EE488 Special Topics in EE
<Deep Learning and AlphaGo>

Sae-Young Chung

Lecture 2

August 30, 2017

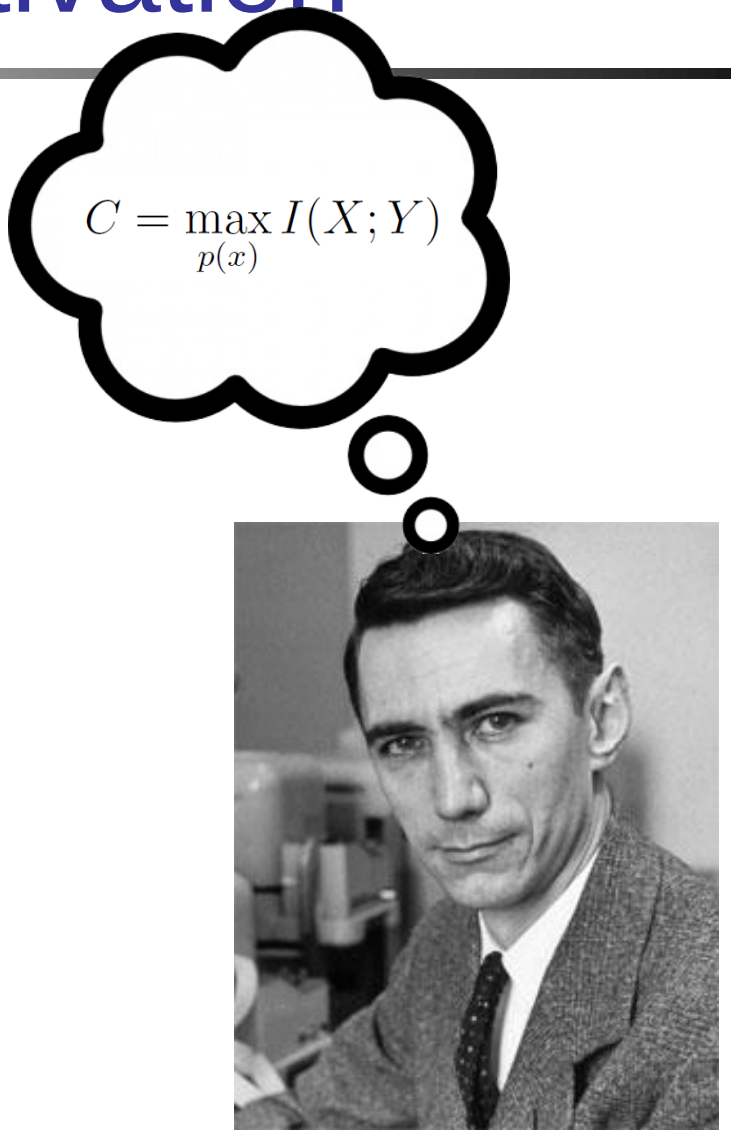
KAIST

Learning Methods for AI

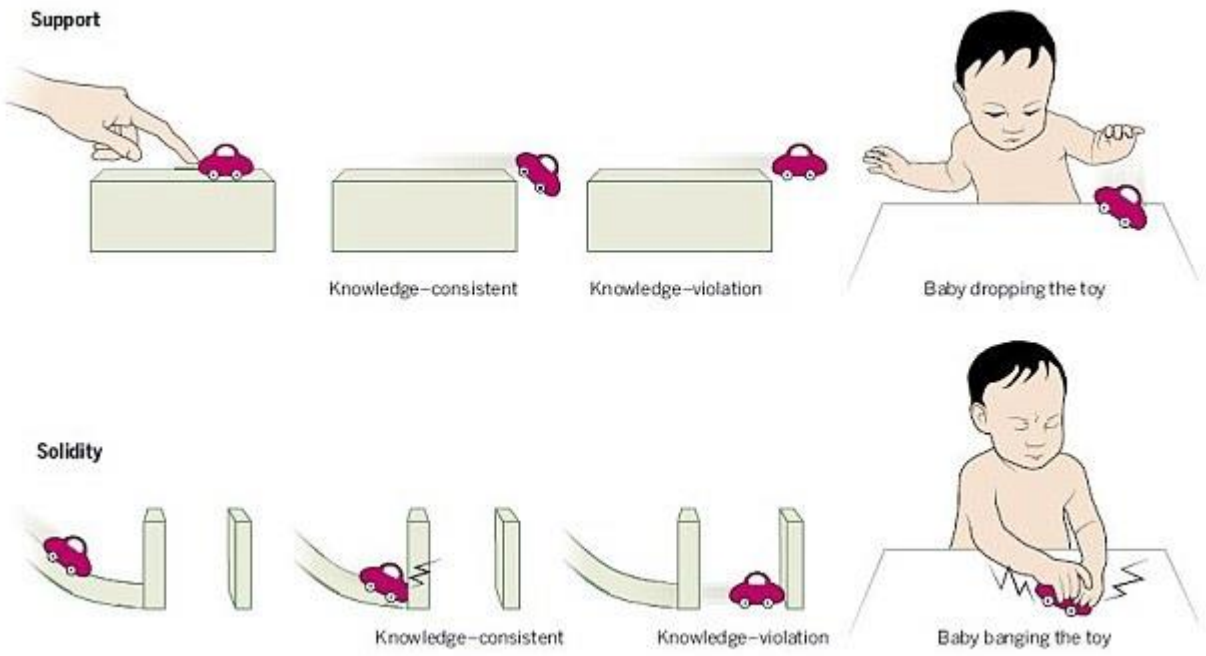
- Supervised learning
- Unsupervised learning
- Semi-supervised learning
- Reinforcement learning
  - Value-based reinforcement learning
  - Policy-based reinforcement learning
  - Actor-critic method
  - Curiosity-driven exploration
  - Intrinsic motivation
  - Empowerment
  - Hierarchical reinforcement learning
  - …
- Transfer learning
- Curriculum (developmental) learning
- Online learning, batch learning
- Learning to learn
- Active learning
- …
- But, we need more

Reinforcement Learning without Reward

Intrinsic Motivation



Active Learning



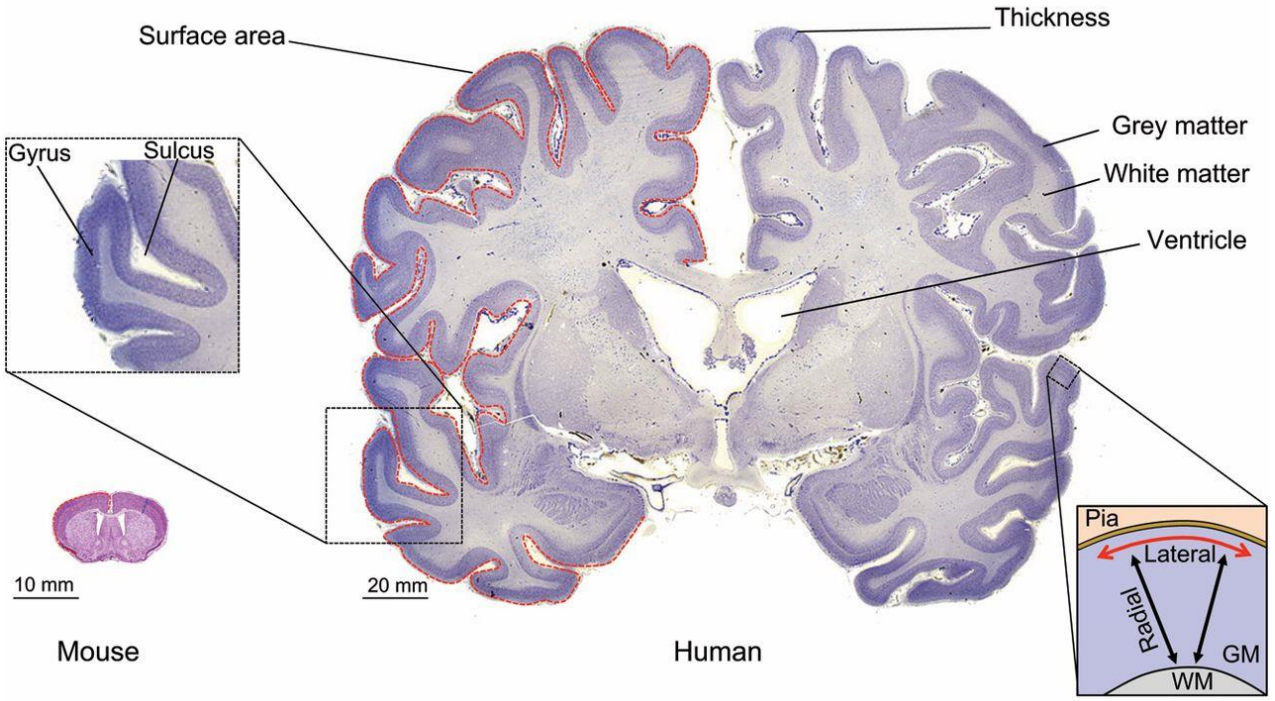
A. E. Stahl, L. Feigenson, Observing the unexpected enhances infants' learning and exploration, Science 03 Apr 2015.

