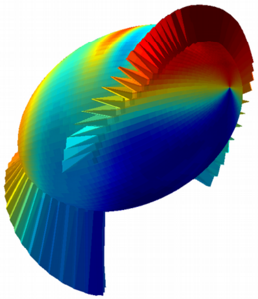
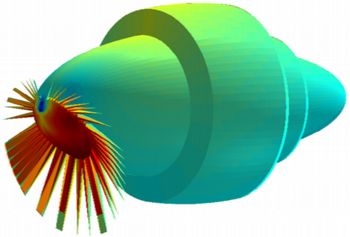
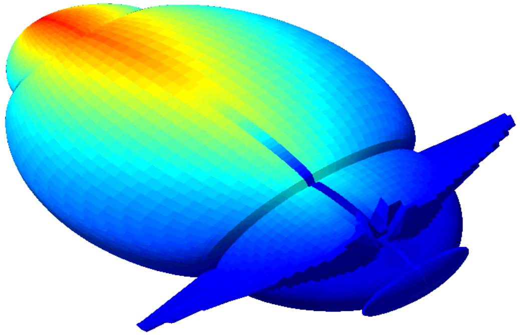
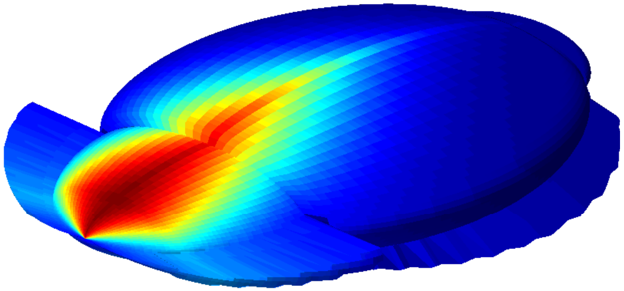




Lecture 1
Intelligent Systems





Artificial intelligence (AI) is everywhere!

But what is AI?

Is AI = Machine Learning?

Is AI = Deep Learning?

Is Machine Learning = Deep Learning?

Are we getting close to having “real” AI?

What is AI?

Building smart tools

Systems that can:

- Intelligently process a lot of information and data.
- Help us make smarter decisions.
- Take care of routine tasks automatically.
- Do skilled tasks like driving, construction, etc., automatically.
- Work as intelligent assistants in surgery, law, education, science, etc.

or

Building artificial minds

Systems capable of:

- Thinking, planning and imagination.
- Innovation and creativity.
- Making truly autonomous choices.
- Having their own values,

What is AI?

Building smart tools

Systems that can:

- Intelligently process a lot of information and data.
- Help us make smarter decisions.
- Take care of routine tasks automatically.
- Do skilled tasks like driving, construction, etc., automatically.
- Work as intelligent assistants in surgery, law, education, science, etc.

or

Building artificial minds

Systems capable of:

- Thinking, planning and imagination.
- Innovation and creativity.
- Making truly autonomous choices.
- Having their own values,

Artificial General Intelligence (AGI)

Or is there no difference?

What is the Mind?

The Mind-Body Problem

Dualism: **Mind and body are different substances.**
Science: **Mind is a property of the physical body.**

AI is based fundamentally on the scientific view of the mind.

http://frithmind.org/socialminds/wp-content/uploads/sites/4/2014/05/Fig-1.jpg

Pattern Recognition

Currently, almost all of AI is based on a **pattern-program framework**

Basic Elements:

- The world has a lot of spatiotemporal structure → **Patterns**.
- Every situation requires specific responses → **Programs**.
- An intelligent system:
  - ✓ Recognizes these patterns in its sensory input = **Perception**.
  - ✓ Maps them into more complex internal patterns = **Cognition**.
  - ✓ Uses these maps to trigger response programs = **Action**.
  - ✓ Responses = choices, decisions, predictions, behaviors, etc.
- Learning = improving pattern recognition and program generation.
- Increased intelligence = more complex perception, cognition, action.

Q1: Is this framework correct?
Q2: How is this framework to be instantiated?

Building Intelligence

Intelligence by Design

- Build computer programs that can do smart things.
- Put them on a computer with sensory input and verbal or mechanical output.
- Fine tune the programs manually.
- Do this separately for all functions of intelligence, and combine.

Intelligence by Learning

- Build a (brain-like) system that can learn from data and produce output.
- Give the system a lot of data and let it learn.
- Test the learning with new data and new situations.
- Do this separately for all functions of intelligence, and combine.

Intelligence by Emergence

- Build a physical system with sensory input, verbal and mechanical output, simple interacting behaviors, and learning.
- Let it learn by experience in the real-world.
- Let the functions of intelligence emerge from the experience and learning.

Building Intelligence

Intelligence by Design

- Build computer programs that can do smart things.
- Put them on a computer with sensory input and verbal or mechanical output.
- Fine tune the programs manually.
- Do this separately for all functions of intelligence, and combine.

Intelligence by Learning

- Build a (brain-like?) system that can learn from data and produce output.
- Give the system a lot of data and let it learn.
- Test the learning with new data and new situations.
- Do this separately for all functions of intelligence, and combine.

Intelligence by Emergence

- Build a physical system with sensory input, verbal and mechanical output, simple interacting behaviors, and learning.
- Let it learn by experience in the real-world.
- Let the functions of intelligence emerge from the experience and learning.

Do these have to be mutually exclusive?

So is this AI?

Memories

Labels

Translations

Decisions

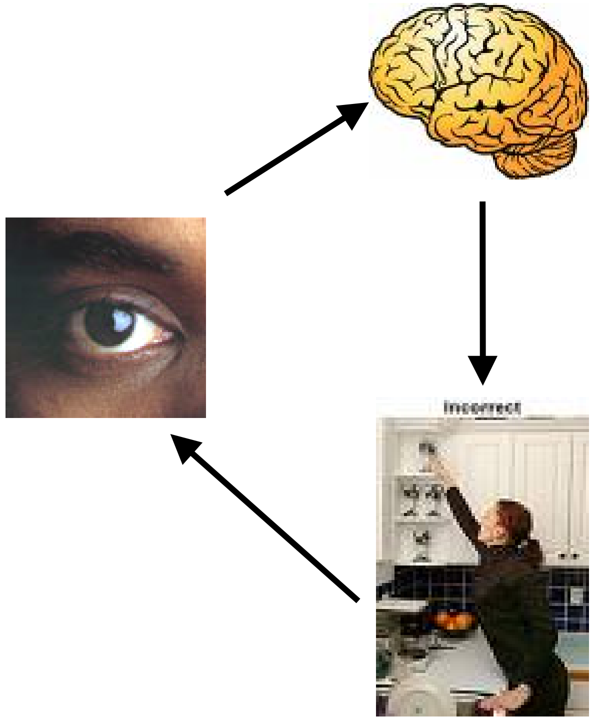
Advice

Explanations

The Classical "Symbolic" View of Intelligence

The mind is abstract computation

The Sense-Think-Act Cycle:



Key Assumption:

Intelligence can be produced by implementing the ***abstract processes*** of the mind as ***algorithms***

Approach:

- Implement algorithms for all the abstract processes.
- Design the control mechanisms to make sure that the right ones are used.

This is a fundamentally ***reductionistic*** approach.

Is the brain really a computer?

Is the brain really a computer?

The Computer Metaphor:

Information processing = computation →

The brain is like a computer:

- Perception = Input.
- Cognition = Processing.
- Action = Output.

Is the brain really a computer?

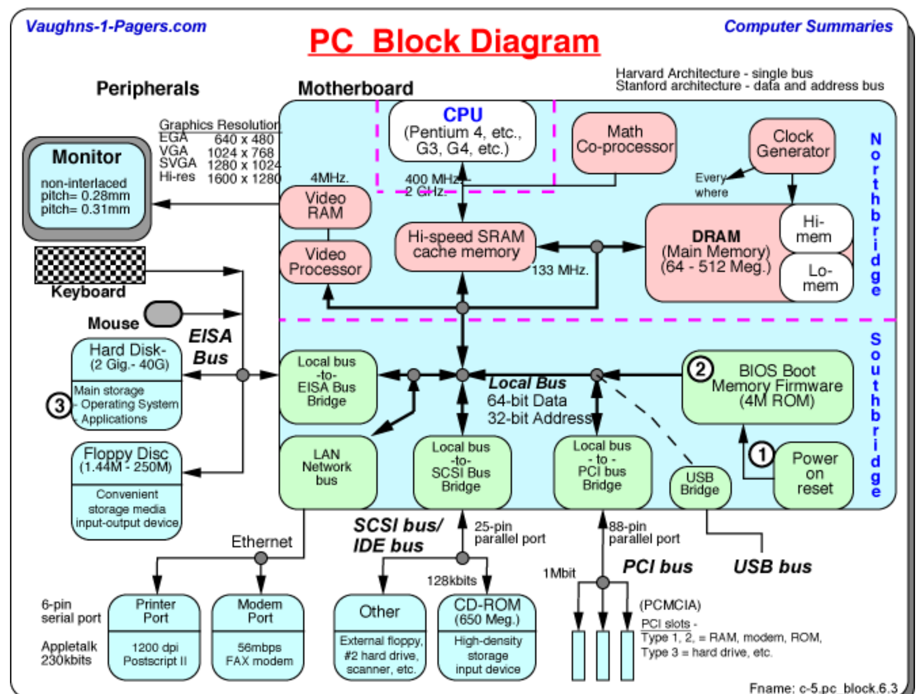
The Computer Metaphor:

Information processing = computation

The brain is like a computer:

- Perception = Input.
- Cognition = Processing.
- Action = Output.

But is it this kind of computer?



http://www.vaughns-1-pagers.com/computer/pc-block-diagram.htm

Both brains and digital computers:

- Are physical systems.
- Use electrical signaling.
- Acquire, process, and output information.
- Store and recall information.
- Are versatile, general-purpose information processors.

Both brains and digital computers:

- Are physical systems.
- Use electrical signaling.
- Acquire, process, and output information.
- Store and recall information.
- Are versatile, general-purpose information processors.

But only the brain:

- Is a self-organizing, adaptive, analog dynamical system.
- Uses “slow” chemical components to achieve “fast” functionality.
- Processes information in distributed networks.
- Stores information extensively rather than locally.
- Accesses memory by association, not address.
- ***Embodies its “programs” in its physics.***

Both brains and digital computers:

- Are physical systems.
- Use electrical signaling.
- Acquire, process, and output information.
- Can store and recall information.
- Are versatile, general-purpose problem solvers.

But only the brain:

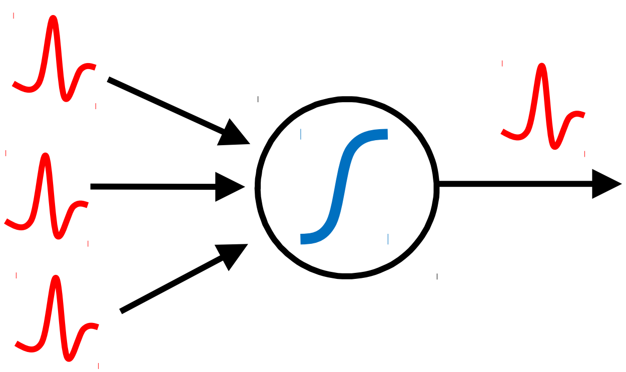
- Is a self-organizing, adaptive, analog dynamical system.
- Uses "slow" chemical components to achieve "fast" functionality.
- Processes information in adaptive networks.
- Stores information extensively rather than locally.
- Accesses memory by association, not address.
- ***Embodies its "programs" in its physics.***

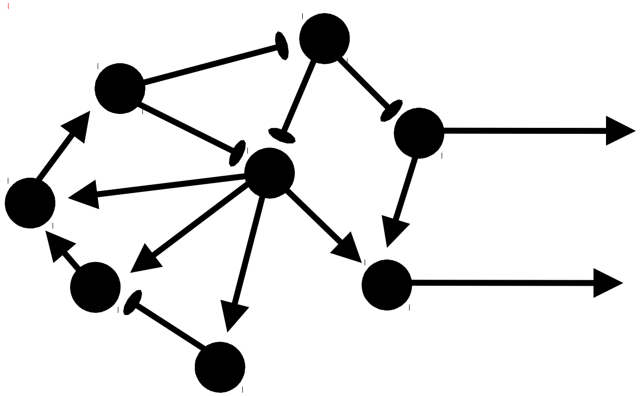
Brains are special – not like ordinary computers…

The Brain-Based View of Intelligence

The mind emerges from the brain

Neural Networks:





Key Assumption:

Intelligence arises from the interaction of many ***neural networks*** operating in parallel.

