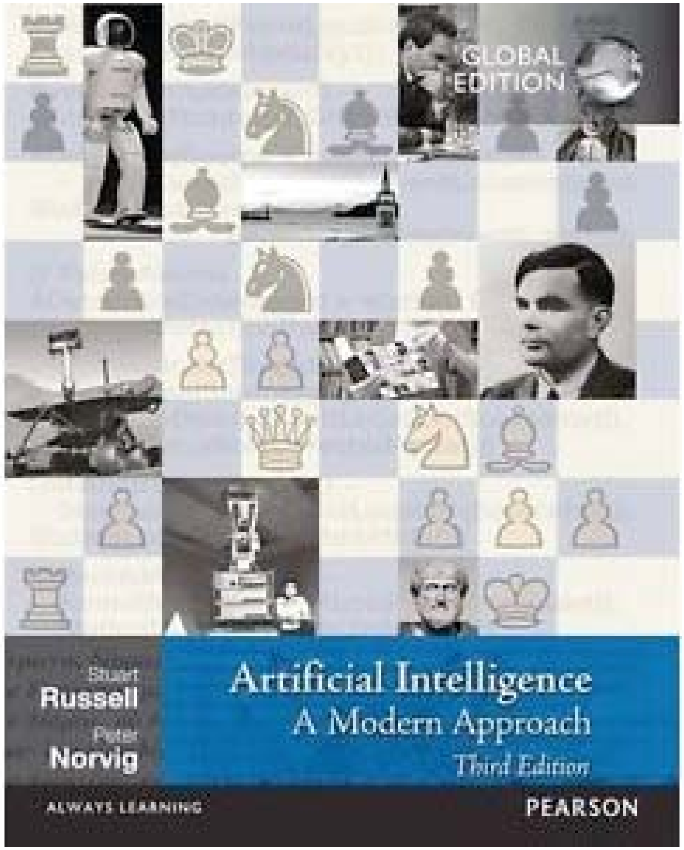
Schultz et al. (1997)

Object recognition



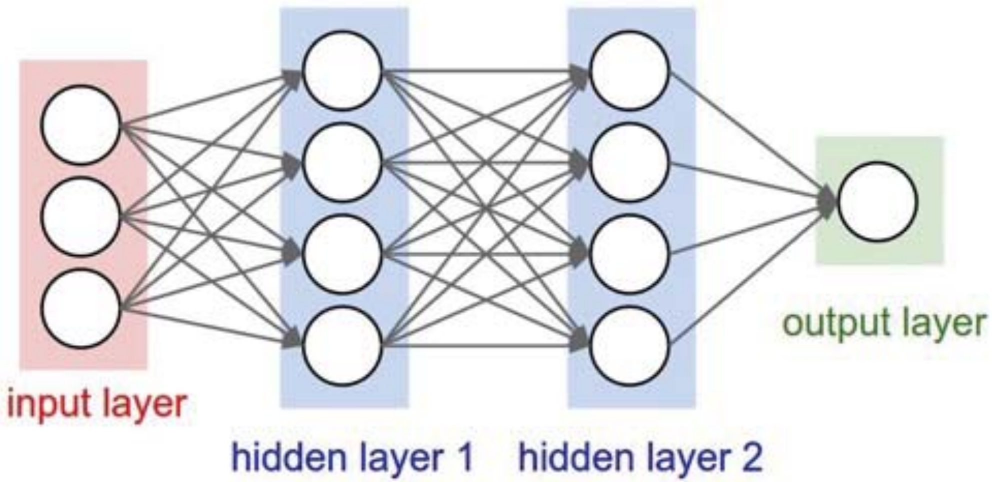
Yamins et al. (2014)

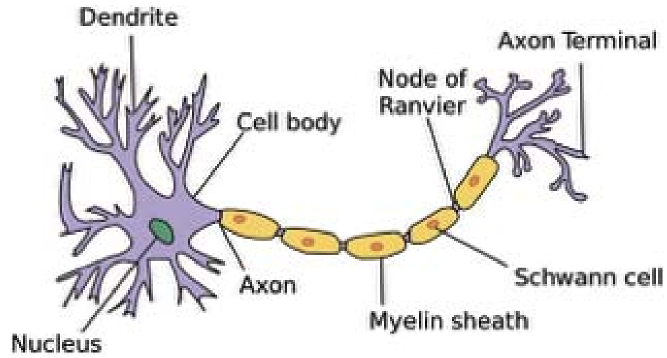
But what has AI learned from neuroscience?