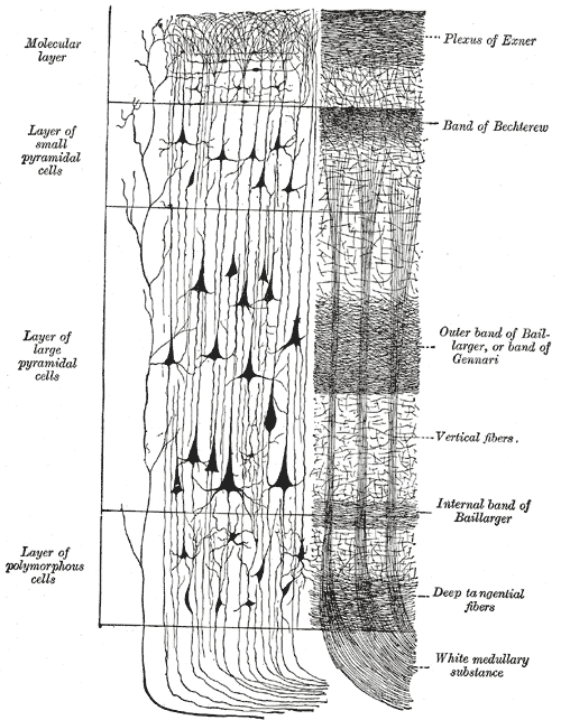
Curiosity-driven Exploration

How to Quantify?

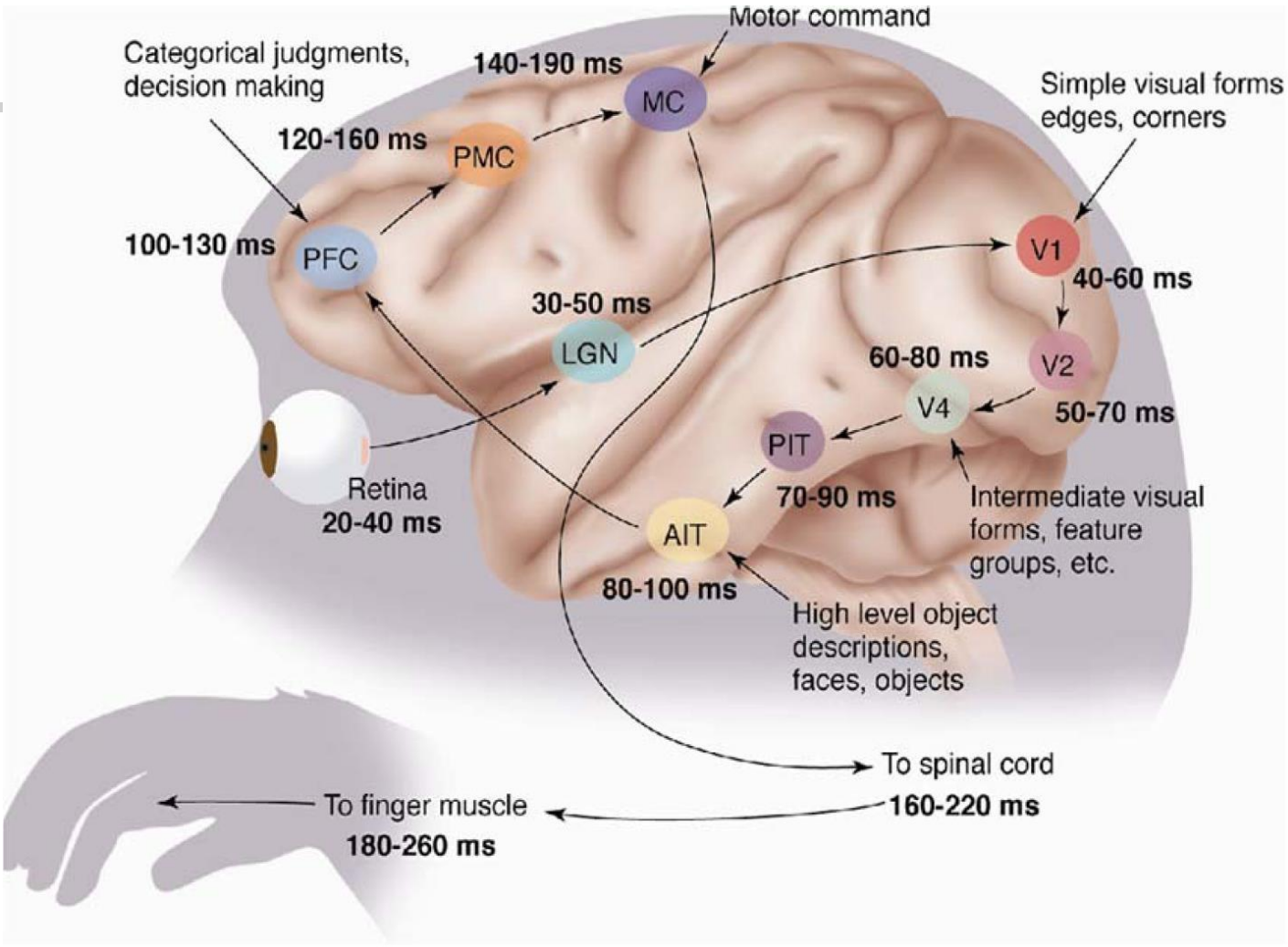
- How to quantify
  - Intrinsic motivation
  - Curiosity
  - Empowerment
  - …
- Information theory provides answers
  - Entropy
  - Cross entropy
  - Causal entropy
  - Mutual information
  - Conditional mutual information
  - Directed mutual information
  - Relative entropy, KL divergence
  - …