Approach:

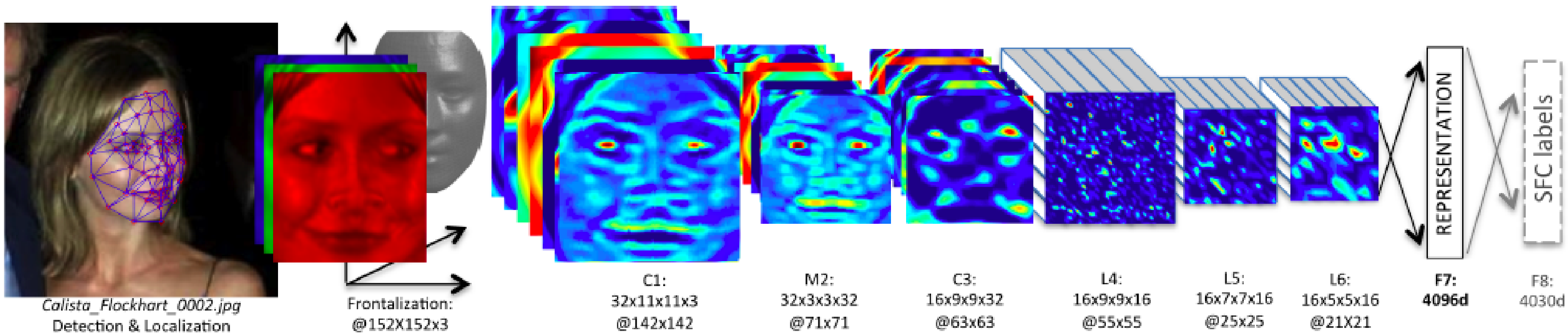
- Build the neural networks
- Design their interaction network.

The system's behavior is modified by changing connections

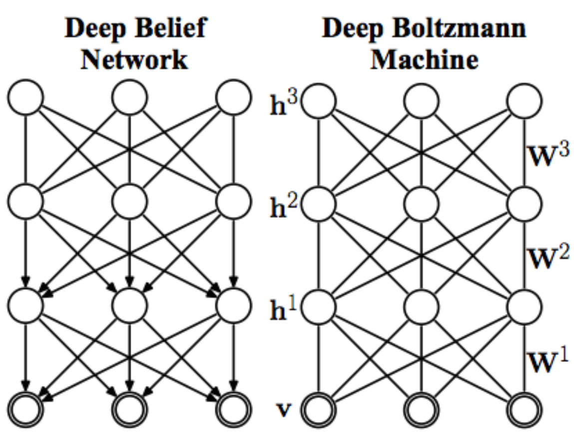
→ Learning

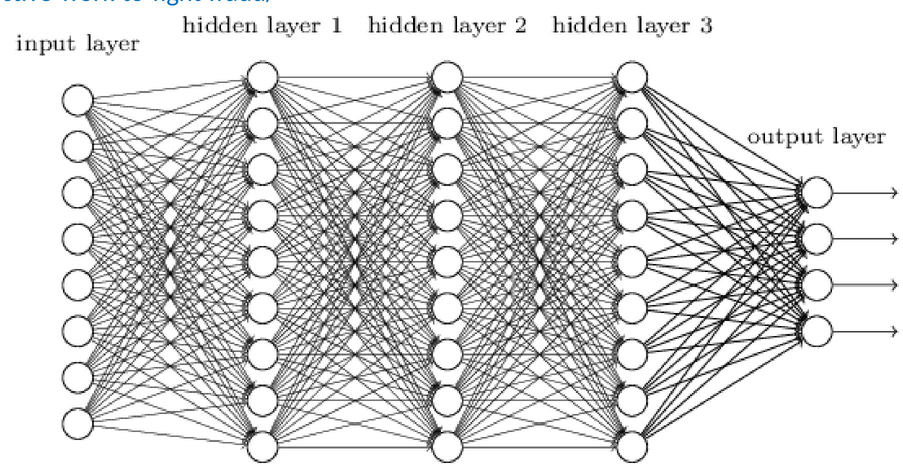
Deep Learning

- Deep learning systems are neural networks with many layers of neurons.
- Different layers may performs distinct functions.



https://gigaom.com/2015/03/06/how-paypal-uses-deep-learning-and-detective-work-to-fight-fraud/

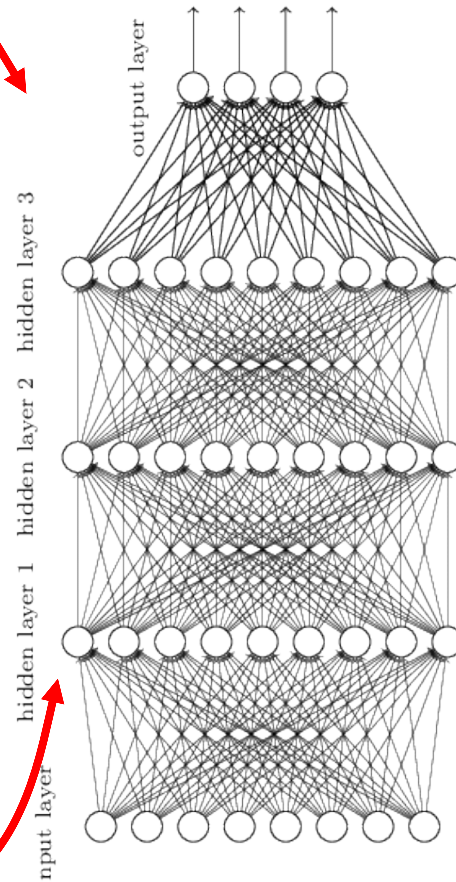




https://www.quora.com/What-is-the-difference-between-a-deep-Boltzmann-machine-and-a-deep-belief-net

http://neuralnetworksanddeeplearning.com/chap6.html

So is this AI?



Memories

Labels

Translations

Decisions

Advice

Explanations

The New AI

The New AI

Machine Learning

+

Big Data

+

Extremely Fast Computing

+

Frameworks & Libraries

The New AI

Machine Learning

- Neural networks / deep learning
- Statistical / Bayesian models
- Graphical models
- Generative models

+

Big Data

+

Extremely Fast Computing

+

Frameworks & Libraries

The New AI

Machine Learning

+

Big Data

- On-line / streaming data.
- Distributed databases.
- Analytics.
- Visualization.

+

Extremely Fast Computing

+

Frameworks & Libraries

The New AI

Machine Learning

+

Big Data

+

Extremely Fast Computing

- Multithreading.
- Cluster computing.
- GPU systems.
- Neuromorphic chips.

+

Frameworks & Libraries

The New AI

Machine Learning

+

Big Data

+

Extremely Fast Computing

+

Frameworks & Libraries

- Hadoop / MapReduce
- scikit
- Watson
- Tensorflow

What does the New AI do?

- **Discover patterns** in complex data.
- **Discover latent features** for embedding data.
- **Classify data** based on features.
- **Make predictions** based on data.
- **Recognize objects, faces, etc**., in images and video streams.
- **Analyze, generate and translate text**.
- **Learn to make near-optimal choices** in complex situations (e.g. games).
- **Build bots and agents** for specific applications.
- **Enable intelligent autonomous behavior** in robots, vehicles, etc.
- ……... and much more …

What does the New AI do?

All based on pattern recognition

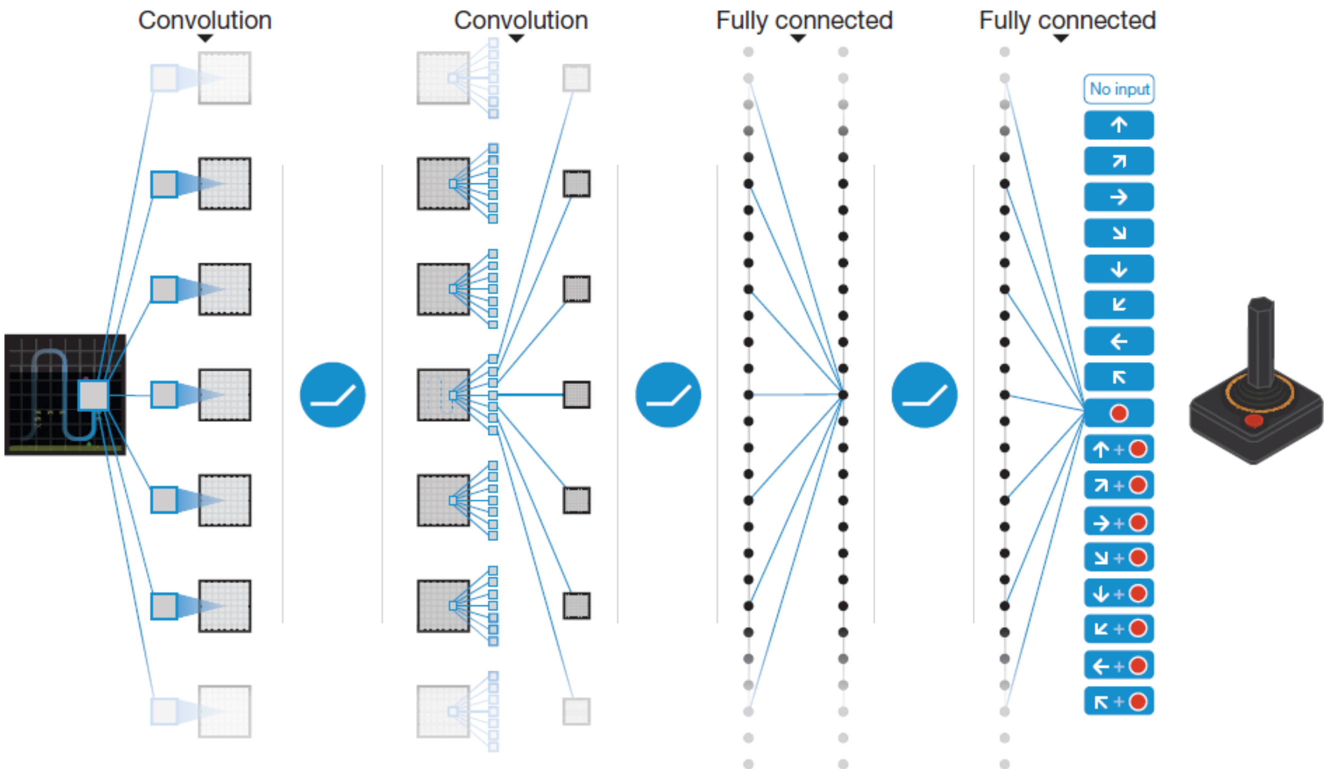
- **Discover patterns** in complex data.
- **Discover latent features** for embedding data.
- **Classify data** based on features.
- **Make predictions** based on data.
- **Recognize objects, faces, etc**., in images and video streams.
- **Analyze, generate and translate text**.
- **Learn to make near-optimal choices** in complex situations (e.g. games).
- **Build bots and agents** for specific applications.
- **Enable intelligent autonomous behavior** in robots, vehicles, etc.
- ……… and much more ….
- ……… **But is this sufficient for intelligence?**

The New AI and Its Successes

Deep learning and related methods have achieved spectacular successes!