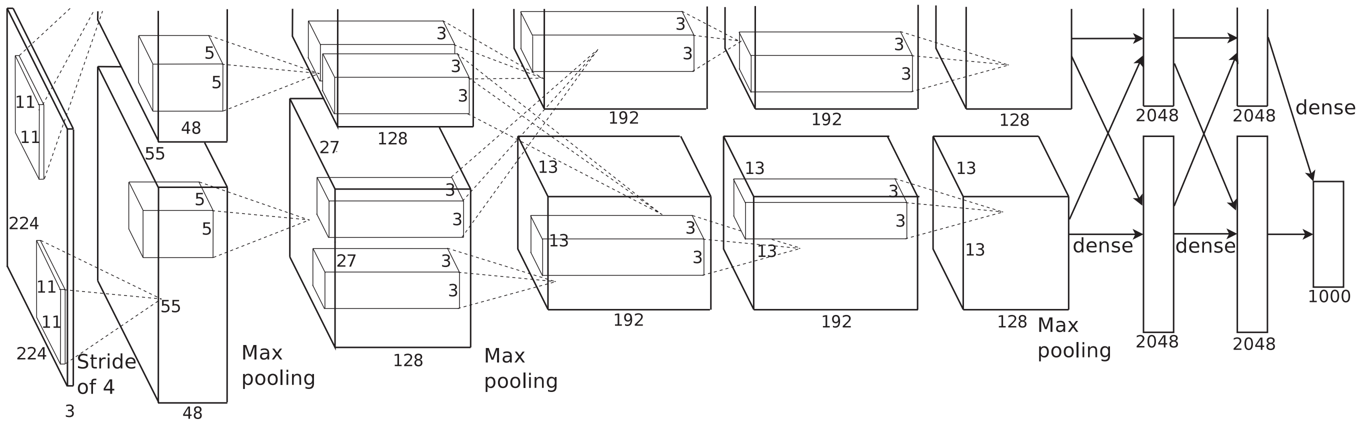


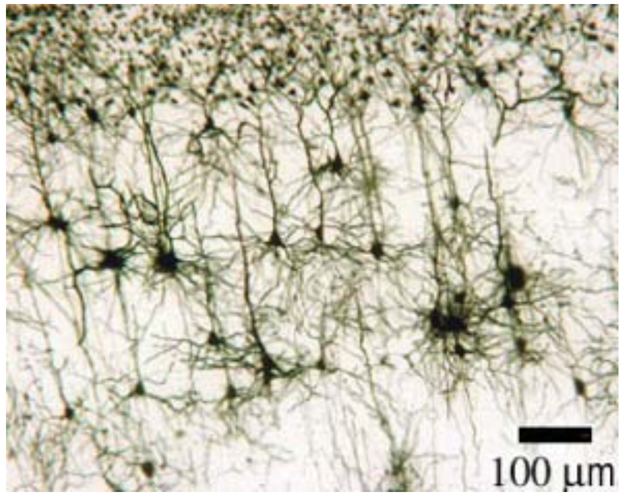
Russell and Norvig: "The quest for 'artificial flight' succeeded when the Wright brothers and others stopped imitating birds and started using wind tunnels and learning about aerodynamics."