Neocortex of Human Brain



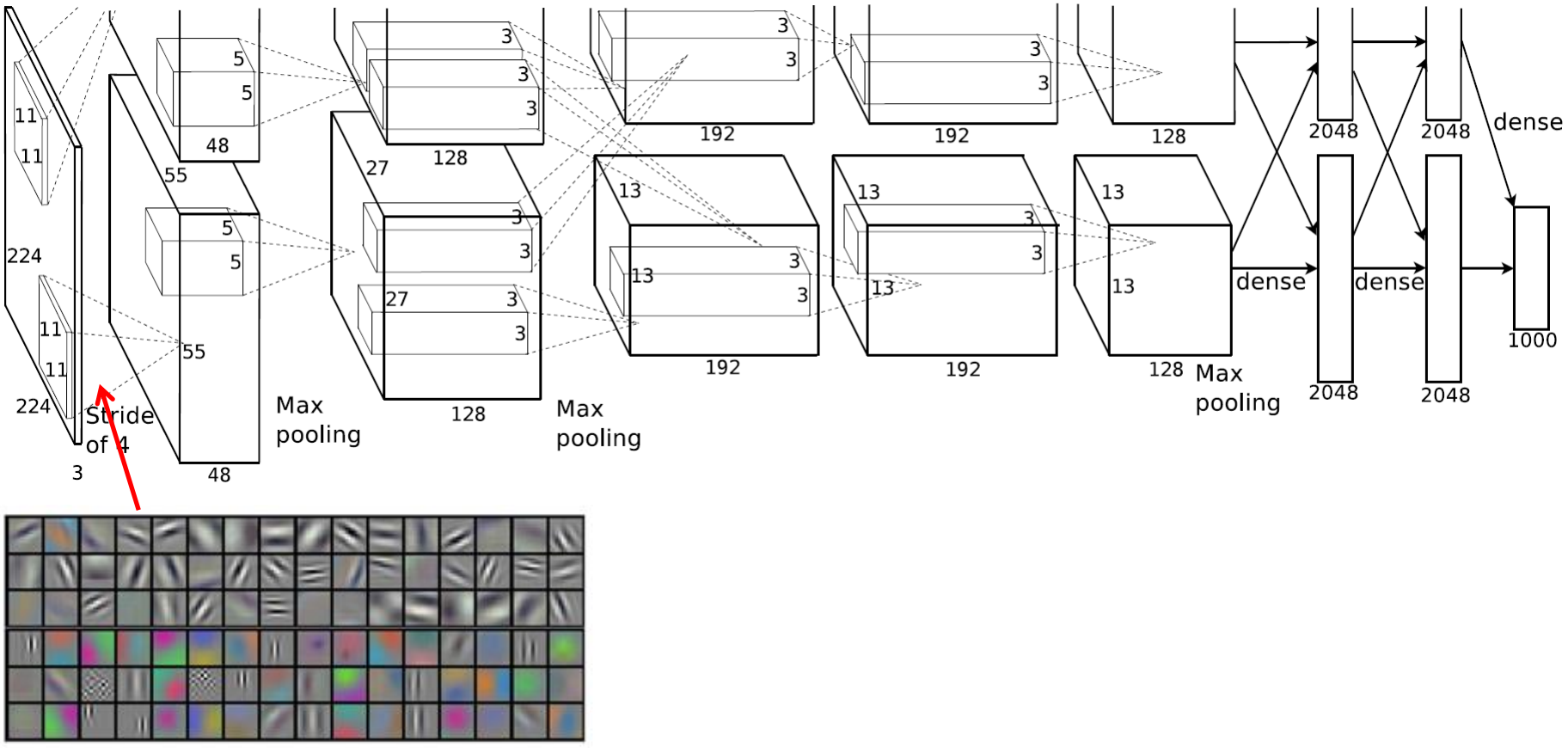


KAIST



Simon Thorpe and Michele Fabre-Thorpe, Seeking Categories in the brain, Science 2001

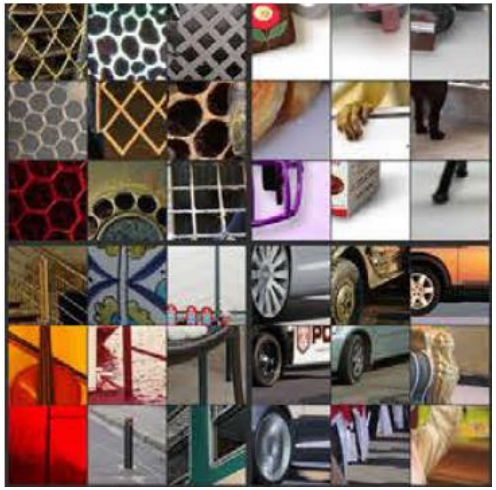

Nurture and Thrive
for
Busy Mom's
Helper

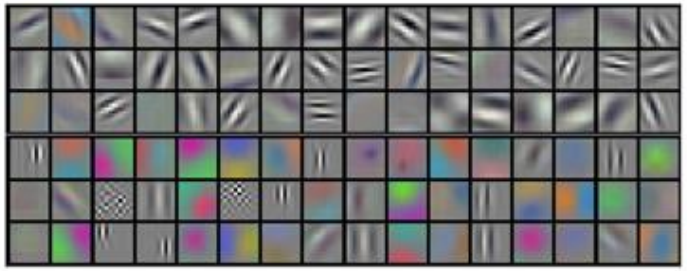

11
11
224
224
3
Stride
of 4
5
5
48
55
48
55
Max
pooling
3
3
3
128
27
128
27
Max
pooling
3
3
3
192
192
128
13
13
13
192
192
128
Max
pooling
dense
2048
2048
dense
dense
2048
2048
1000

Layer 1

Layer 2

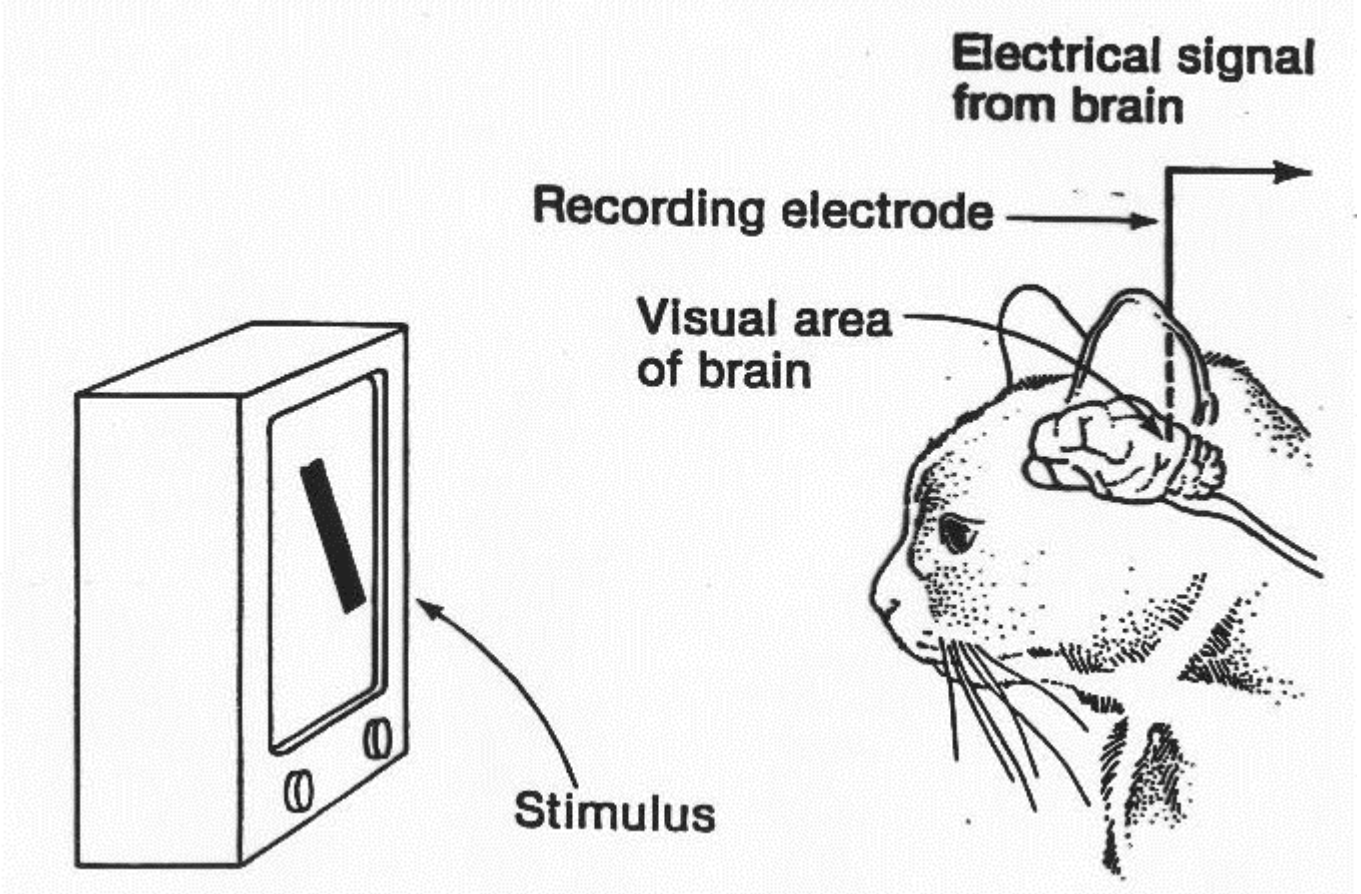
Layer 3





Matthew Zeiler and Rob Fergus, Visualizing and understanding convolutional networks, 2014