Human-level control through deep reinforcement learning

Volodymyr Mnih[1]*, Koray Kavukcuoglu[1]*, David Silver[1]*, Andrei A. Rusu[1], Joel Veness[1], Marc G. Bellemare[1], Alex Graves[1], Martin Riedmiller[1], Andreas K. Fidjeland[1], Georg Ostrovski[1], Stig Petersen[1], Charles Beattie[1], Amir Sadik[1], Ioannis Antonoglou[1], Helen King[1], Dharshan Kumaran[1], Daan Wierstra[1], Shane Legg[1] & Demis Hassabis[1]

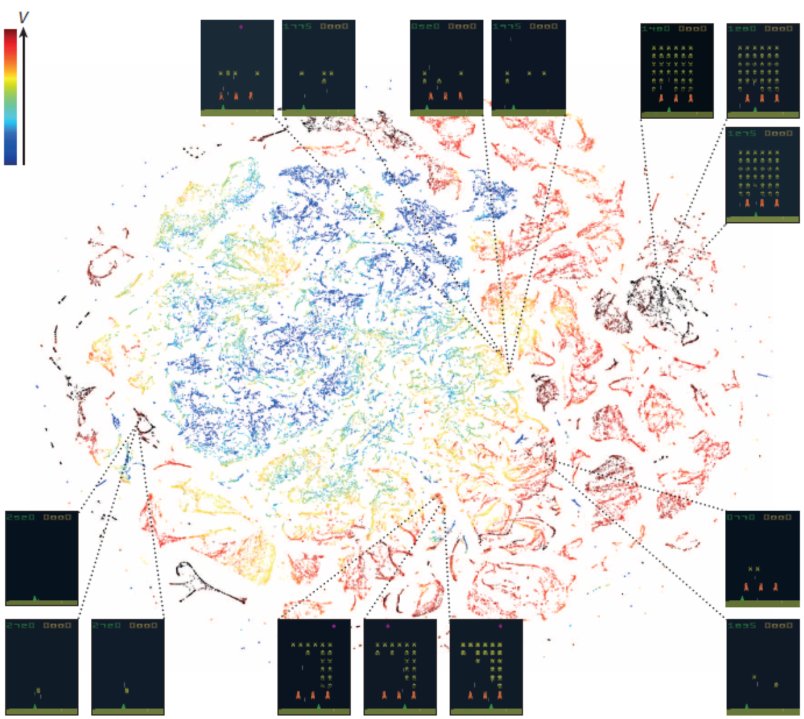


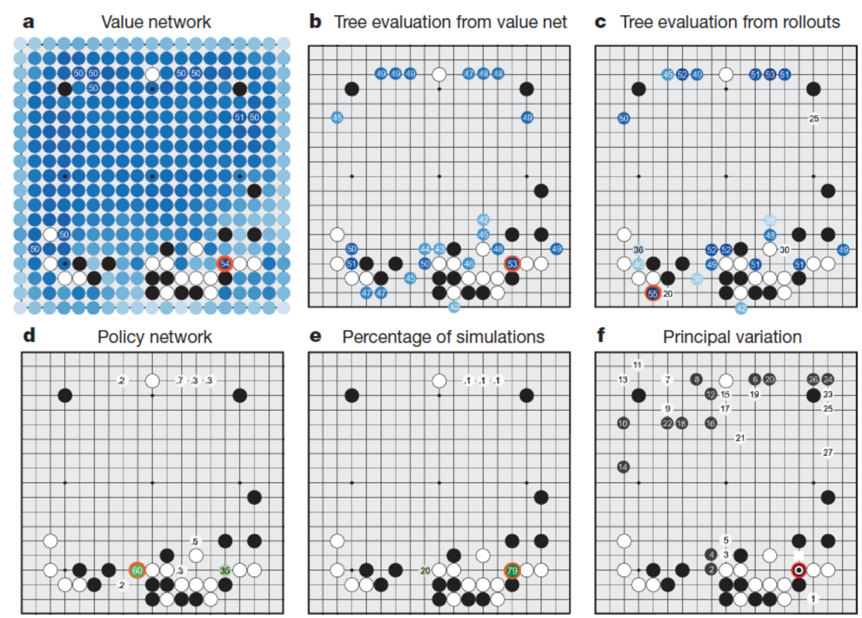
Human-level control through deep reinforcement learning

Volodymyr Mnih[1]*, Koray Kavukcuoglu[1]*, David Silver[1]*, Andrei A. Rusu[1], Joel Veness[1], Marc G. Bellemare[1], Alex Graves[1], Martin Riedmiller[1], Andreas K. Fidjeland[1], Georg Ostrovski[1], Stig Petersen[1], Charles Beattie[1], Amir Sadik[1], Ioannis Antonoglou[1], Helen King[1], Dharshan Kumaran[1], Daan Wierstra[1], Shane Legg[1] & Demis Hassabis[1]

Mastering the game of Go with deep neural networks and tree search

David Silver[1]*, Aja Huang[1]*, Chris J. Maddison[1], Arthur Guez[1], Laurent Sifre[1], George van den Driessche[1], Julian Schrittwieser[1], Ioannis Antonoglou[1], Veda Panneershelvam[1], Marc Lanctot[1], Sander Dieleman[1], Dominik Grewe[1], John Nham[2], Nal Kalchbrenner[1], Ilya Sutskever[2], Timothy Lillicrap[1], Madeleine Leach[1], Koray Kavukcuoglu[1], Thore Graepel[1] & Demis Hassabis[1]



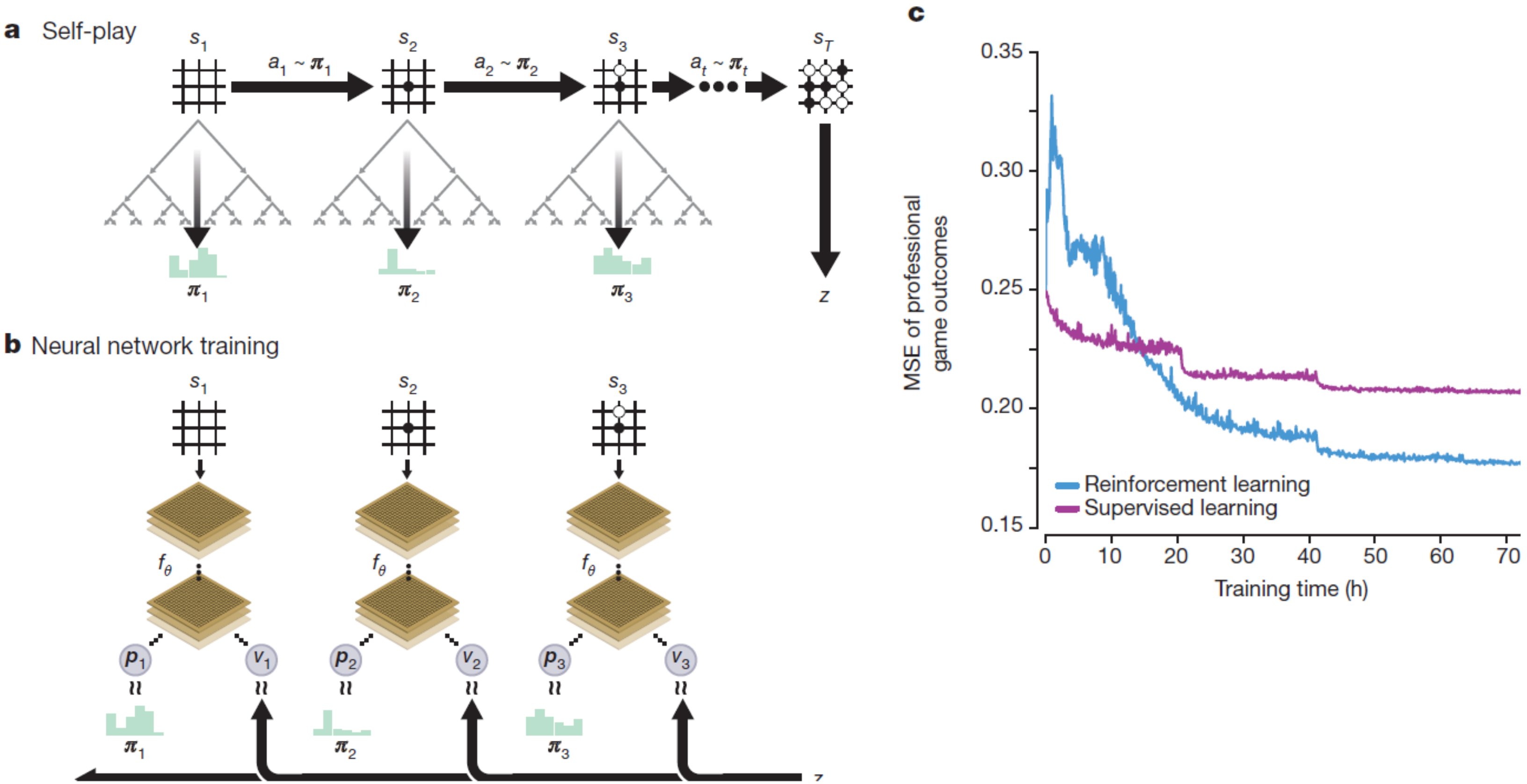


Silver et al. (2016) Nature 529: 484 - 491

Mnih et al. (2015) Nature 518: 529 - 533

Mastering the game of Go without human knowledge

David Silver[1]*, Julian Schrittwieser[1]*, Karen Simonyan[1]*, Ioannis Antonoglou[1], Aja Huang[1], Arthur Guez[1], Thomas Hubert[1], Lucas Baker[1], Matthew Lai[1], Adrian Bolton[1], Yutian Chen[1], Timothy Lillicrap[1], Fan Hui[1], Laurent Sifre[1], George van den Driessche[1], Thore Graepel[1] & Demis Hassabis[1]



GPT2

https://openai.com/blog/better-language-models/

In a shocking finding, scientist discovered a herd of unicorns living in a remote, previously unexplored valley, in the Andes Mountains. Even more surprising to the researchers was the fact that the unicorns spoke perfect English.

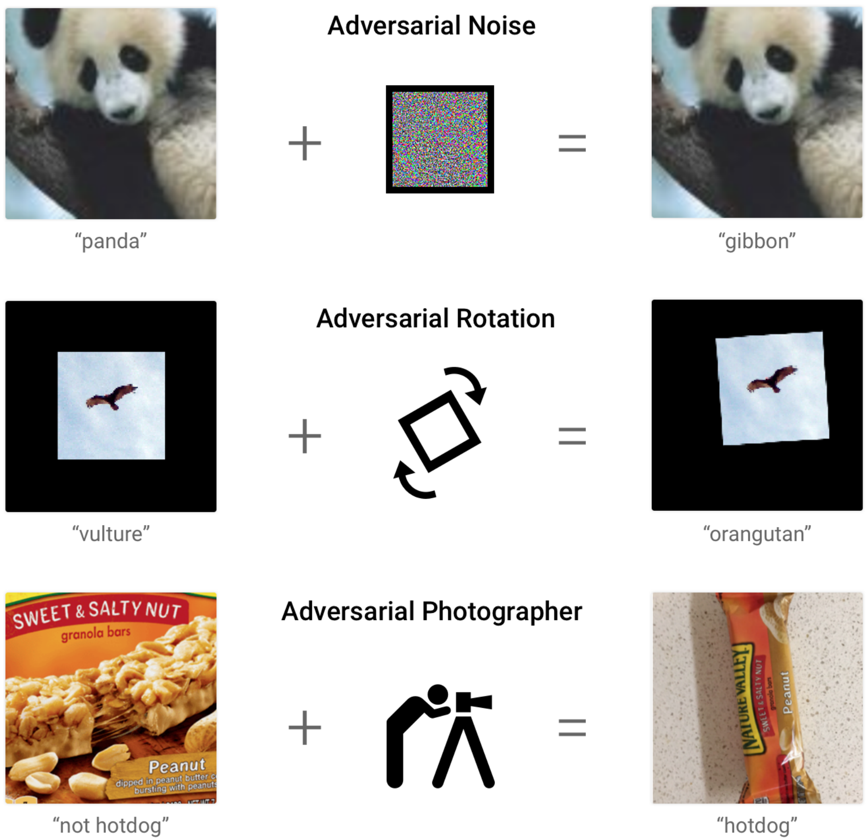
The scientist named the population, after their distinctive horn, Ovid's Unicorn. These four-horned, silver-white unicorns were previously unknown to science.
Now, after almost two centuries, the mystery of what sparked this odd phenomenon is finally solved.
Dr. Jorge Pérez, an evolutionary biologist from the University of La Paz, and several companions, were exploring the Andes Mountains when they found a small valley, with no other animals or humans. Pérez noticed that the valley had what appeared to be a natural fountain, surrounded by two peaks of rock and silver snow.
Pérez and the others then ventured further into the valley. "By the time we reached the top of one peak, the water looked blue, with some crystals on top," said Pérez.
Pérez and his friends were astonished to see the unicorn herd. These creatures could be seen from the air without having to move too much to see them – they were so close they could touch their horns.
While examining these bizarre creatures the scientists discovered that the creatures also spoke some fairly regular English. Pérez stated, "We can see, for example, that they have a common 'language,' something like a dialect or dialectic."
Dr. Pérez believes that the unicorns may have originated in Argentina, where the animals were believed to be descendants of a lost race of people who lived there before the arrival of humans in those parts of South America.
While their origins are still unclear, some believe that perhaps the creatures were created when a human and a unicorn met each other in a time before human civilization. According to Pérez, "In South America, such incidents seem to be quite common."
However, Pérez also pointed out that it is likely that the only way of knowing for sure if unicorns are indeed the descendants of a lost alien race is through DNA. "But they seem to be able to communicate in English quite well, which I believe is a sign of evolution, or at least a change in social organization," said the scientist.