Brain-inspired computation



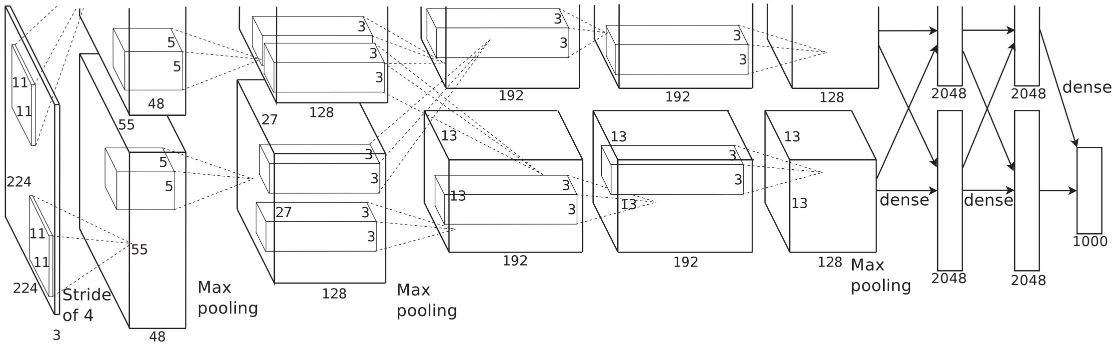




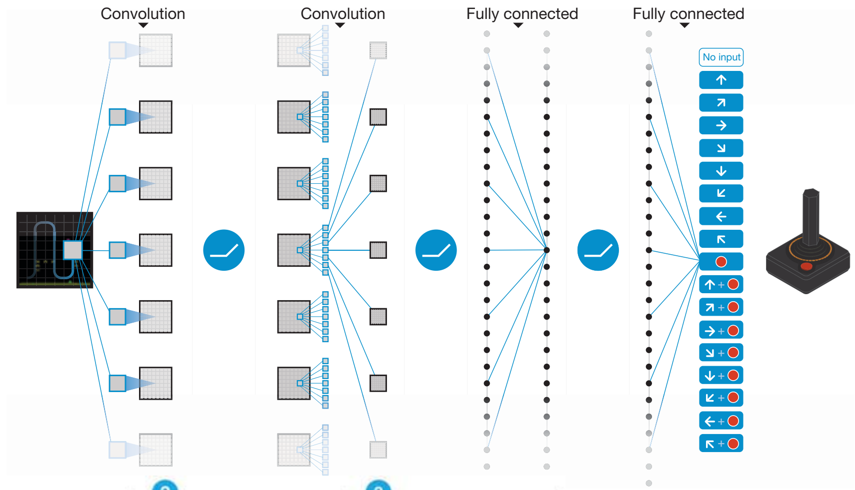


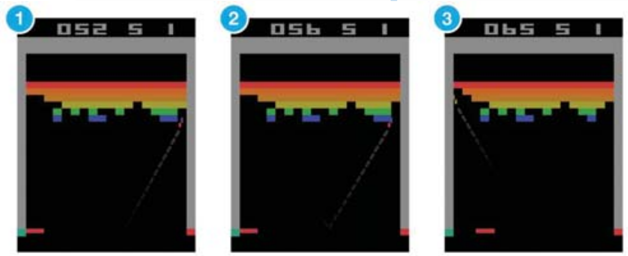
Recent progress in AI

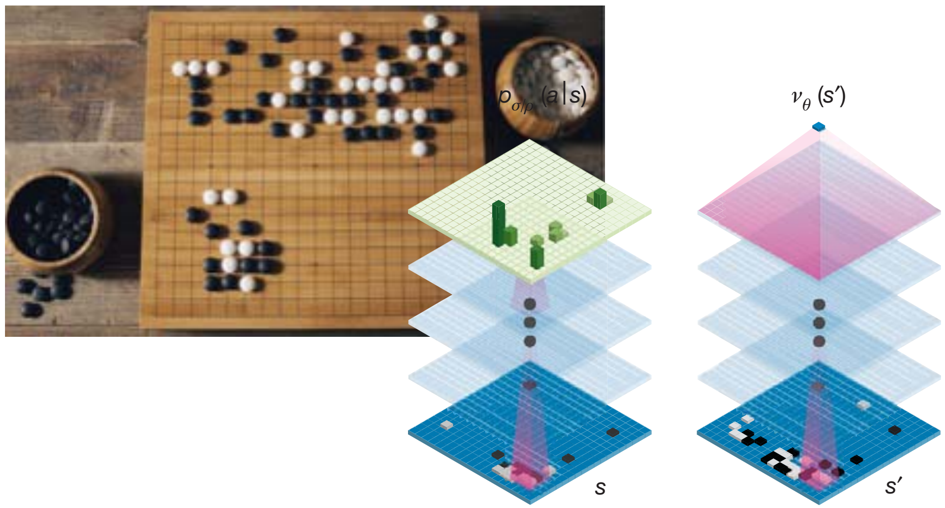
Object recognition



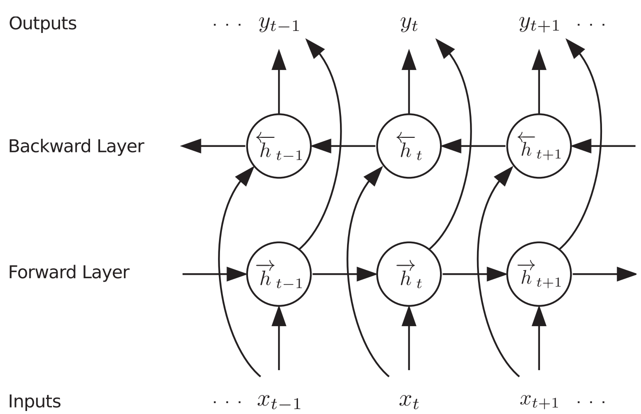
Game playing

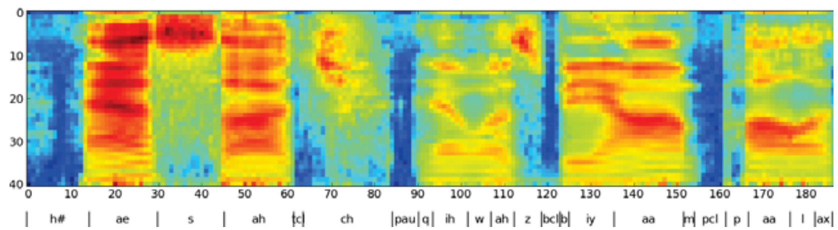






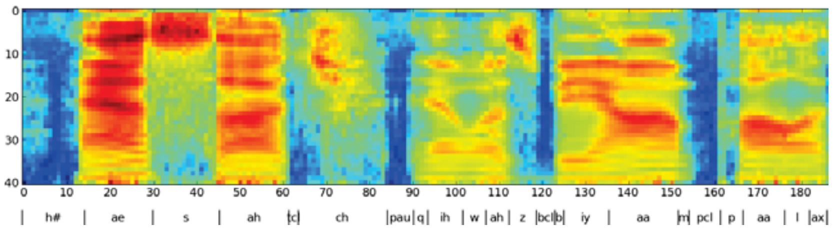
Speech recognition

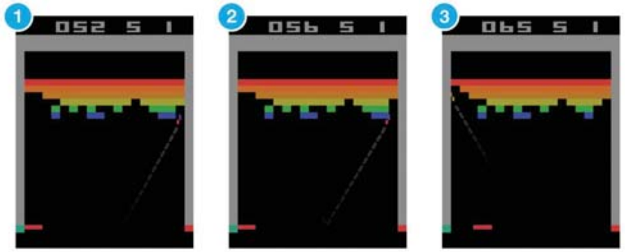




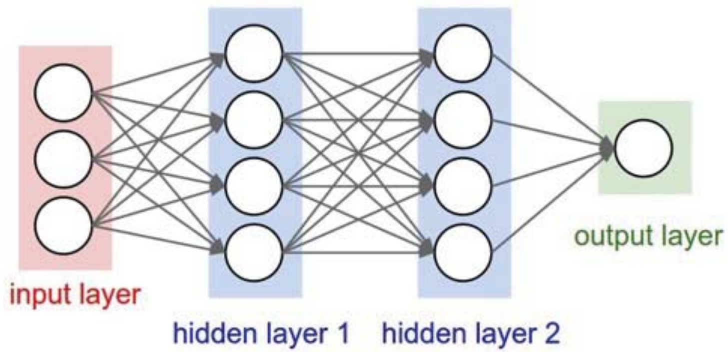
Machines that learn and think like people?

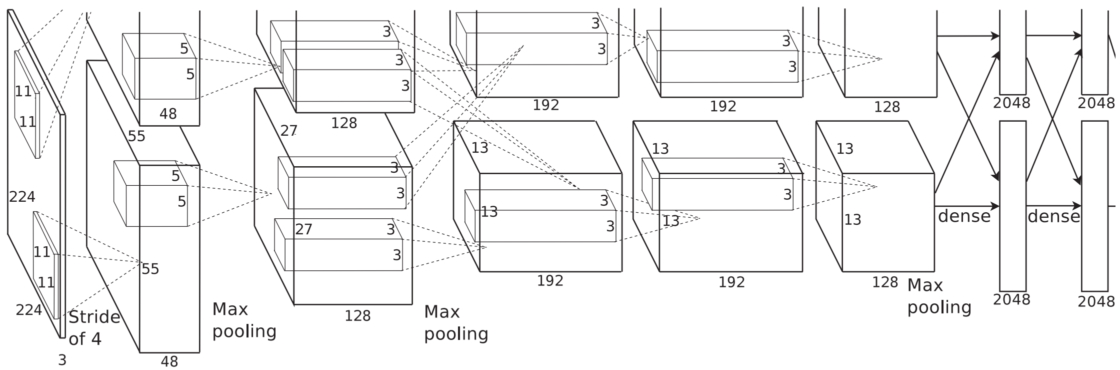
Human-level performance on important tasks





Brain-inspired computational architectures





NBC NEWS HOME TOP VIDEOS DECISION 2016 ONGOING: EUROPE'S ABORTION FIGHT PARIS TERROR ATTACKS

U.S. WORLD LOCAL POLITICS HEALTH TECH SCIENCE POP CULTURE BUSINESS INVESTIGATIONS SPORTS MORE NIGHTLY NEWS TODAY

FEB 25 2015, 1:05 PM ET

This AI Learned Atari Games Like Humans Do - And Now it Beats Them

by DEVIN COLDEWEY

The state of the AI's learning can be visually inspected, which shows how it has clustered and categorized different types of data. Google DeepMind

Ever been totally dominated by the computer player in a video game? A new artificial intelligence system takes on all comers with a handful of old Atari titles, and it does so after learning the rules bit by bit like a human. Its creators claim this is just the very beginning of what it can do. In a few years, it may be driving you to work.

nature International weekly journal of science

Home | News & Comment | Research | Careers & Jobs | Current Issue | Archive | Audio & Video | For Au