Hubel-Wiesel Experiment



Jennifer Aniston Neuron

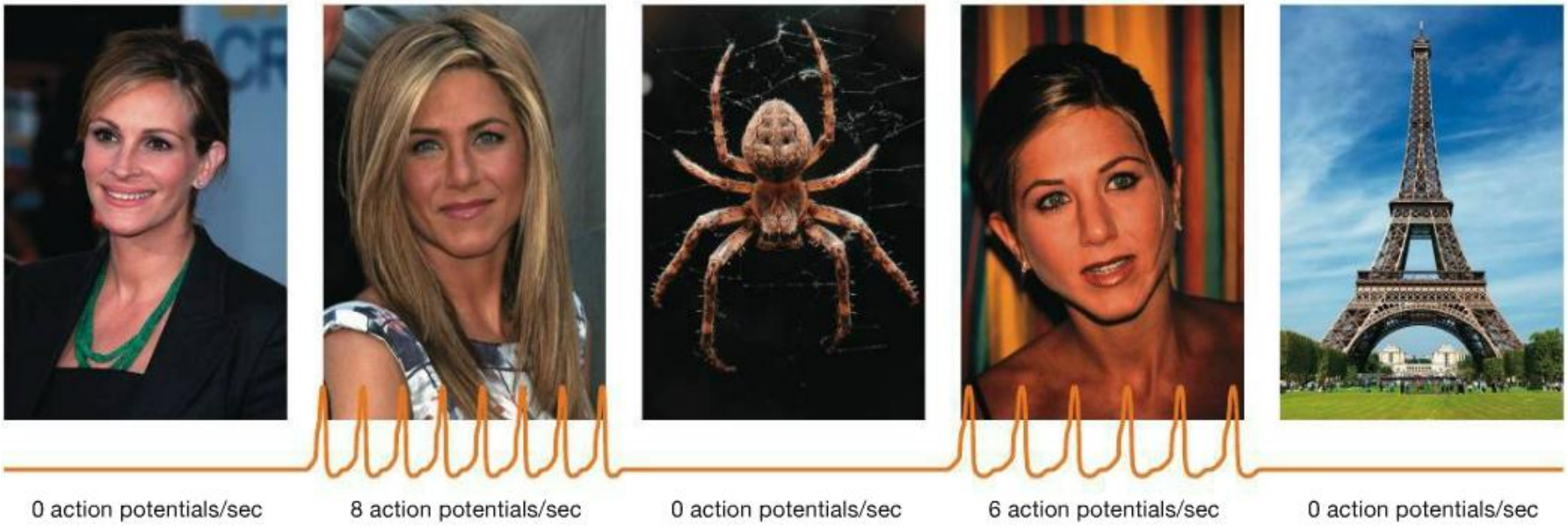


FIGURE 46.22 Single-Neuron Recording Reveals that Some Neurons in the Brain Recognize Specific Concepts. The graphs below each image show how a single neuron fires in response to images of actress Jennifer Aniston but not to other images.

DATA: Quiroga, R. Q., L. Reddy, G. Kreiman, et al. 2005. *Nature* 435: 1102–1107.

The Nobel Prize in Physiology or Medicine 2014

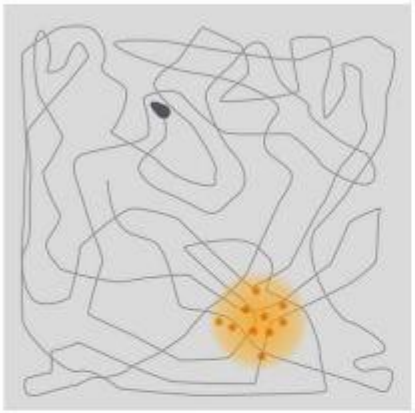


Fig. 1

John O'Keefe

John O'Keefe discovered, in 1971, that certain nerve cells in the brain were activated when a rat assumed a particular place in the environment. Other nerve cells were activated at other places. He proposed that these "place cells" build up an inner map of the environment. Place cells are located in a part of the brain called the hippocampus.

May-Britt Moser and Edvard I. Moser

May-Britt och Edvard I. Moser discovered in 2005 that other nerve cells in a nearby part of the brain, the entorhinal cortex, were activated when the rat passed certain locations. Together, these locations formed a hexagonal grid, each "grid cell" reacting in a unique spatial pattern. Collectively, these grid cells form a coordinate system that allows for spatial navigation.

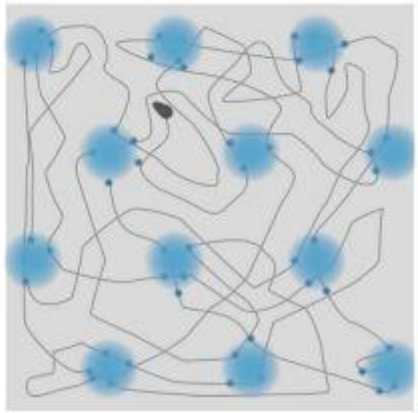
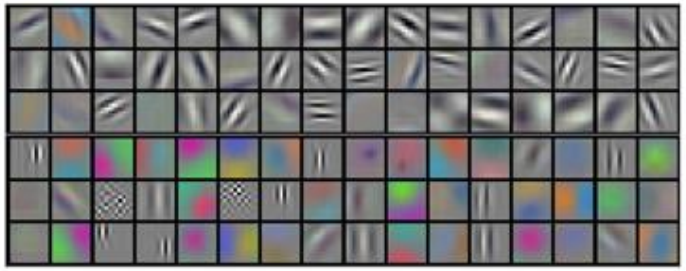
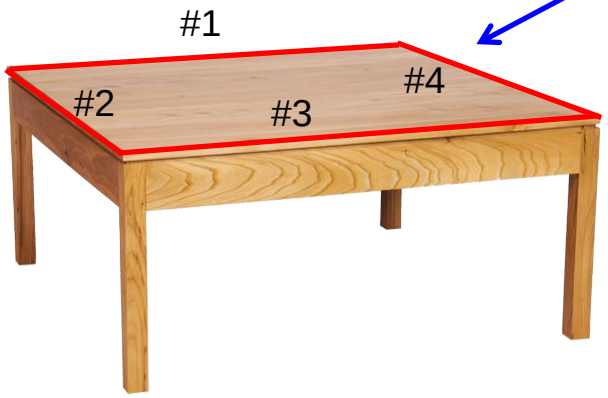


Fig. 2


Nurture and Thrive
for
Busy Mom's
Helper

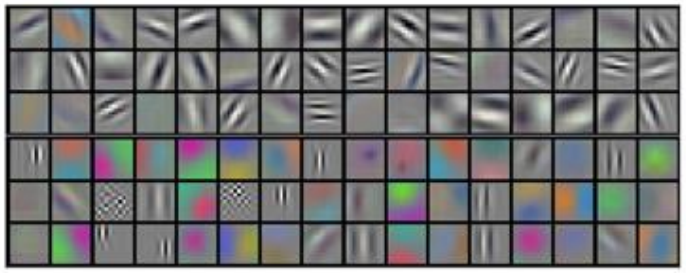


Handcrafted Algorithm vs. DL



If there are 4 edges such that #1 and #3 are almost parallel and #2 and #4 are almost parallel, then it is a rectangle.

Hand-crafted algorithm by humans: takes a lot of time and effort to come up with algorithms and implement them.