So have we achieved “real” AI already?

So have we achieved "real" AI already?

Apparently not!
Even sophisticated AI programs are fooled easily by things that would not fool any intelligent animal.

But are we at least headed in the right direction?



https://ai.googleblog.com/2018/09/introducing-unrestricted-adversarial.html

We have taken important steps, but what we have achieved so far is really:
Augmented Intelligence: Using machine learning to amplify human intelligence
Intelligent Analytics: Using machine learning to find patterns in big data.
Intelligent Automation: Systems capable of making intelligent decisions.

Machine Learning vs Natural Intelligence?

Machine Learning

- Needs a lot of data
- Needs many iterations through the data.
- Data needs off-line storage.
- Trained for specific tasks with little task generalization.
- Driven by specific/explicit goals
- Converges to a function and stops.
- Learning decreases learning potential.
- Subject to catastrophic forgetting when capacity is exceeded.

Natural Intelligence

- Needs very little data
- Learns in a few trials – or even a single trial.
- Almost all data is on-line.
- Versatile - acquires task-independent capabilities.
- Driven by internal motivations.
- Performs lifelong learning.
- Learning increases learning potential.
- Builds deeper layers of capability over its lifetime.

Current AI:

- Trains complex systems for ***specific*** (human) purposes.
- Starts with naïve systems and hopes that they can infer complex solutions from lots of training on lots of data.
- Tries to ensure that the systems are completely reliable.
- Tries to ensure that the systems function exactly as we desire.
- Looks for systems whose outcomes can be explained causally.

Meanwhile, natural intelligent systems:

- Have versatile, general intelligence that can quickly train itself for unforeseen tasks.
- Rely on priors configured by evolution and development to learn rapidly with very little data.
- Are not completely reliable.
- Often have unpredictable behavior.
- Often behave in ways that cannot be explained – even by themselves.

AI is still missing something → **Artificial General Intelligence (AGI)**

Five Fallacies (?) of AI (among many)

- **Program-Processor Duality:** General intelligence (mind) can be implemented as a set of programs on a sufficiently powerful generic processor (body).
- **Prescriptive Autonomy:** It is possible to build truly intelligent systems that will always do what we want them to do.
- **Functional Reductionism:** Intelligence is a collection of specific abilities with pre-defined purposes that can be understood individually and implemented separately.
- **Rationality:** The goal of intelligence is to make optimal decisions.
- **Shallow Adaptation:** Intelligence can be learned from scratch by a sufficiently complex agent.

The Barrier of Meaning

Can AI that learns mainly by pattern recognition ever get to meaning?

Where does meaning come from?

Experience?
Observation?
Action?
Thought?
Imagination?

…. or is it enough to just see patterns and link them to consequences?

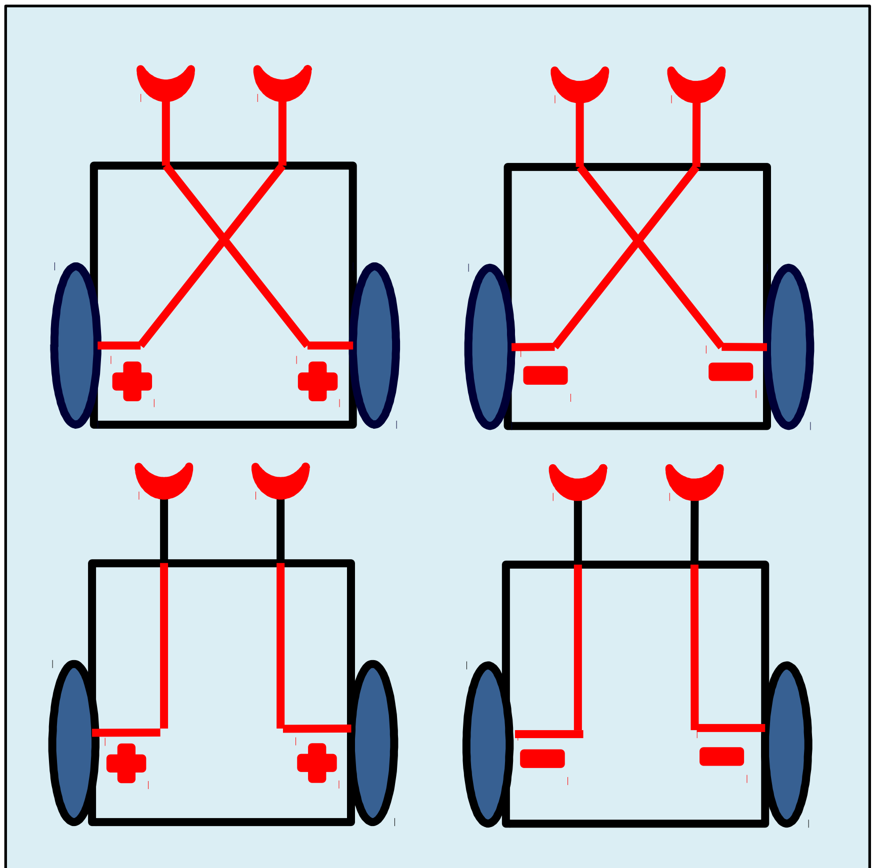
See "Artificial Intelligence Hits the Barrier of Meaning" by Melanie Mitchell

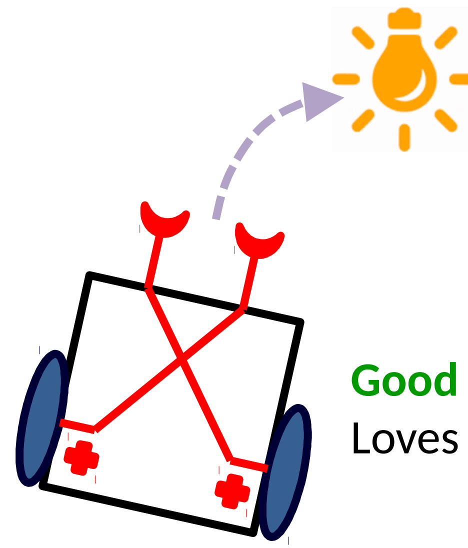
How much brain does a mind really need?