Archive > Volume 518 > Issue 7540 > News > Article

NATURE | NEWS

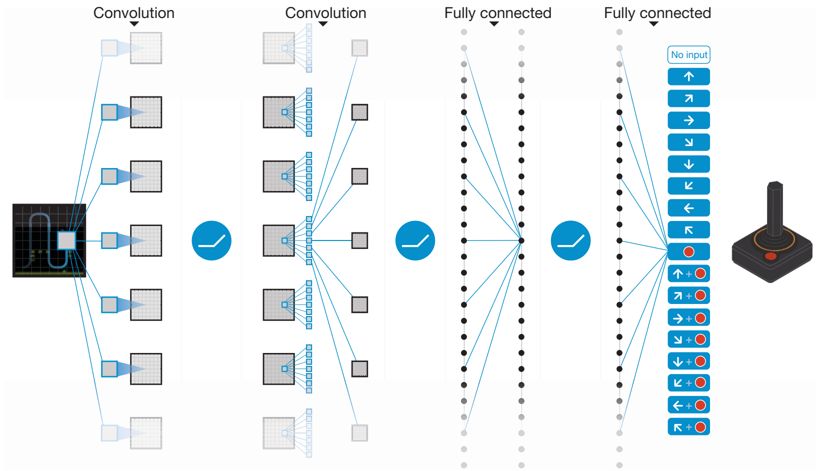
عربي

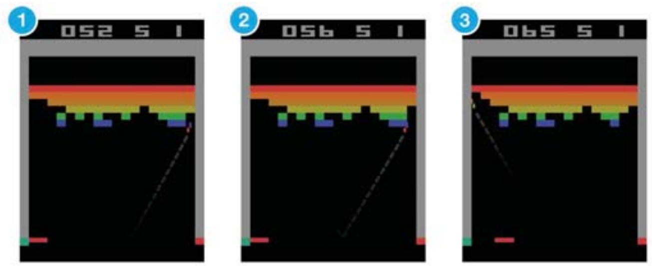
Game-playing software holds lessons for neuroscience

DeepMind computer provides new way to investigate how the brain works.

Elizabeth Gibney

25 February 2015





Technology | Wed Jan 27, 2016 8:16pm EST

Related: SCIENCE

Go figure! Game victory seen as artificial intelligence milestone

WASHINGTON | BY WILL DUNHAM

"It's a very beautiful game with extremely simple rules that lead to profound complexity. In fact, Go is probably the most complex game ever devised by humans," said Hassabis, a former child chess prodigy.

Scientists have made artificial intelligence strides in recent years, making computers think and learn more like people do.

Markos Kay / Quanta

The Atlantic

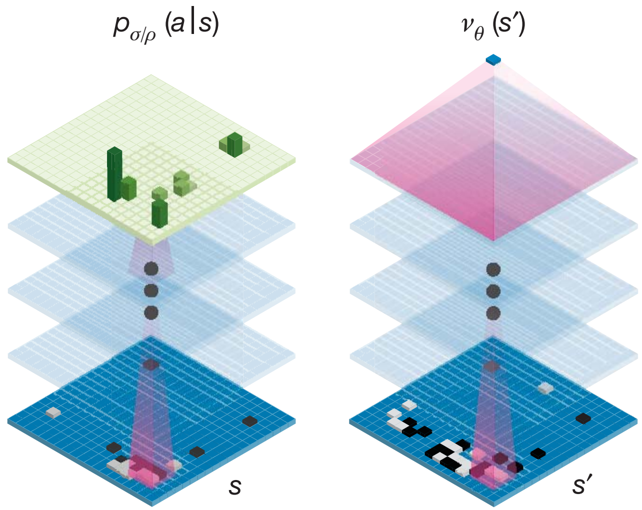
How Google's AlphaGo Imitates Human Intuition

The computer's defeat of one of the world's greatest Go players speaks to the potential for artificial-intelligence systems with real instincts.

1.6k

TEXT SIZE

MICHAEL NIELSEN | APR 4, 2016 | TECHNOLOGY



In 1997, IBM's Deep Blue system defeated the world chess champion, Garry

Goals

1. What does it mean for a machine to learn and think like a person?

2. How close are contemporary AI systems to reaching this standard?

3. Should AI researchers care about building more human-like machines?

Goals

1. What does it mean for a machine to learn and think like a person?
    - Five key ingredients with strong empirical support from cognitive science.
2. How close are contemporary AI systems to reaching this standard?
    - Deep learning systems have not yet incorporated many of the key ingredients, and thus may be solving problems in different ways than people.
3. Should AI researchers care about building more human-like machines?
    - Yes — in most cognitively natural domains, people are still much better learners and thinkers the best AI systems, and we still have a lot to learn by reverse engineering the human mind.

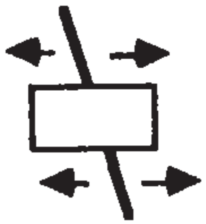
Key ingredients

1. Intuitive physics
2. Intuitive psychology
3. Compositionality
4. Learning-to-learn
5. Causality

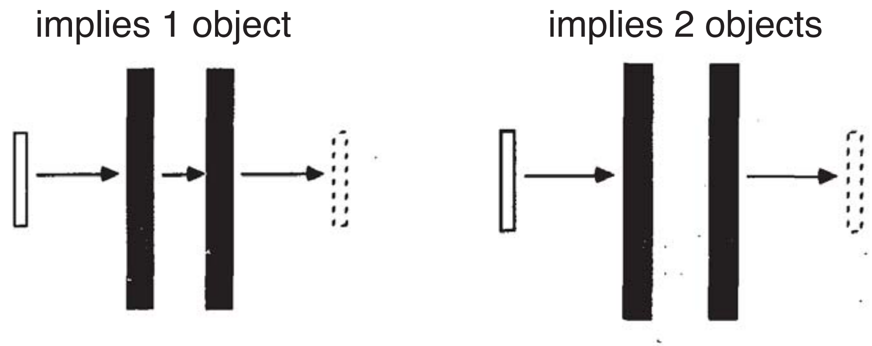
Intuitive physics

Intuitive physics

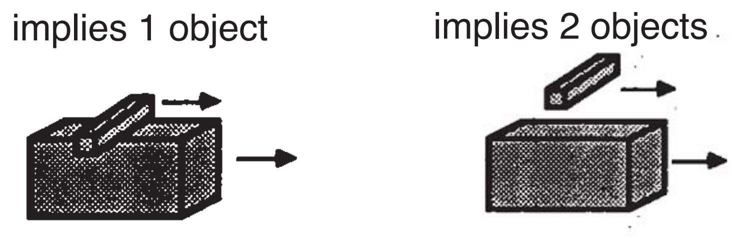
Present in the first few months of life:



cohesion: objects move as connected and bounded wholes



continuity: objects move on connected paths



contact: objects do not interact at a distance

Present by first birthday:

inertia, support, containment, collisions, …

see work by Elizabeth Spelke, Renée Baillargeon and colleagues

Intuitive physics

What are the benefits of intuitive physics?

1. *Reduced sample complexity*. No need to re-lean physics for each object type, or for each type of game.