"Algorithm" automatically learned from data. Very fast

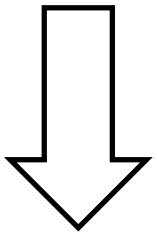
KAIST

Handcrafted Algorithm vs. DL

```
QUICKSORT(A, p, r)
1  if p < r
2      q = PARTITION(A, p, r)
3      QUICKSORT(A, p, q − 1)
4      QUICKSORT(A, q + 1, r)
```

```
PARTITION(A, p, r)
1  x = A[r]
2  i = p − 1
3  for j = p to r − 1
4      if A[j] ≤ x
5          i = i + 1
6          exchange A[i] with A[j]
7  exchange A[i + 1] with A[r]
8  return i + 1
```

Massive training data



Training

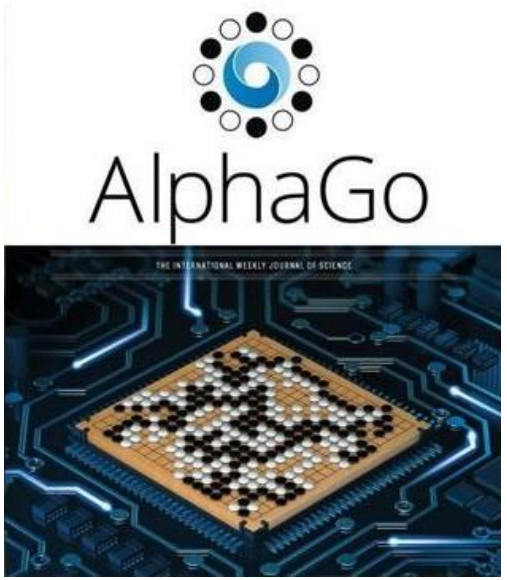
Neural network parameters

Simple descriptive complexity

Extreme descriptive complexity

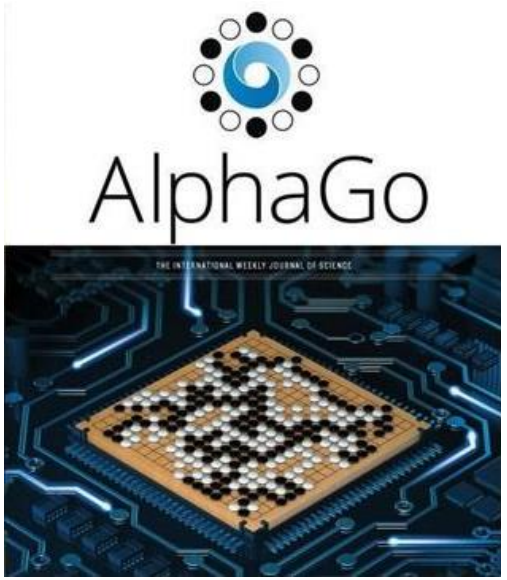
- 4,000 year history of Go
- Still never-played positions and strategies

- New “algorithms” (Go strategies) found automatically within months after a lot of training



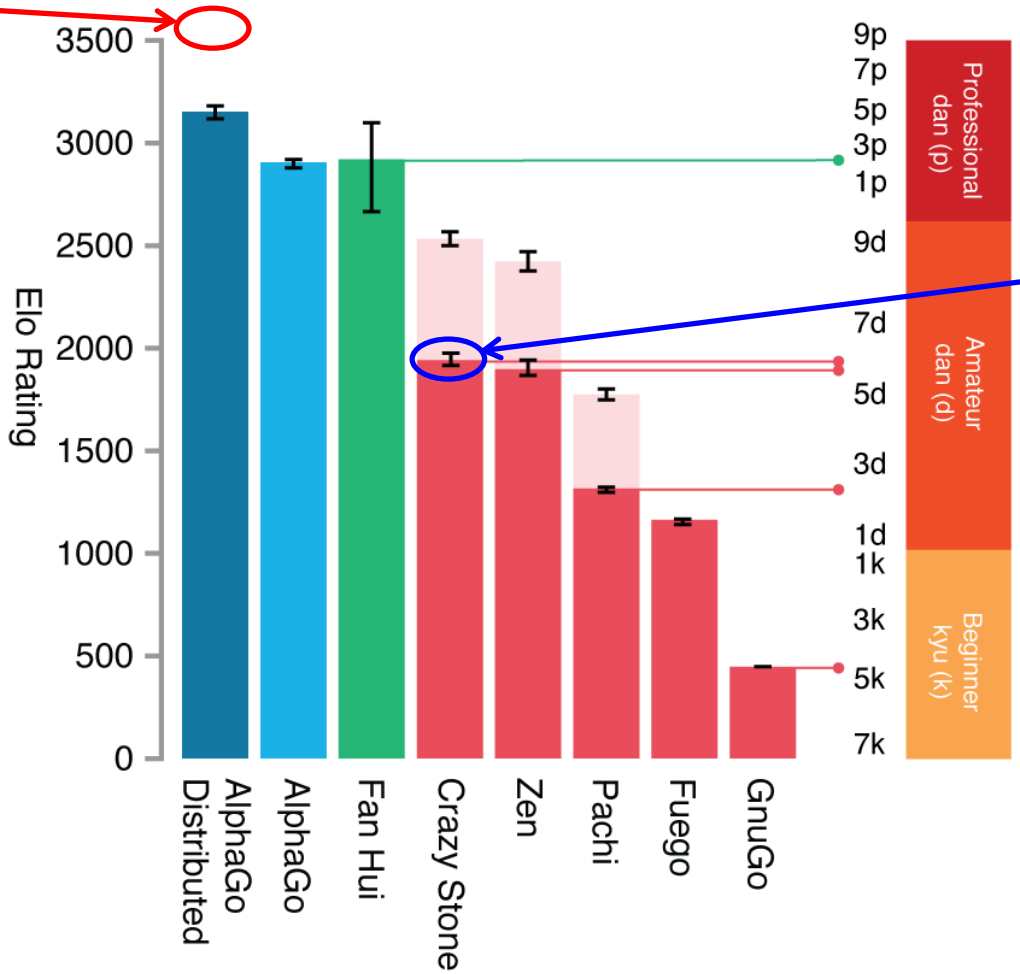
- Millions of lines of code
- Millions of man-hours
- Almost zero intelligence

- No coding required during learning
- “Algorithms” and “codes” found automatically
- Intelligence emerges automatically



AlphaGo (2016)

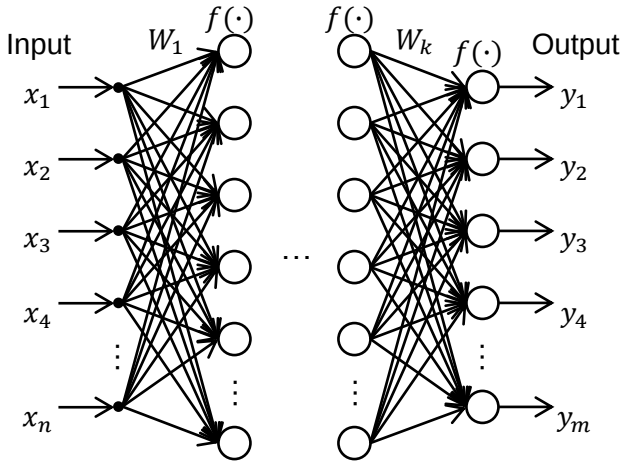
AlphaGo learning by itself



Best Go program based on hand-crafted algorithms

KAIST

Paradigm Shift



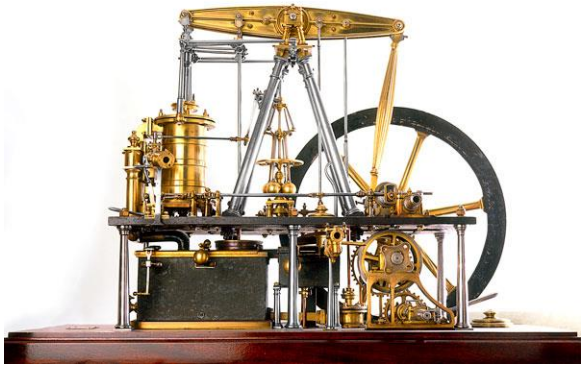
Paradigm Shift

“Machine learning. This is the next transformation.” “I’m a programmer who sort of got lucky at Google. But the programming paradigm is changing. Instead of programming a computer, you teach a computer to learn something and it does what you want.”