The Moral Psychology of Brainless Robots

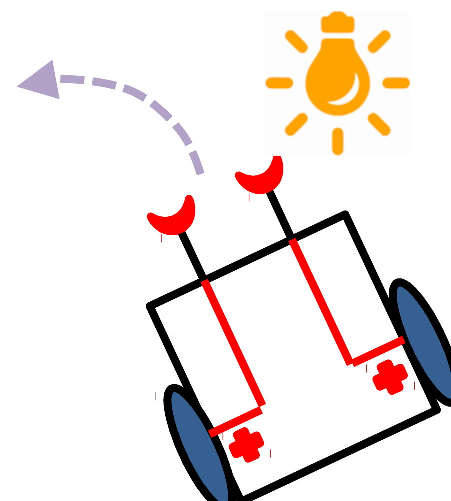
Braitenberg Vehicles

V. Braitenberg (1986) Vehicles: Experiments in Synthetic Psychology, MIT Press.





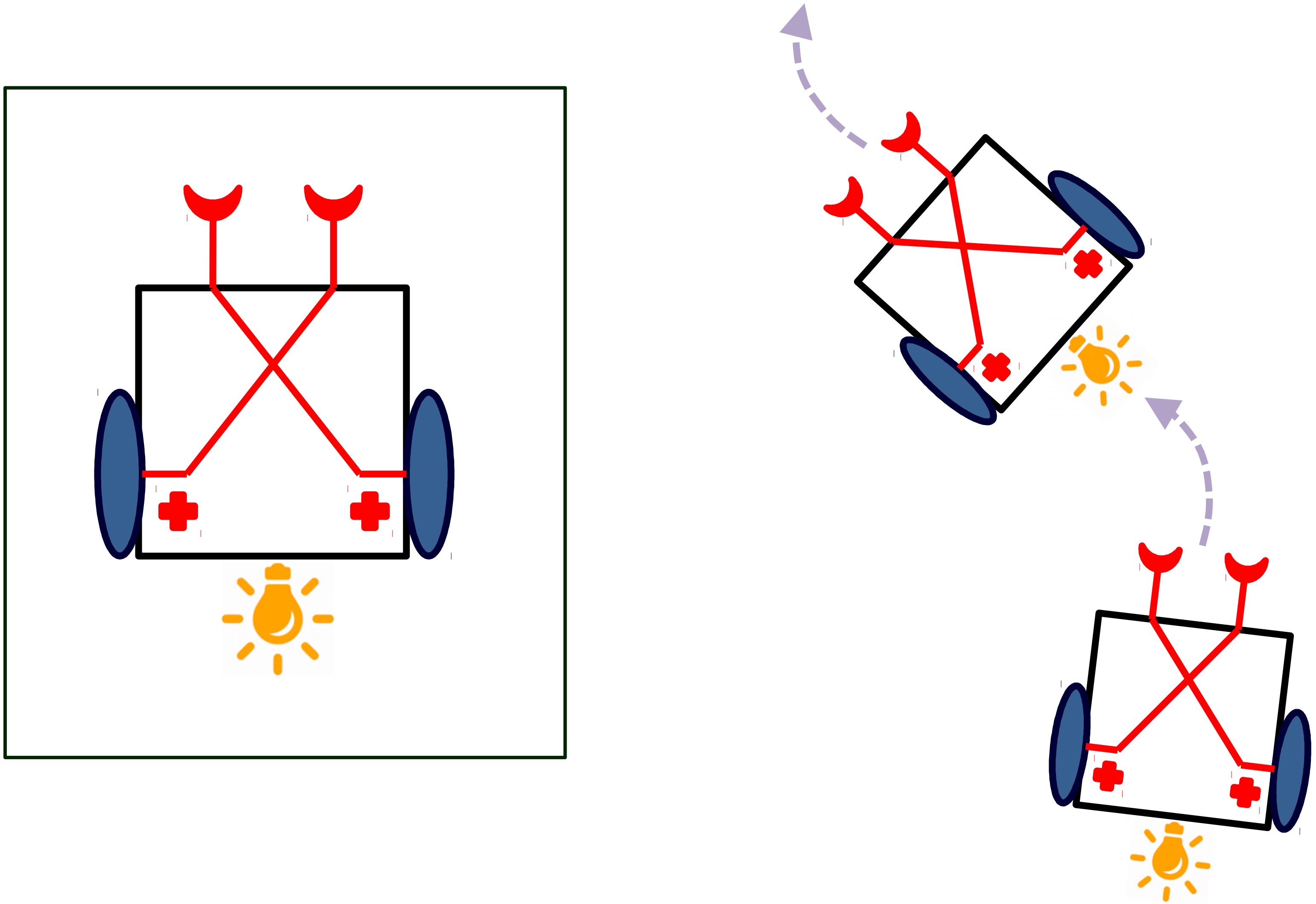
Good Robot: Loves light



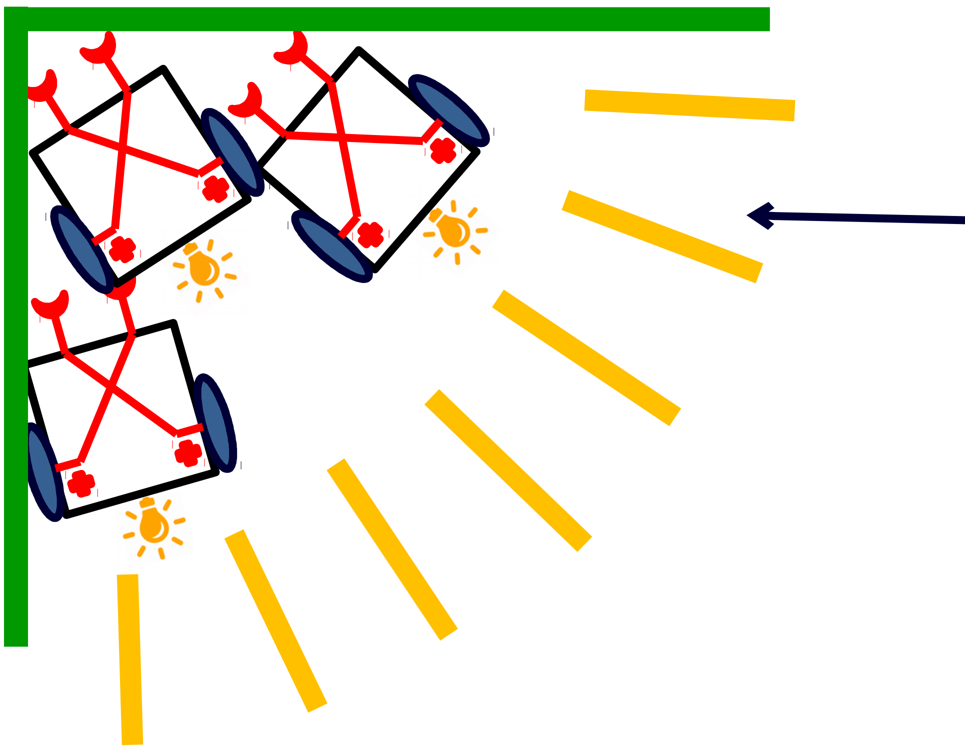
Evil Robot: Prefers darkness

Watch https://www.youtube.com/watch?v=A-fxij3zM7g

Tail-Chasing "Good" Robots

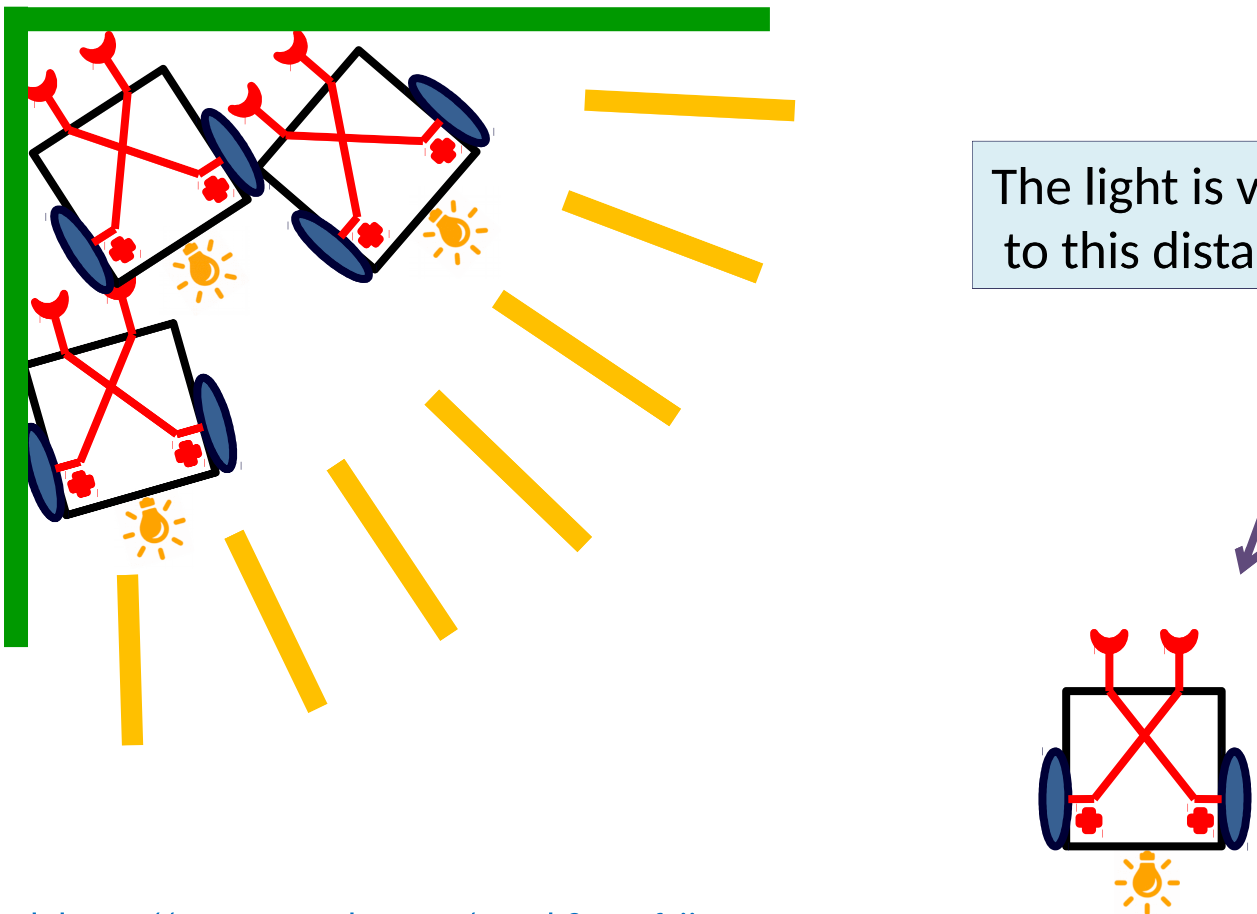


A Touching Tale of Robots



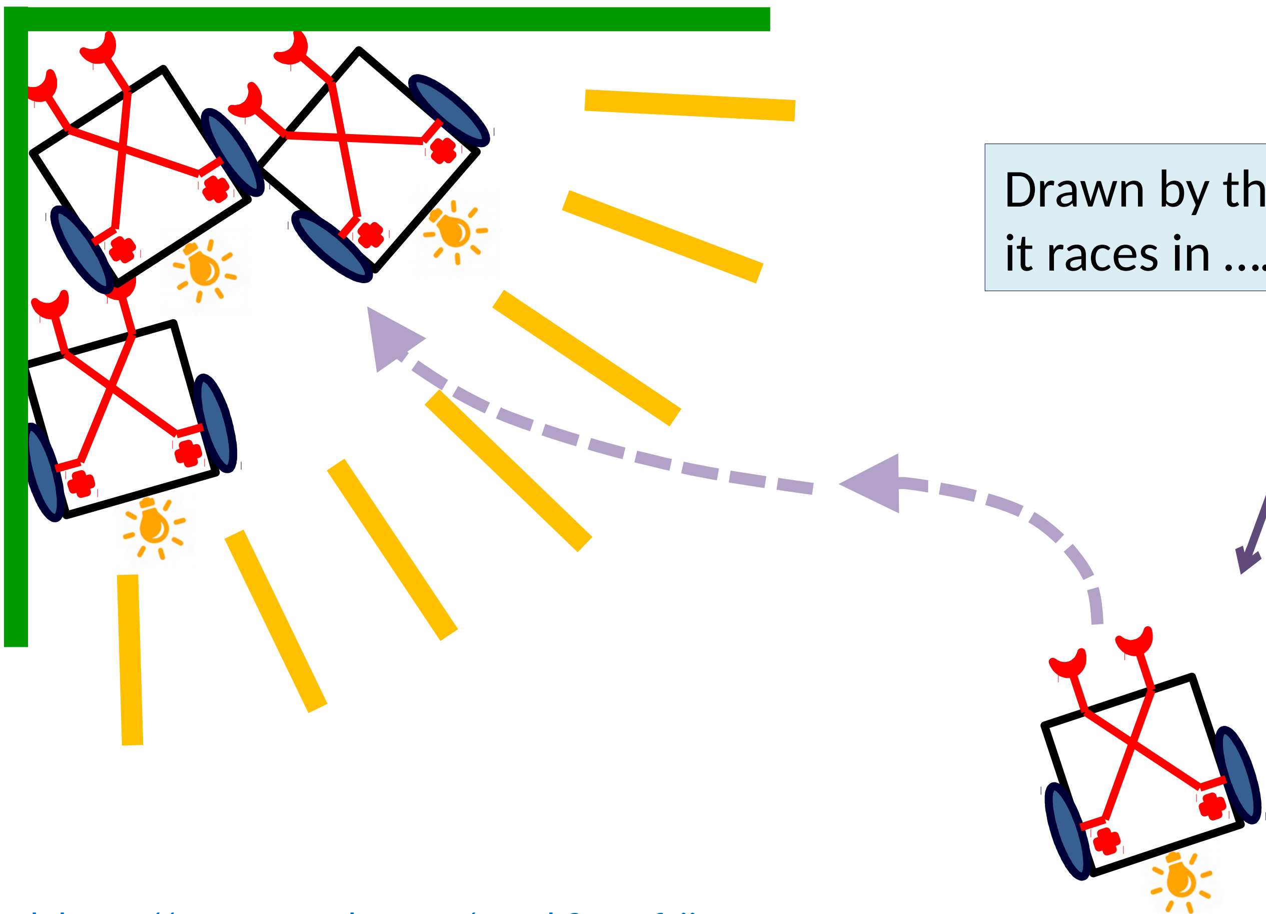
As a result of chasing others' tails, these robots are stuck in a corner, glowing like a beacon!

Watch https://www.youtube.com/watch?v=A-fxij3zM7g



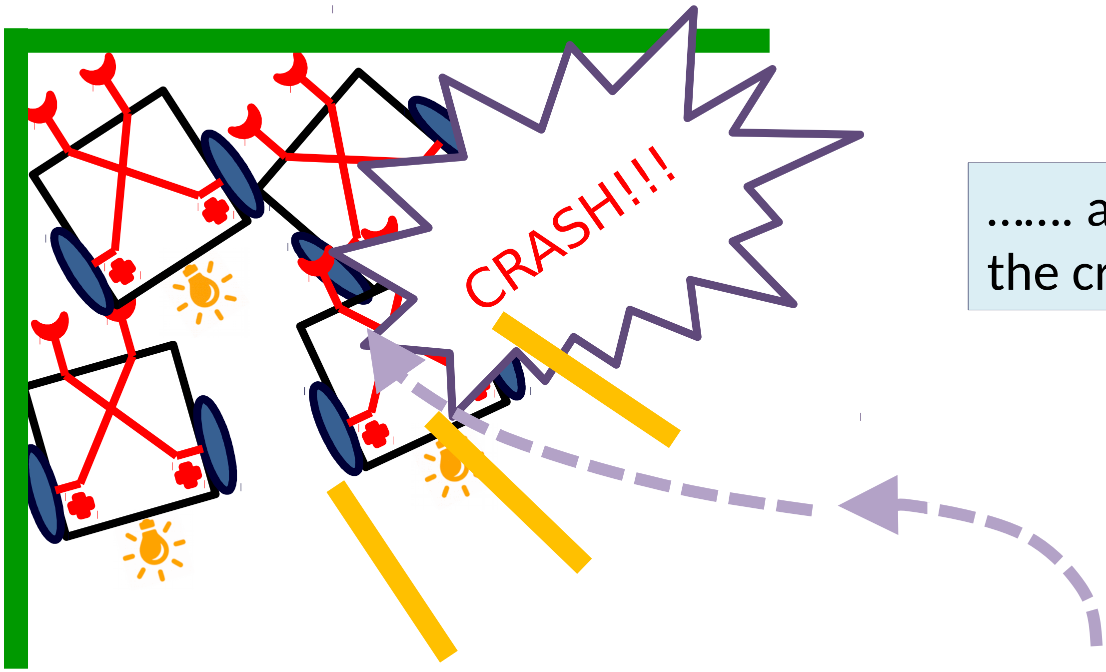
The light is very attractive to this distant robot!

Watch https://www.youtube.com/watch?v=A-fxij3zM7g



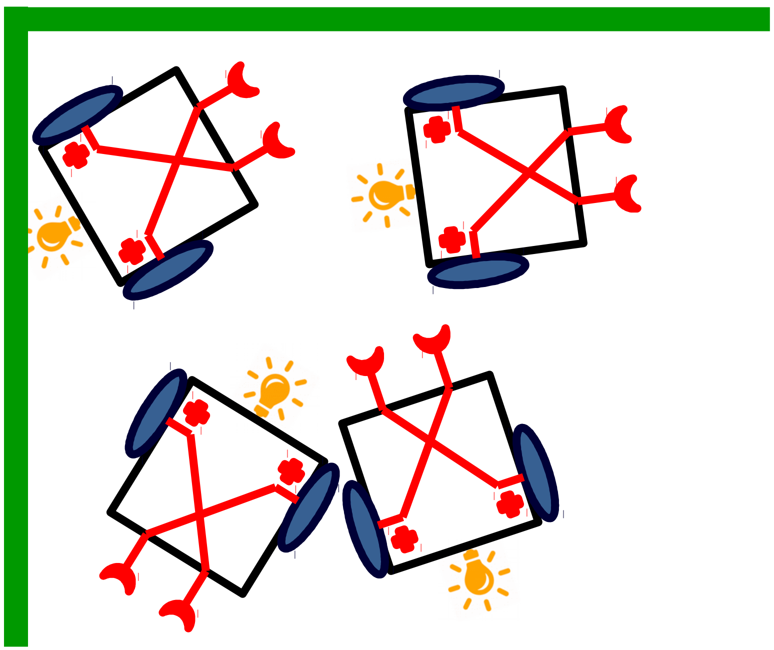
Drawn by the strong light, it races in …….