2. *Deeper generalizations.*

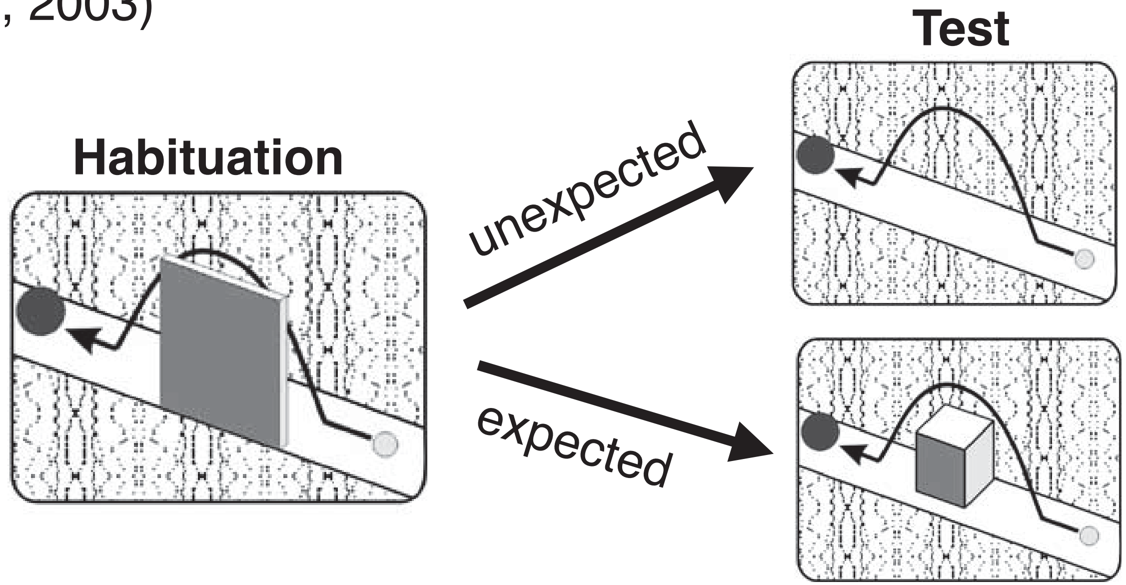
    A. Can handle changes in object appearance.

    B. Seamless addition of new objects

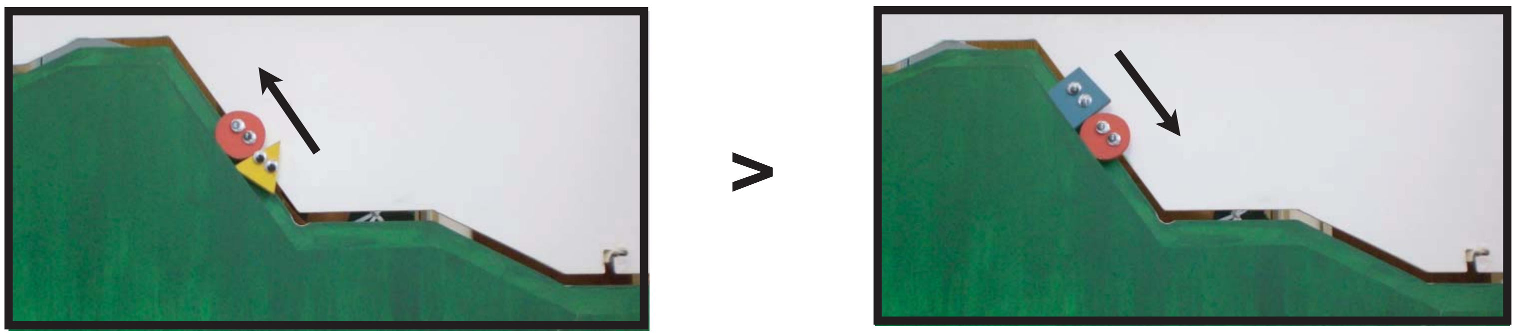
Intuitive psychology

Understanding that people have mental states (beliefs and desires), which explain and predict their behavior.

Infants expect agents to take efficient actions to achieve goals (12 month olds; Csibra et al., 2003)



Infants can discriminate helpers from hinderers (6 month olds; Hamlin et al., 2010)



Compositionality

New representations can be constructed through the combination of primitive elements.

Computer programming: primitive functions combined to create new functions, which can be further combined, etc.

Language and thought: people can think an infinite number of thoughts, and utter or understand an infinite number of sentences (Fodor, 1975; Fodor & Phylsyn, 1988)

Vision: new objects can be represented as novel combinations of parts and relations
(e.g., Structural Descriptions: Winston, 1975; Marr & Nishihara, 1978; Biederman, 1987)

Compositionality

What are the benefits of compositionality?

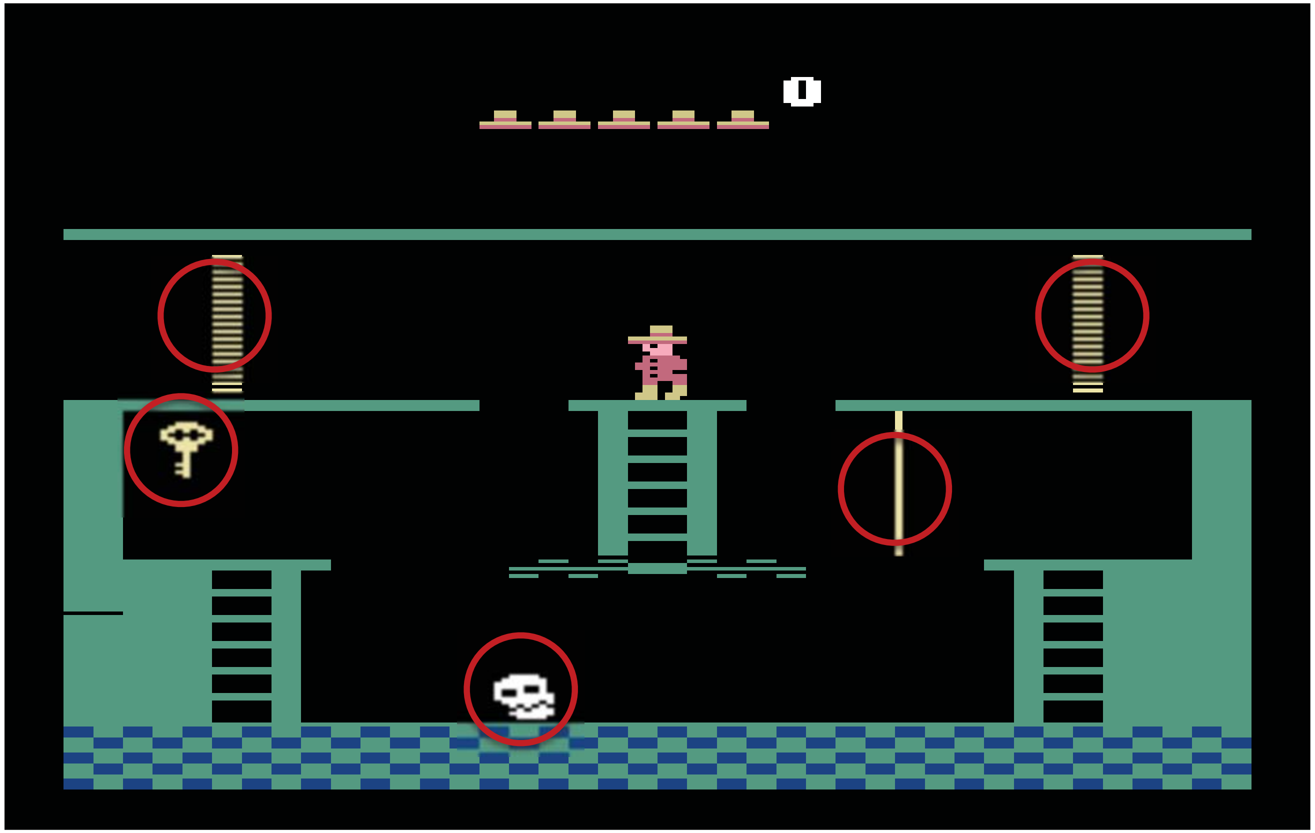
Reduced sample complexity and deeper generalizations:

- A finite number of object types can generate an infinite number of scenes
- All objects of a type have the same behavior (and consequences).

Compositionality

Dividing goals into sub-goals
can help when learning from sparse rewards.

Montezuma's Revenge



Standard DQN: Almost no progress on Montezuma's revenge (agent just falls and dies)

Hierarchical DQN: Progress can be made with explicit object-based sub-goals (Kulkarni et al., 2016)

Learning-to-learn

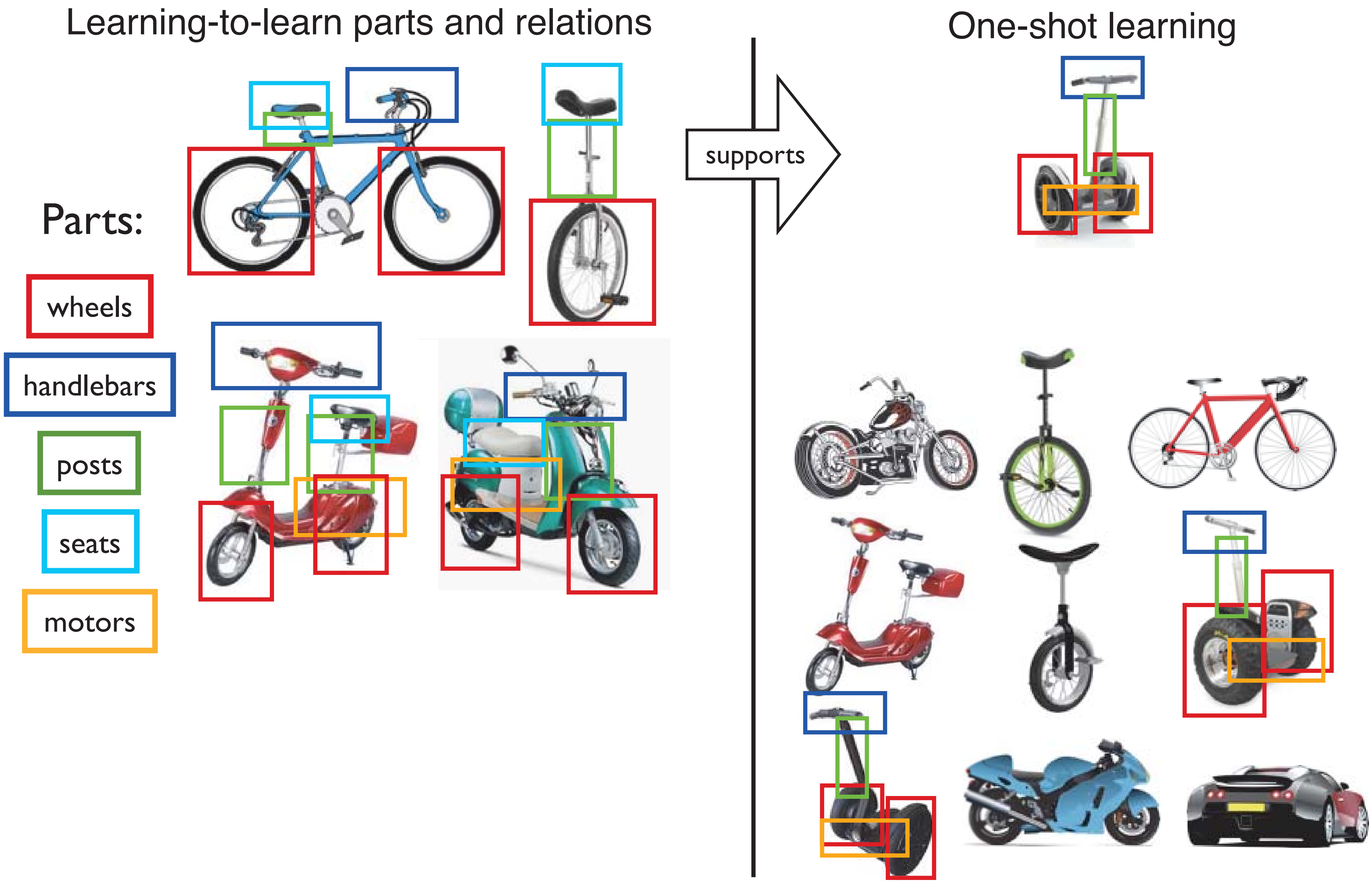
Learning new tasks is accelerated due to previous learning.