– Eric Schmidt, CEO of Google, during a keynote speech during Google's Cloud Computing Platform tech conference (2016/03/23)

New Era

Energy
1700's ~



Steam engine
Electricity

Simple information processing
1940's ~

Computers
Internet

Complicated information processing
2010's ~

Artificial Intelligence

ICML 2017 Sponsors

Diamond

Platinum

ICML 2017 Sponsors

Gold

Silver: Recruit Communications, Voleon Group, D. E. Shaw, Cubist Systematic Strategies, Bosch, Aitrics, Western Digital

Bronze: Cisco, Adobe, Toyota

Not in the List

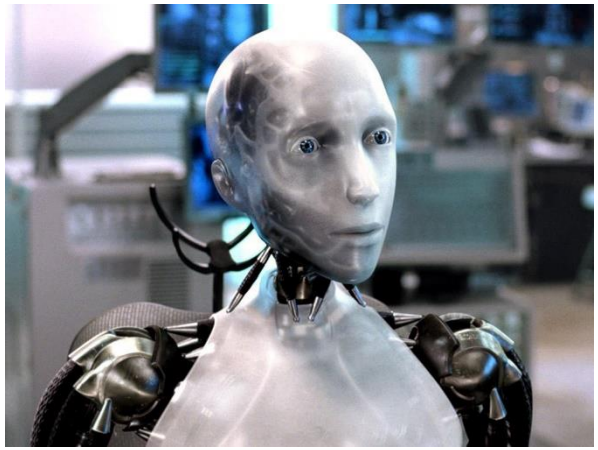
World's Largest Companies

2017 Second quarter

Rank	Country	Company	Market cap (M$)
1	USA	Apple Inc.	749,124
2	USA	Google (Alphabet)	628,610
3	USA	Microsoft	528,778
4	USA	Amazon.com	466,471
5	USA	Berkshire Hathaway	418,880
6	USA	Johnson & Johnson	357,310
7	USA	Facebook	357,176
8	China	Alibaba Group	356,390
9	China	Tencent	344,879
10	USA	ExxonMobil	341,947

Companies investing heavily in AI research

Why No Strong AI Yet?



KAIST

Artificial Intelligence

- Premature optimism & AI winters

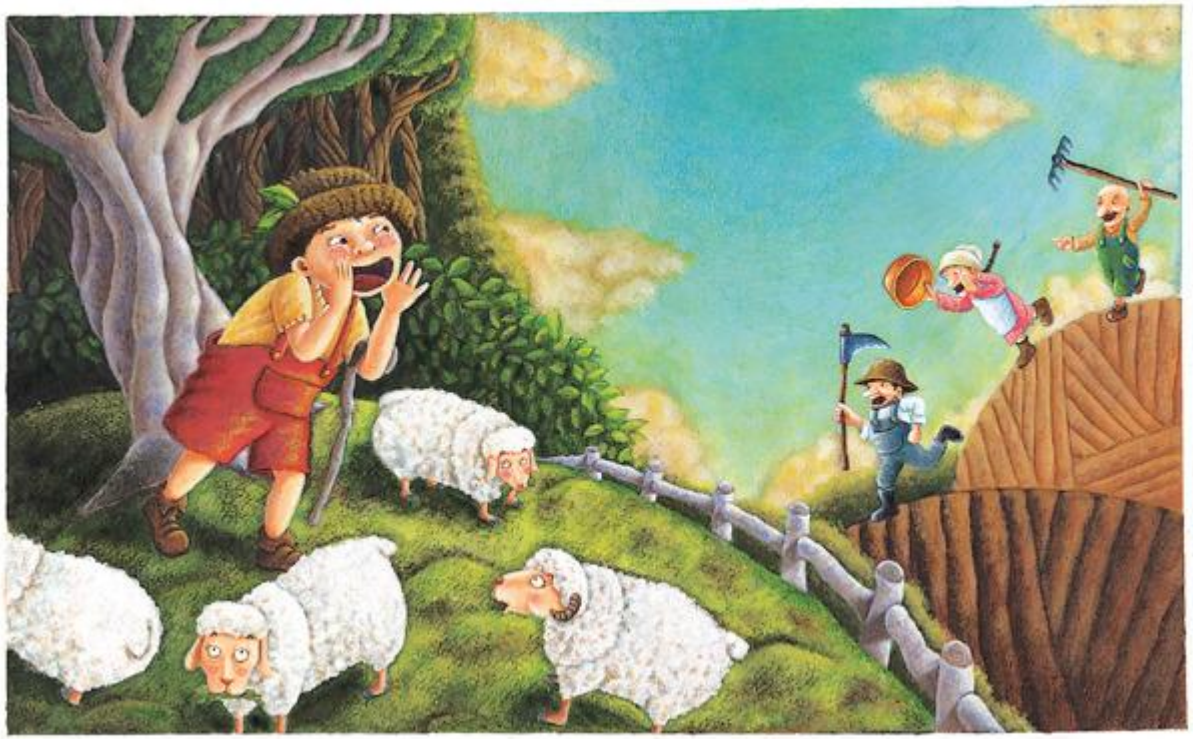
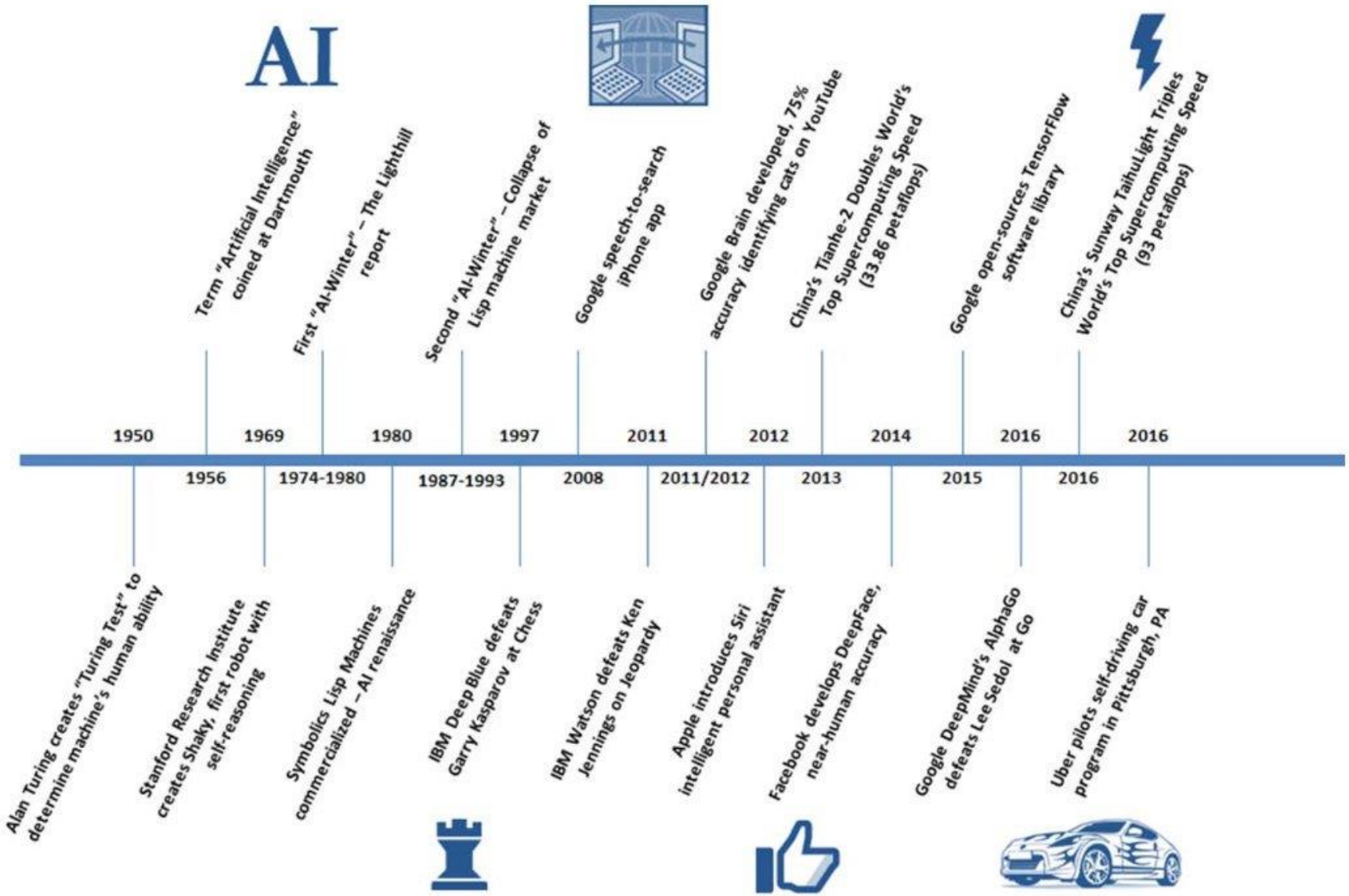