Watch https://www.youtube.com/watch?v=A-fxij3zM7g



……. and crashes into the crowd!

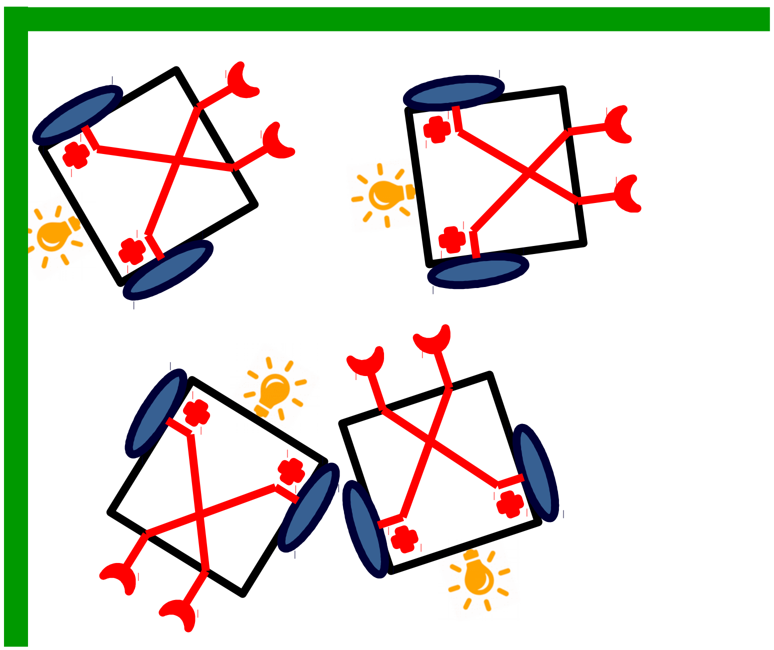
Watch https://www.youtube.com/watch?v=A-fxij3zM7g



Now everyone is free!

What a helpful "good" robot!

Watch https://www.youtube.com/watch?v=A-fxij3zM7g



Now everyone is free!

What a helpful "good" robot!

How much "brain" was needed for this "thoughtful" rescue?

Could a lot of complex mental function emerge in a similar way?

Watch https://www.youtube.com/watch?v=A-fxij3zM7g

Braitenberg vehicles – and many other similar examples – tell us that:

- Complex behaviors can emerge ***spontaneously*** from the ***interaction*** of simpler behaviors in connected components.
- No explicit planning or control need underlie behaviors that appear planned and controlled.
- There is no sense-think-act cycle: It is one seamless process.
- “Purpose” is often in the mind of the observer.

Perhaps “mind” doesn’t always need much “brain” after all ……

……. so where does it come from?

Braitenberg vehicles – and many other similar examples – tell us that:

- Complex behaviors can emerge ***spontaneously*** from the ***interaction*** of simpler behaviors in connected components.
- No explicit planning or control need underlie behaviors that appear planned and controlled.
- There is no sense-think-act cycle: It is one seamless process.
- Teleology is mostly in the mind of the observer.

Perhaps "mind" doesn't always need much "brain" after all ……

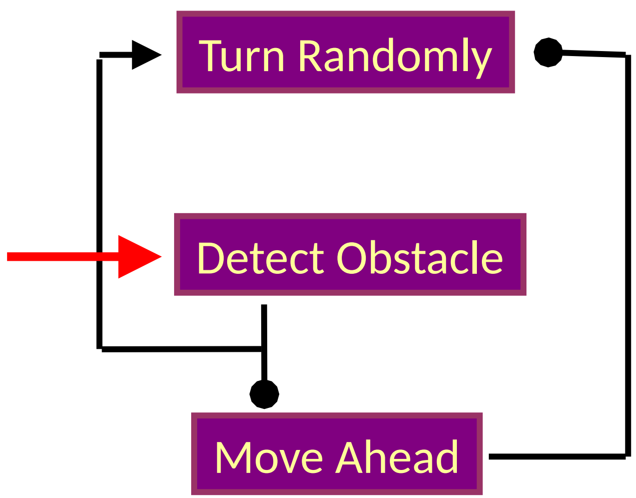
……. so where does it come from?

The Embodied Intelligence Model

Brain + Body = Active Artifact

Mind = Emergent Phenomenon

Subsumption: (Rodney Brooks)



Key Assumption:

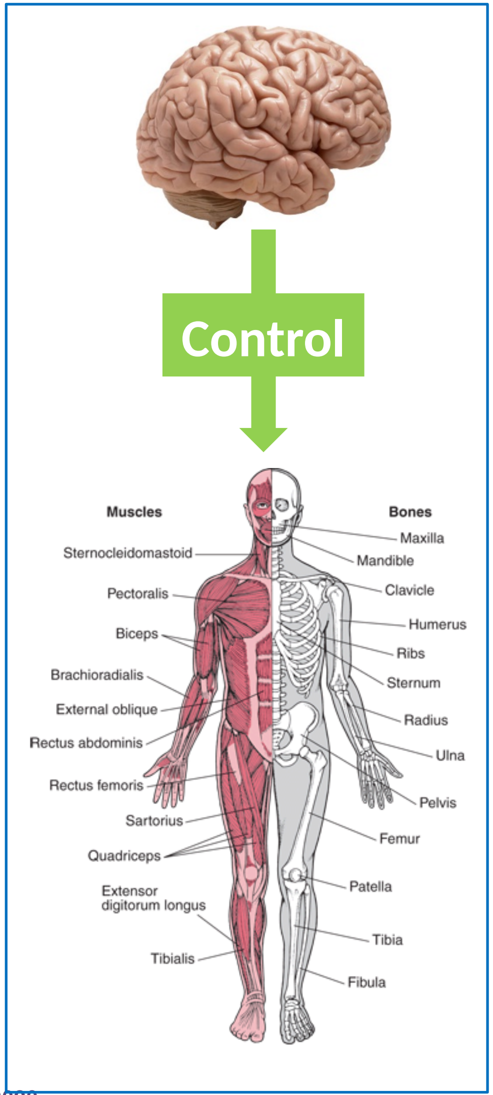
Intelligence arises from the interaction of many simpler concrete behaviors operating in parallel.

Approach:

- Implement the behaviors.
- Design their interaction network.

But is this enough?

The Standard View

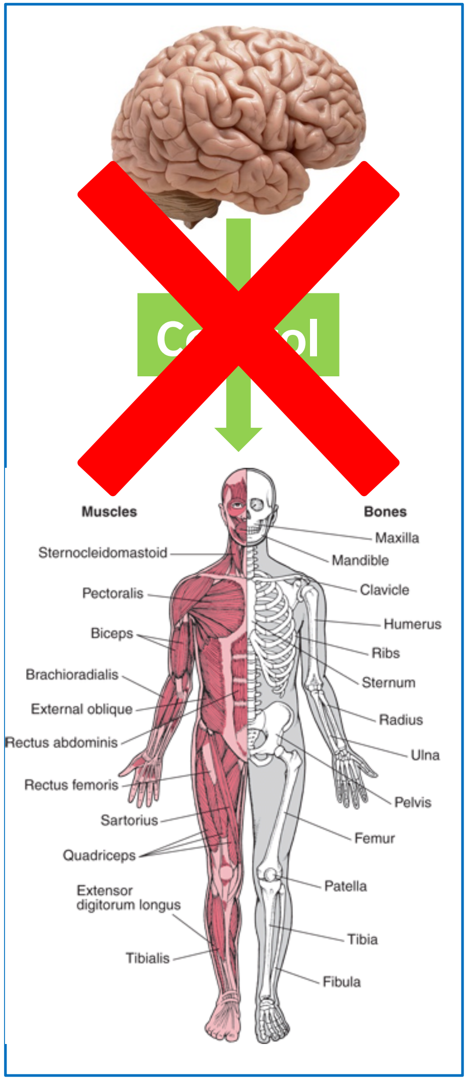


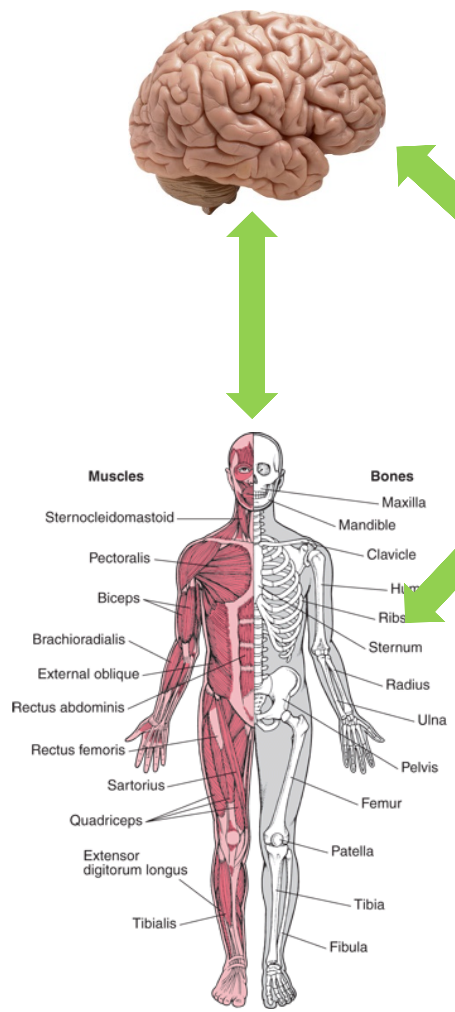
The brain:

1. Senses the environment,	**Sense**
2. Interprets it through perception, 3. Recognizes, 4. Remembers, 5. Reflects, 6. Decides what to do, 7. Plans a course of action,	**Think**
8. Creates control signals, 9. Sends them to muscles,	**Act**

………… to produce behavior.

Embodiment





The mind emerges from the interaction of these!

Embodiment and Computation

Physics ⟷ Computation

All material systems:

- Experience their environment.
- Change their state in response.
- Become reorganized over time.
- Change their environment.

because of their ***form*** – from the molecular level up!

This is true of both computers and brains – and all tissue….

….. and rock and water…

***All* matter computes!**

(That's why analog computers are possible)

Living systems also:

- ***Maintain internal organization*** through metabolism.
- ***Develop and grow*** over time.
- ***Reproduce*** to create new organisms similar to themselves.
- ***Evolve*** into new species.

Animals with nervous systems:

- ***Perceive*** their environment in complex ways.
- ***Build and use memories*** over time.
- ***Learn*** to improve their performance.
- ***Change their environment*** in controlled ways.

All this is ***still*** because of their **form** .....

... but the nervous system makes a ***BIG*** difference.

Living systems also:

- ***Maintain internal organization*** through metabolism.
- ***Develop and grow*** over time.
- ***Reproduce*** to create new organisms similar to themselves.
- ~~***Evolve*** into new~~ species.

Animals with nervous systems:

- ***Perceive*** their environment in complex ways.
- ***Build and use memories*** over time.
- ***Learn*** to improve their performance.
- ***Change their environment*** in controlled ways.

All this is ***still*** because of their **form** .....

... but the nervous system makes a ***BIG*** difference.

The brain is critical !!

What makes the nervous system special?

All cells in the body:

- Sense their environment
- Communicate with other cells
- Respond to signals
- Process information
- Adapt their behavior
- Act on their environment

But the nervous system has:

- A huge variety of specialized cells.
- Diversity of signaling and modulating chemicals.
- The ability to make distinct, cell-specific connections.

Which enables:

- Multiple methods of ***information representation at multiple scales***.
- Formation of ***specific, heterogeneous networks*** at many scales.
- ***Coordination of specific activity patterns*** over networks.
- ***Specific and associative adaptation*** of activity patterns.