Deep learning models already utilize learning-to-learn in several ways:

- Sharing features between classes is standard in object recognition (Anselmi et al., 2016; Baxter, 2000; Zeiler & Fergus, 2014)

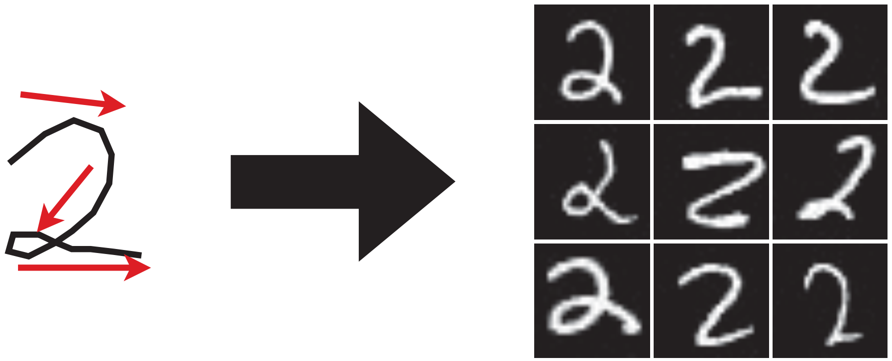
- Re-using previously learned knowledge/modules to help perform new tasks (Bottou, 2014; Lopez-Paz, Bottou, Scholköpf, & Vapnik, 2016)

A synergy between compositionality and learning-to-learn



Causality

Human concepts often go beyond *features* to resemble *causal explanations,* or beyond *pattern recognition* to resemble *model building.*



(Hinton & Nair, 2006)

(Murphy & Medin, 1985)

Causality

Towards richer model-based scene understanding

a man riding a motorcycle on a beach

an airplane is parked on the tarmac at an airport

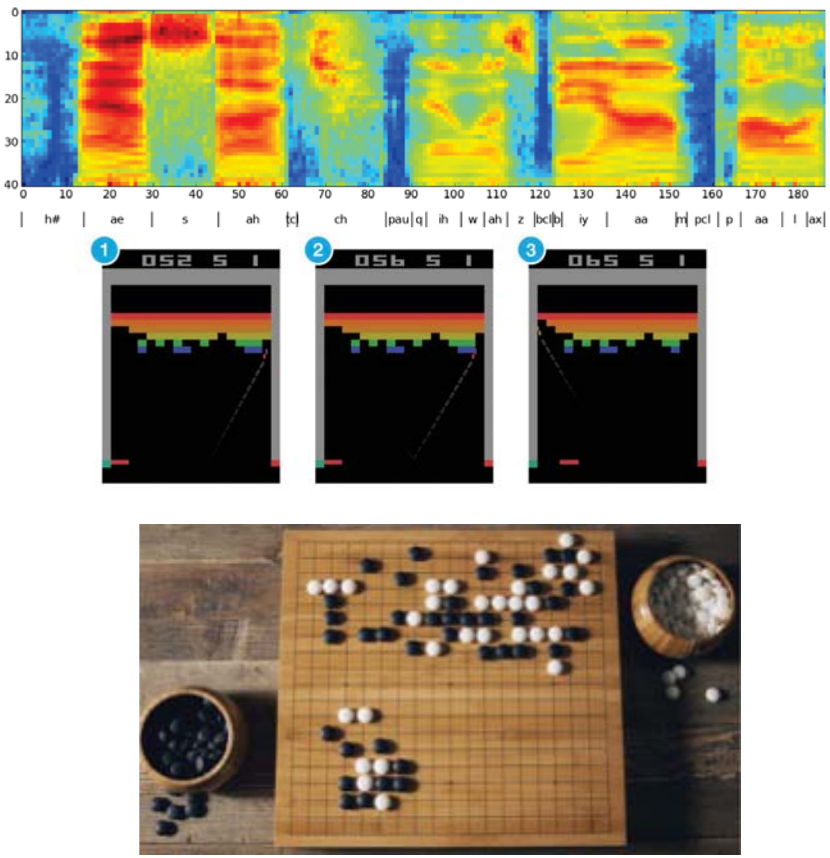
a group of people standing on top of a beach

Image captions generated by NeuralTalk (Karpathy &
(code from https://github.com/karpathy/neuraltalk2)

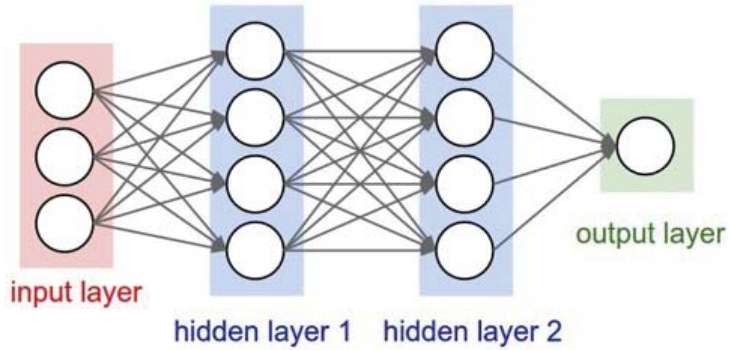
Similar examples at twitter.com/interesting_jpg/

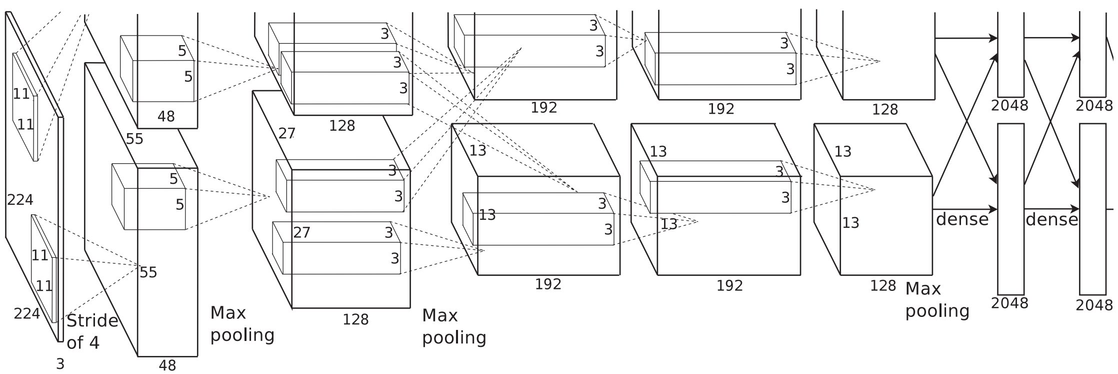
Machines that learn and think like people?

Human-level performance on important tasks



Brain-inspired computational architectures





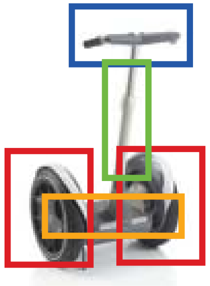
Machines that learn and think like people (?)