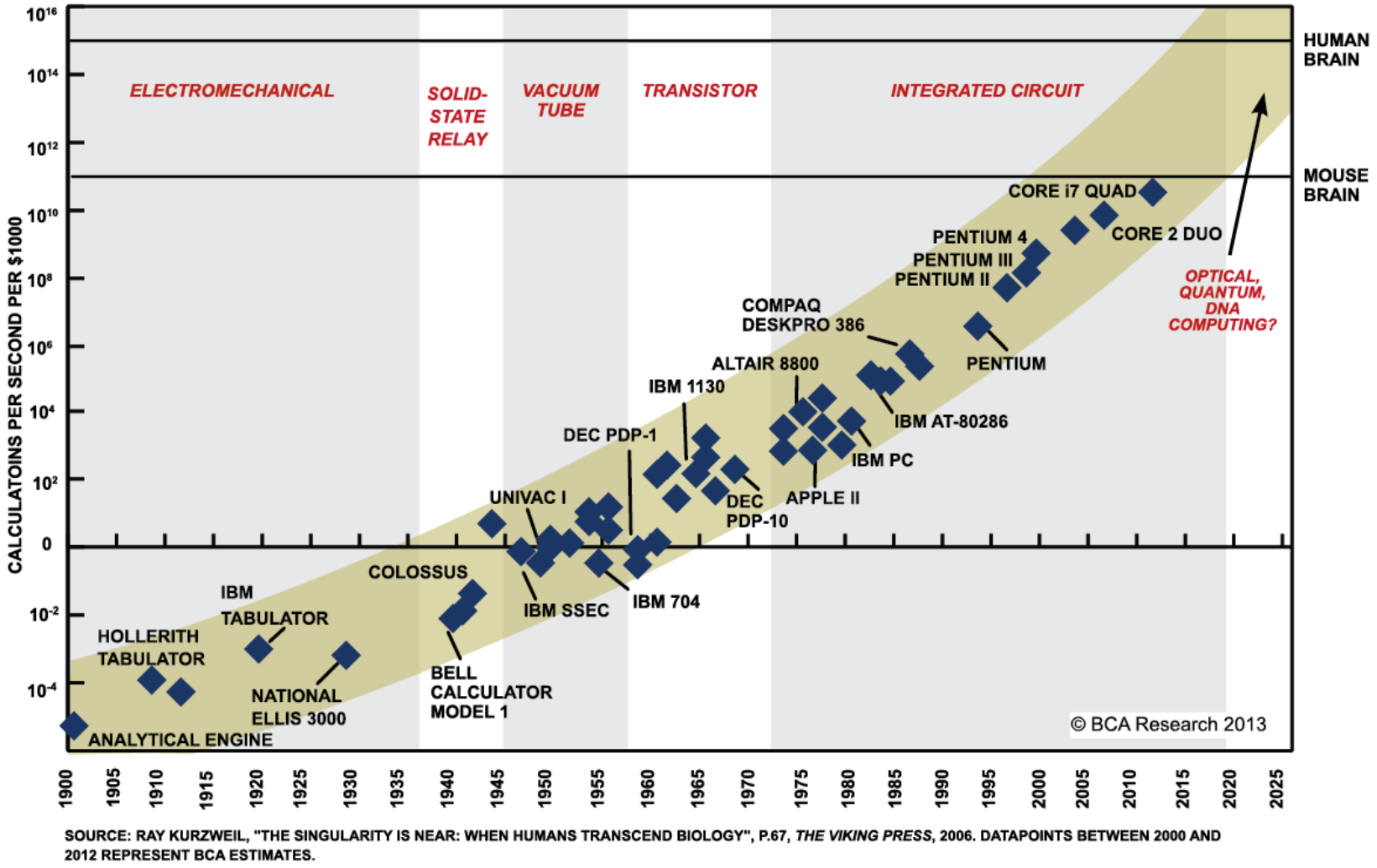
Exhibit 8: Evolution of AI: 1950-Present



Source: Company data, Goldman Sachs Global Investment Research

source goldman sachs via @mikequindazzi

Hardware is Not There Yet



SOURCE: RAY KURZWEIL, "THE SINGULARITY IS NEAR: WHEN HUMANS TRANSCEND BIOLOGY", P.67, *THE VIKING PRESS*, 2006. DATAPOINTS BETWEEN 2000 AND 2012 REPRESENT BCA ESTIMATES.

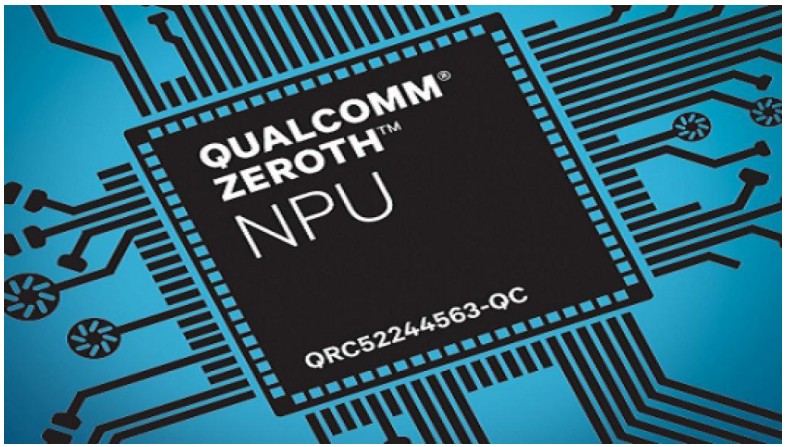
Supercomputer vs. human brain

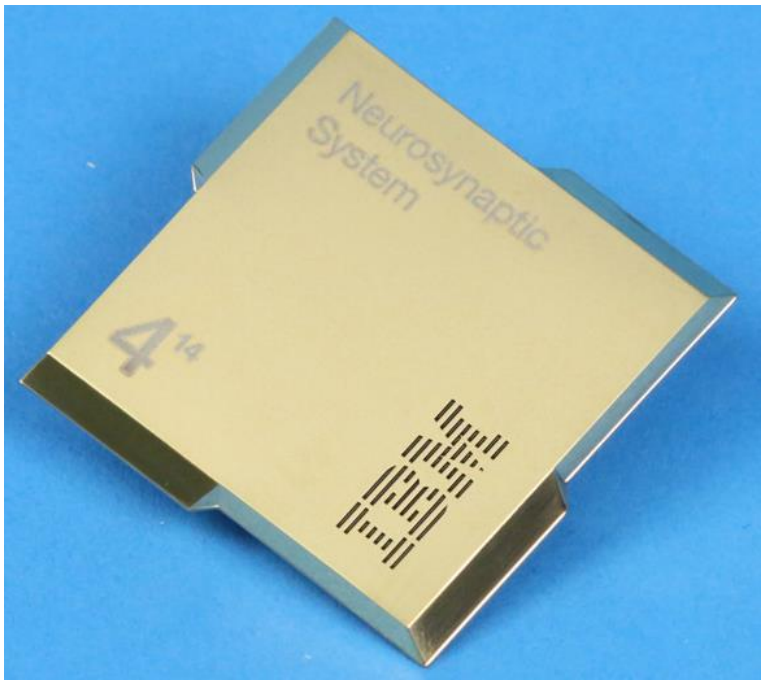
	Human brain	Tianhe-2
Components	86,000,000,000 neurons* 150,000,000,000,000 synapses	3,120,000 cores
Computational power	?	33.86 Peta FLOPS
Memory	2.5PB?	1.375PB
Storage		12.4PB
Power consumption	~20 Watts	24 MWatts

Human: each neuron is slow (up to 100 m/s), massive parallelism
Supercomputer: each core is fast (GHz), light speed communication (200,000~300,000 km/s), mid-scale parallelism