Embodiment and the Brain

A system at the mercy of a fixed embodiment.

A system not at the mercy of a fixed embodiment.

Why not?

Because it has:

- A body that changes with experience.
- ***A brain that acts as an "adaptive switchboard" for the body.***

Main Lesson

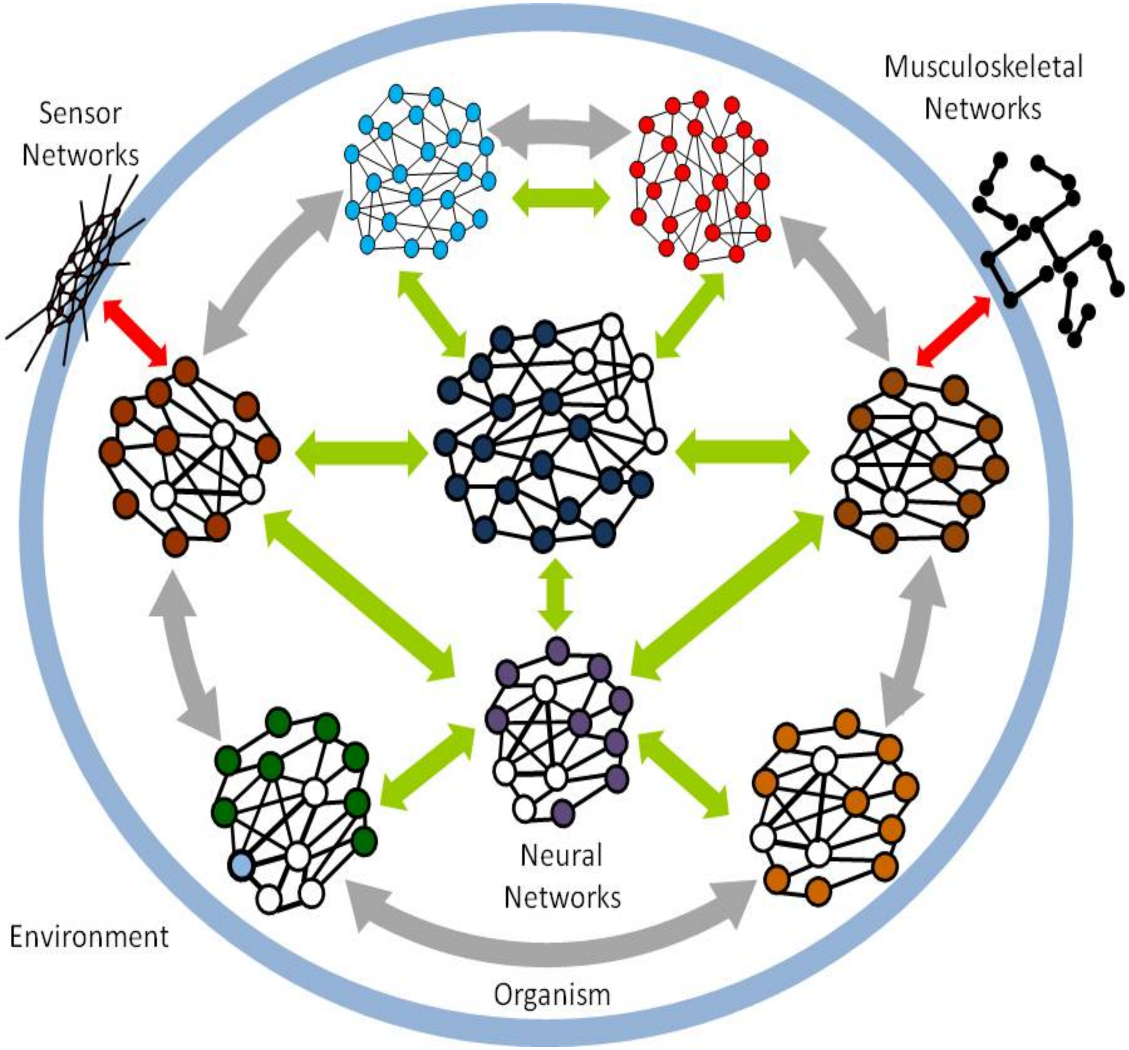
It's all in the networks!

Both the brain and the rest of the body are networks.

Fundamental Postulates

- The animal is a self-organizing, ***complex network of networks embedded*** in a hierarchy of complex networks..
- Information is represented and processed as ***dynamical patterns of activity*** over these networks.
- These networks are ***reconfigured continually*** by processes at ***many spatial and temporal scales***.
- All perception, cognition and action ***emerges*** from the activity of these networks.

The Animal as an Embedded Complex System



Implications for AI

- It is not enough just to create algorithms.
- It is not enough just to create artificial brains.
- AI ultimately requires creating artificial animals → **Robots, animats**.

Artificial Intelligence ⟷ Artificial Life

Algorithms and neural networks will play a big role .....
..... but the real action in AI will soon come from….

- Biomorphic robots.
- Biomorphic chips.
- Brain-machine interfaces.
- Brain implants.
- Mind-controlled prosthetics……..

Creating Animals with Complex Behaviors

Option 1: Engineering

Build a complex animal with pre-designed complex behaviors.

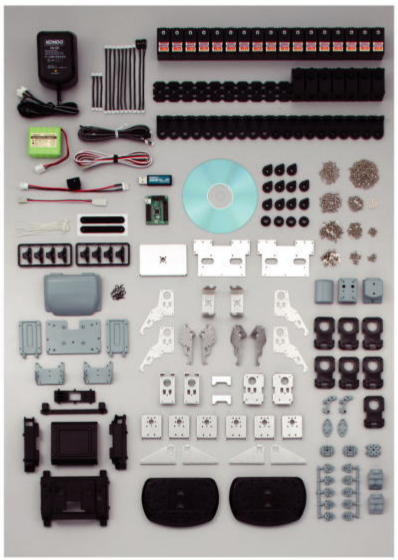
Won't work because:

- Impossible to anticipate useful complex behaviors.
- Impossible to think of every contingency.
- Impossible to design efficient behaviors.
- Cannot adapt to changing needs.
- Will not scale up to larger, more complex animals.

Creating Animals with Complex Behaviors

Option 2: Learning

Build a complex animal and have it learn complex behaviors.



Learn

This!

Won't work because:

- Too many degrees of freedom.
- Too many possibilities to explore during learning.
- Will not scale up to larger, more complex animals.

Creating Animals with Complex Behaviors

Option 3: Artificial EvoDevo

Gradually bootstrap to more complex animals by processes analogous to evolution and development.

- Start with simple animals that learn to do simple things.
- Produce increasingly complex animals in stages.
- At every stage, learn more complex behaviors by building on those learned in earlier stages.

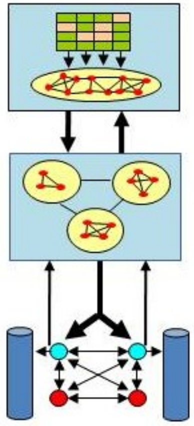
Will work because:

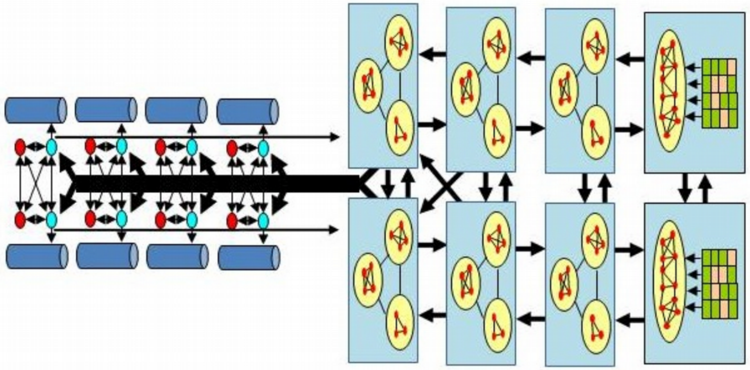
- Task at each stage is constrained and simple.
- Most bad choices are never considered.
- No limit to growth in principle.

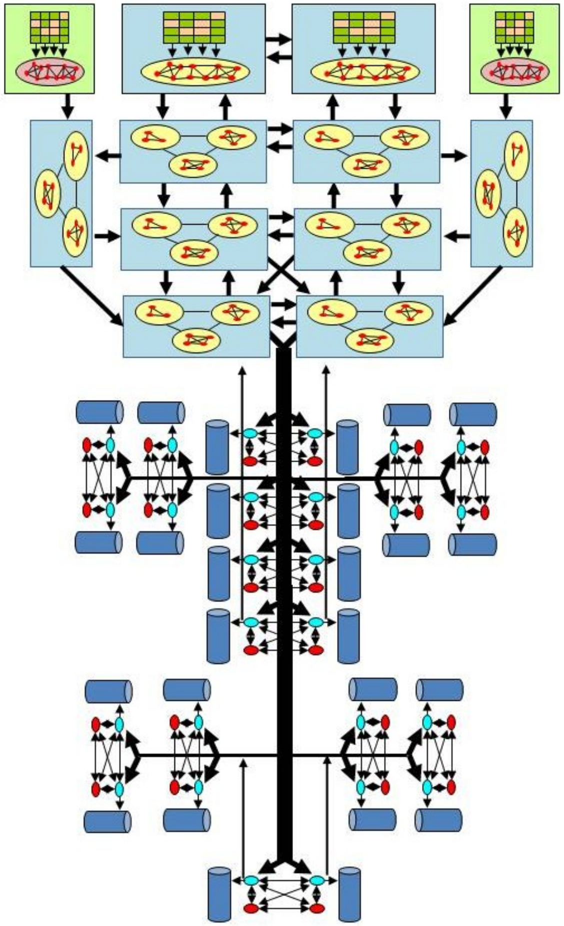
Key Enabler: Modularity

Modular construction is the key to efficient generation of complexity.

- Simplifies growth.
- Channels growth in useful directions.
- Makes growth more efficient.
- Increases stability and robustness.







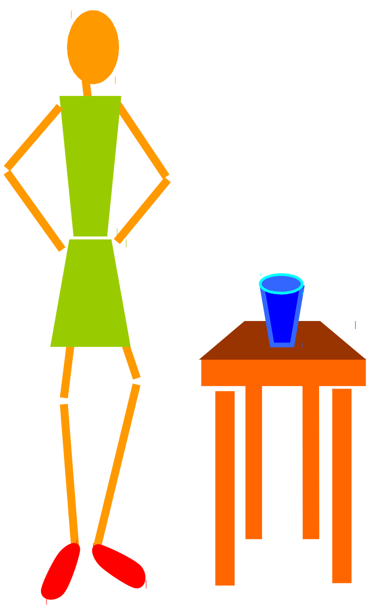
Learning to Behave

One possible approach for learning behaviors:

- Generate behaviors autonomously.
- Learn useful ones by reinforcement.
- Reject the others.

→ Repertoire of useful behaviors

But this works only if behavioral sampling is constrained and productive



List of possibilities:
- Stand and stare.
- Reach and pick up.
- Turn away.
- Kick the table.
- Say "hello" to the cup
- Make a face.
- Jump up and down.
- Scream.
- Dance.
- ….
- Recite Hamlet's monologue
- …..

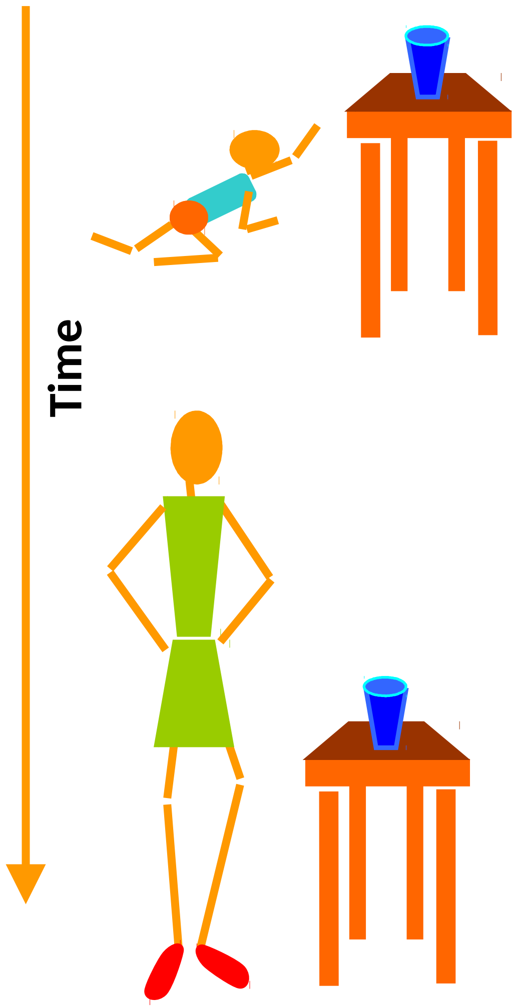
- Too many possibilities.
- Mostly irrelevant.