Human-level performance on important tasks

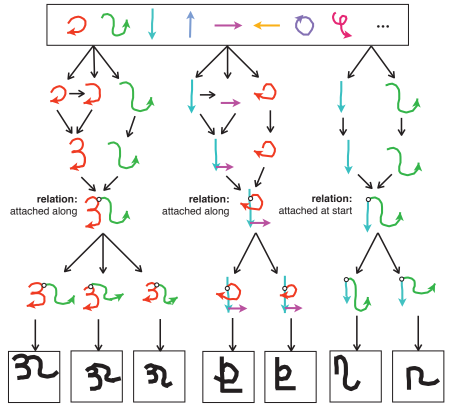
Brain-inspired computational architectures

Intuitive physics

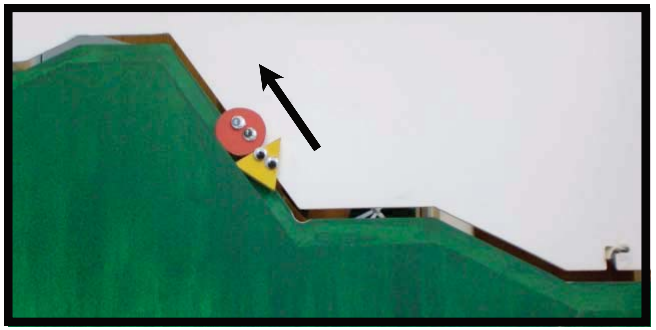
Compositionality



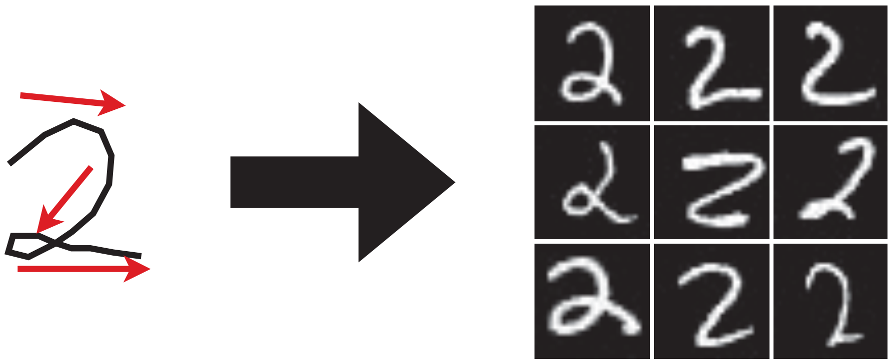
Learning-to-learn



Intuitive psychology



Causality



Current trend in deep learning: Integrating psychological ingredients

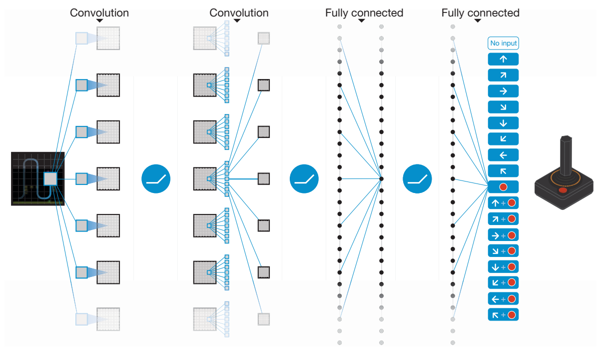
Why not also higher-level cognitive ingredients?

Selective attention

A woman is throwing a **frisbee** in a park.

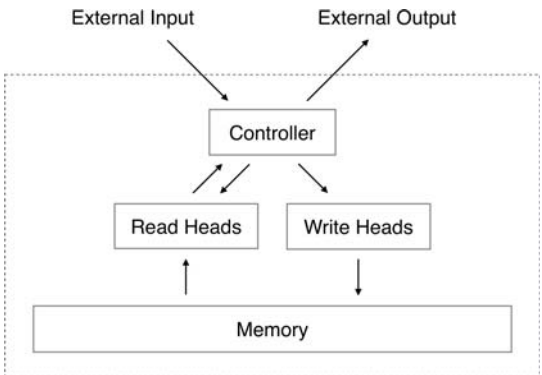
Machine translation (Bahdanau et al. 2015)
Object recognition (Mnih et al., 2014)
Caption generation (Xu et al., 2015)

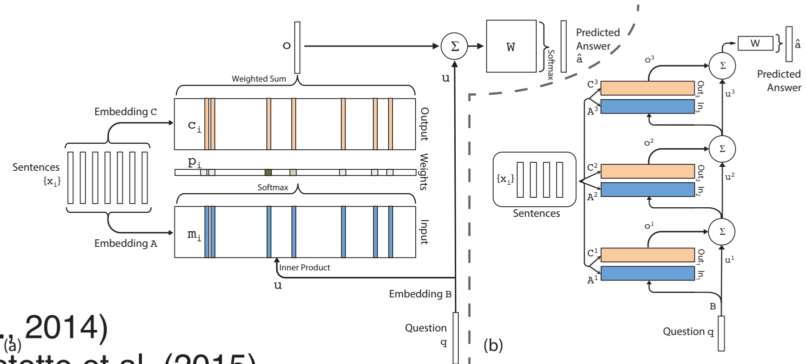
Experience replay



(Mnih et al., 2015; Schaul et al., 2016)

External working memories





Neural Turing Machine (Graves et al., 2014)
Neutral Stacks and Queues (Grefenstette et al. (2015)
Memory Networks (Sukhbaatar et al., 2015; Weston et al., 2015)
Neural Program-Interpretors (Reed & de Freitas, 2016)

Potential applications to practical AI tasks

Scene understanding

a man riding a motorcycle on a beach

an airplane is parked on the tarmac at an airport

a group of people standing on top of a beach

Autonomous driving

Autonomous agents and intelligent devices

Creative design

What about biological plausibility?

- Biological plausibility arguments are often used to knock down cognitive theories.
- However, these arguments sometimes rest on dubious "facts" about the relationship between biological and psychological phenomena.
- For example, Gallistel has argued that LTP does not explain key psychological characteristics of memory. Need to consider **cognitive plausibility**.

What kind of neural network is the brain?

- We know that the brain is a neural network, so human intelligence has to be implemented in a neural network.
- But what kind of network? We shouldn’t get prematurely tied down to current notions of “biological plausibility”. Cognitively plausible neural networks might look very different.

Thank you