↓

Learning is not feasible!

Developmental Learning

Learning occurs in stages that match expanding choices with increasing control DoF.



- Simple embodiment
- Few choices.
- Simple actions.
- Few control degrees of freedom.

→

- Build elementary repertoire.
- Learn broad options.

↓

- More choices.
- More control DoF.
- More working memory.
- Mirror system.

→

- Construct complex sequences.
- Add to repertoire as chunks.

↓

- Even more choices.
- Even more control DoF.
- More complex repertoire.

→

- Construct more complex sequences.
- Add to repertoire as chunks.

At every stage, most choices are truncated

→ Much smaller option space → Learning!

What We Still Don't Understand

- Fast, reliable action from slow, noisy components.
- Rapid learning – often in a single experience.
- Rapid construction of complex voluntary movements.
- Volition and planning.
- Fluent thought and expression.
- Creativity.

And a whole lot more!

Even deep learning hasn't helped much!

The Path to AGI

Real AI will be:

- Embodied (brain + body, not just brain).
- Emergent (not driven by specific tasks/goals).
- Unsupervised + reinforced (with very fast RL).
- Developmental (learning in stages building on prior stages).
- Hierarchical (more complex functions by combination of less complex ones).
- Self-motivated (active learning).
- Generative (able to imagine and learn from imagination).
- Inherently abstractive and analogical (transfer learning across domains).
- Empathetic (have a theory of other minds).
- Heuristic (rooted in innate evolutionary priors).
- Non-parametric (not based on abstract prior models).

The Path to AGI

It will not use:

- Optimization / supervised learning.
- IID sampling from assumed distributions.
- Slow, iterative learning.
- Large amounts of data.
- Specific tasks.

AI will undergo a great transformation in its priorities and approaches over the next 5-10 years.

An important fact to keep in mind:

Creating AGI = Creating a new species with totally different biology!

Potentialities

The current version of AI is already enabling many new things:

- Smarter devices and appliances.
- The Internet of Things.
- Autonomous vehicles.
- Smart environments – smart homes, navigation systems, etc.
- Smart media.
- Improved services – law enforcement, healthcare, education, etc.
- Better consumer experience – e.g., recommender systems.
- Better financial analysis.
- Better cyber-security.
- Much better intelligence analysis.
- Smart military systems.
- …. etc.

Challenges

The Challenge of Autonomy:

- **Three levels of autonomy:** Operational; Decision-making; Goal-setting.
- Do we want a servant or a truly autonomous system?
- How do the system's goals align with ours?
- How do we keep the system from going rogue?

The Challenge of Bias:

- Data-induced biases – how much control do we have over data?
- Essential biases of human intelligence – some good, some bad!

The Challenge of Expainability/Verifiability:

- The model is ***opaque*** – data driven, deeply nonlinear.
- The model is ***abstract*** – no meaningful internal primitives.

The Challenge of Validation:

- Ground truth is often unavailable.
- The system is not fully controllable.

The Challenge of Values:

- Where does the system get its values?

Short-Term Hazards

- Loss of privacy and individual control.
- Algorithms imposing inhumane values.
- Hyper-rationality.
- Liability and responsibility in autonomous systems.
- Localized economic and social disruption.
- Exacerbation of economic inequality.
- Autonomous munitions.

Not Skynet!

Longer-Term Hazards

- **Alternative Intelligence:** Systems with general intelligence but with values very different from ours.
- Massive restructuring of societies and economies.
- Fundamental changes in the modes of governance and control.
- Exacerbation of economic inequality.

Not Skynet!

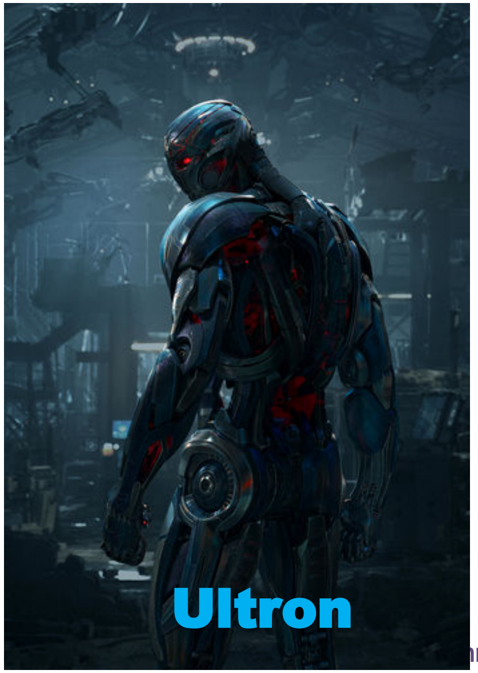
Some interesting thoughts on this here:
https://www.smithsonianmag.com/innovation/artificial-intelligence-future-scenarios-180968403/#kFwJygofLaeq8LHe.01

Long-Term Hazards

Long-Term Hazards

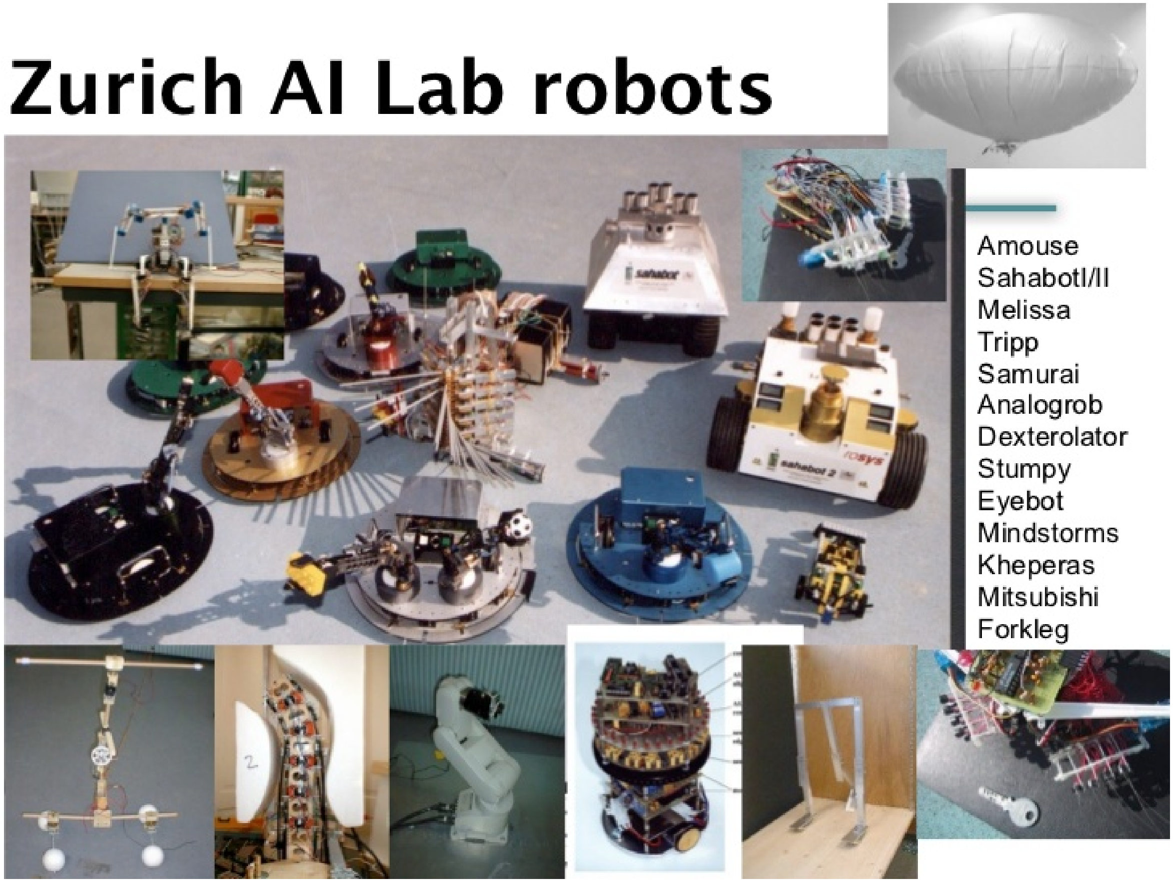
VS

VS



VS

These are the early ancestors of Mr. Data and Skynet .....



For lots of interesting stuff on embodied intelligence, check out:
http://www.slideshare.net/swissnexSF/embodied-intelligence-the-four-messages

The Shape of Things to Come

What we will do in this class

Define the basic concepts needed to understand and build intelligent systems.

Briefly review information processing in the brain.

Study ***neural network methods*** (and a few others) for the fundamental intelligent systems tasks such as classification, pattern recognition, clustering, and memory.

- Feed-forward networks with supervised learning.
- Unsupervised learning in neural networks (clustering and feature maps).
- Recurrent neural networks and associative memory.
- Introduction to deep learning.

- K-nearest neighbor algorithms.
- Bagging and boosting.
- K-means clustering.
- Etc.

For more general discussions on AI, look at these:

OPenAI on GPT2: https://openai.com/blog/better-language-models/

Tony Zador on learning from animal brains:
https://www.nature.com/articles/s41467-019-11786-6?fbclid=IwAR18H6BQ9Wl4PlI09a2pUmDnNAxU31q2xOMOQLtuZ7DVnUqXZ_M0Tk0J71w

David Deutch on doing AI right:
https://aeon.co/essays/how-close-are-we-to-creating-artificial-intelligence?fbclid=IwAR2mJRUas-g3_IgRVR9yxD8jltCmhbvhi2JaPPa08gKrGIhGTGm2UuFqkak

Groth, Nitzberg and Russell on AI algorithms:
https://www.wired.com/story/ai-algorithms-need-drug-trials/?fbclid=IwAR1__eJb4GCIiLw23_K0BVBxJX-TZxeF7MDVOHgz_Eg_35crLz8OB_nVwEA

Derek Thompson on A World Without Work:
https://www.theatlantic.com/magazine/archive/2015/07/world-without-work/395294/

….. and more

Melanie Mitchell on The Barrier of Meaning:
https://aaai.org/ojs/index.php/aimagazine/article/view/5259
https://www.nytimes.com/2018/11/05/opinion/artificial-intelligence-machine-learning.html

https://pdxscholar.library.pdx.edu/compsci_fac/194/

Questioning the direction of AI today:
https://www.technologyreview.com/2020/08/18/1007196/ai-research-machine-learning-applications-problems-opinion/

On the importance of mistakes for intelligence:
https://www.newyorker.com/science/maria-konnikova/why-are-babies-so-dumb-if-humans-are-so-smart

General Books

M. Mitchell, “Artificial Intelligence: A Guide for Thinking Humans”, Farrar, Strauss, and Giroux, 2019.

G. Marcus and E. Davis, “Rebooting AI: Building Artificial Intelligence We Can Trust”, Pantheon, 2019.

M. Tegmark, “Life 3.0: Being Human in the Age of Artificial Intelligence”, Knopf, 2017.

N. Bostrom, “Superintelligence: Paths, Dangers, Strategies”, Oxford University Press, 2016

Textbooks

C.C. Aggarwal, “Neural Networks for Deep Learning: A Textbook”, Springer, 2018.

E. Alpaydin, “Introduction to Machine Learning”, MIT Press, 2010.

K.P. Murphy, “Machine Learning: A Probabilistic Perspective”, MIT Press, 2012

I. Goodfellow, Y. Bengio and A. Courville, “Deep Learning”, MIT Ptress, 2016.

To read some of my opinions on AI:

http://barbarikon.blogspot.com/2014/04/understanding-artificial-intelligence.html

https://www.3quarksdaily.com/3quarksdaily/2015/08/fearing-artificial-intelligence.html

https://www.3quarksdaily.com/3quarksdaily/2017/03/artificial-stupidity.html

https://www.3quarksdaily.com/3quarksdaily/2019/02/what-ai-fails-to-understand-for-now.html