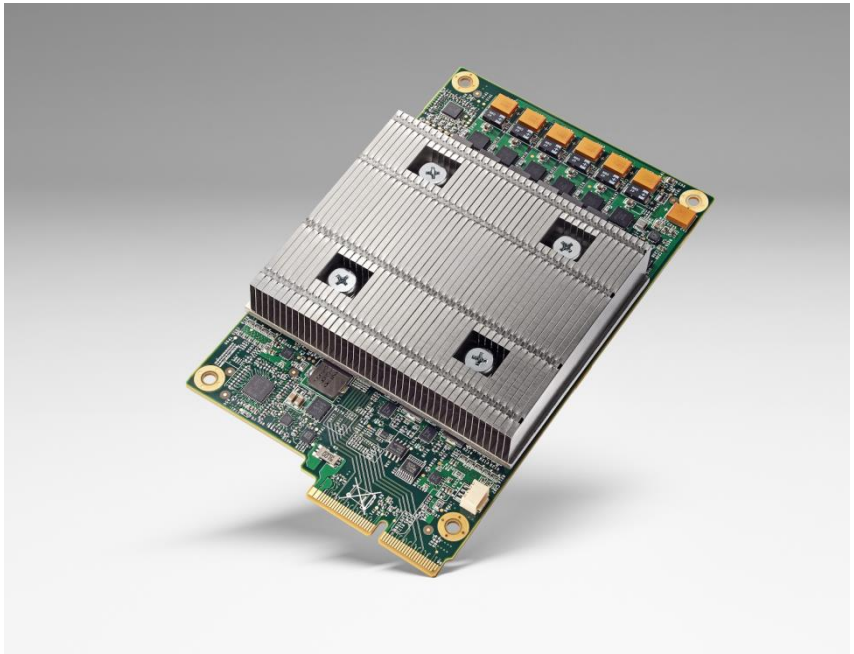
*86.1±8.1 billion neurons and 84.6±9.8 billion non-neuron cells (F. Azevedo, et al., Equal numbers of neuronal and nonneuronal cells make the human brain an isometrically scaled-up primate brain, J Comp Neurol. 2009)

Neural Hardware





Tensor Processing Units



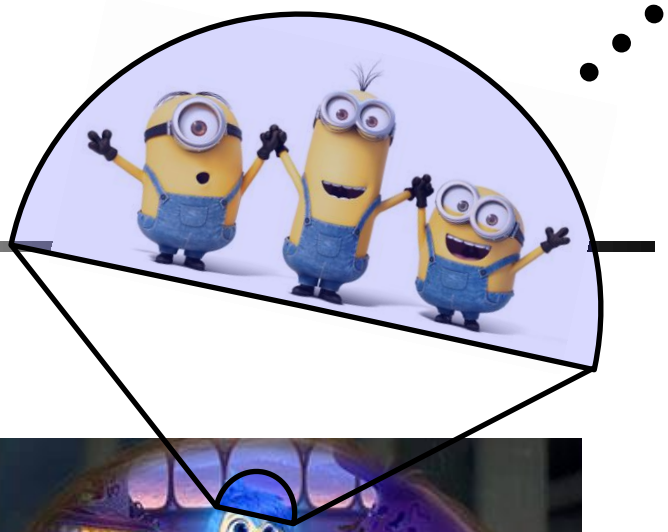
Why No Strong AI Yet?

- Hardware is not there yet
  - Even today, world's fastest supercomputer can barely match the computing power of a single human brain
- No general architecture for human-like brain
  - We only have specialized architectures good for specific tasks – CNN, RNN, limited memory, attention mechanism, …
  - Human brain is much more complicated
- No general training method for general architecture
  - We only have simple training methods for specific tasks – supervised learning, unsupervised learning, reinforcement learning, …
  - Human brain is not trained that way

Some People Say

Strong AI is not possible because humans have soul but machines don’t.

Soul?



Mysterious driver?

Humans have intuition

||

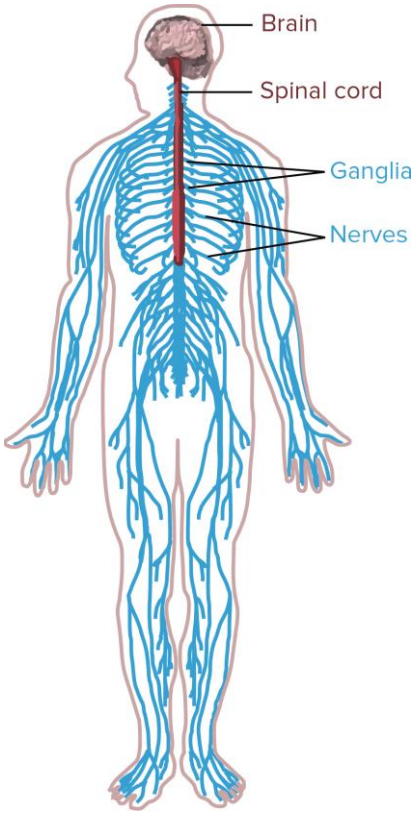
I have no idea how humans think

Humans have soul

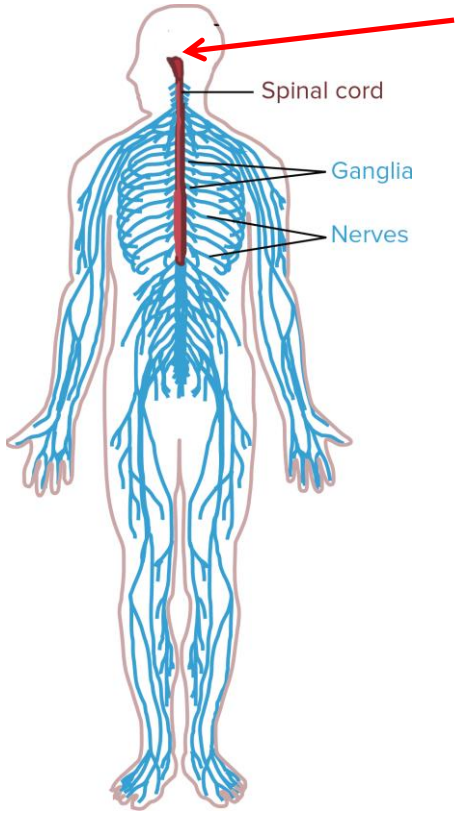
||

I have no idea how humans think

If a human brain follows the law of physics



If we have soul that does not follow the law of physics, then we don't need a brain



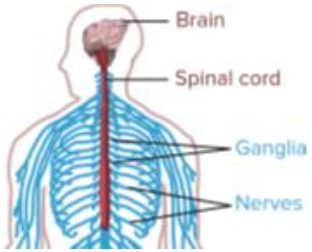
because signals can magically appear here and control our body.

What Most People Believe

Normal brain

You can talk, think, move normally.

Alzheimer's disease

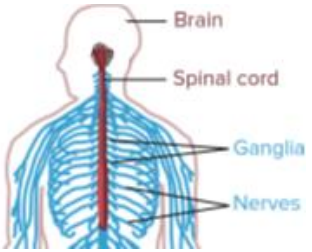


Brain shrinks

Loss of memory

Changed personality

Brain death



No cognition

No more talking, thinking, moving

Death

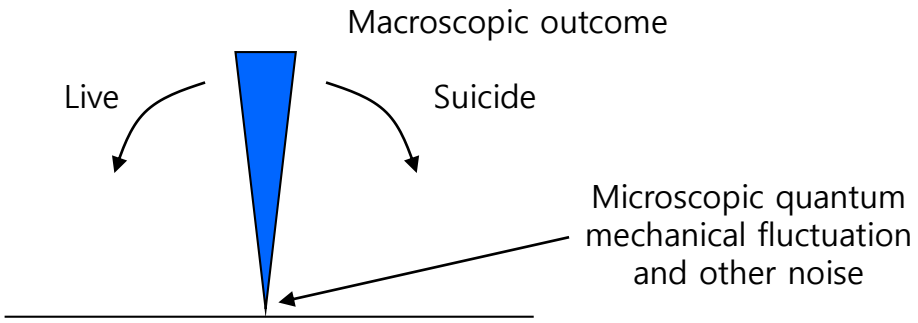
All brain functions restored by magic

All memory restored by magic

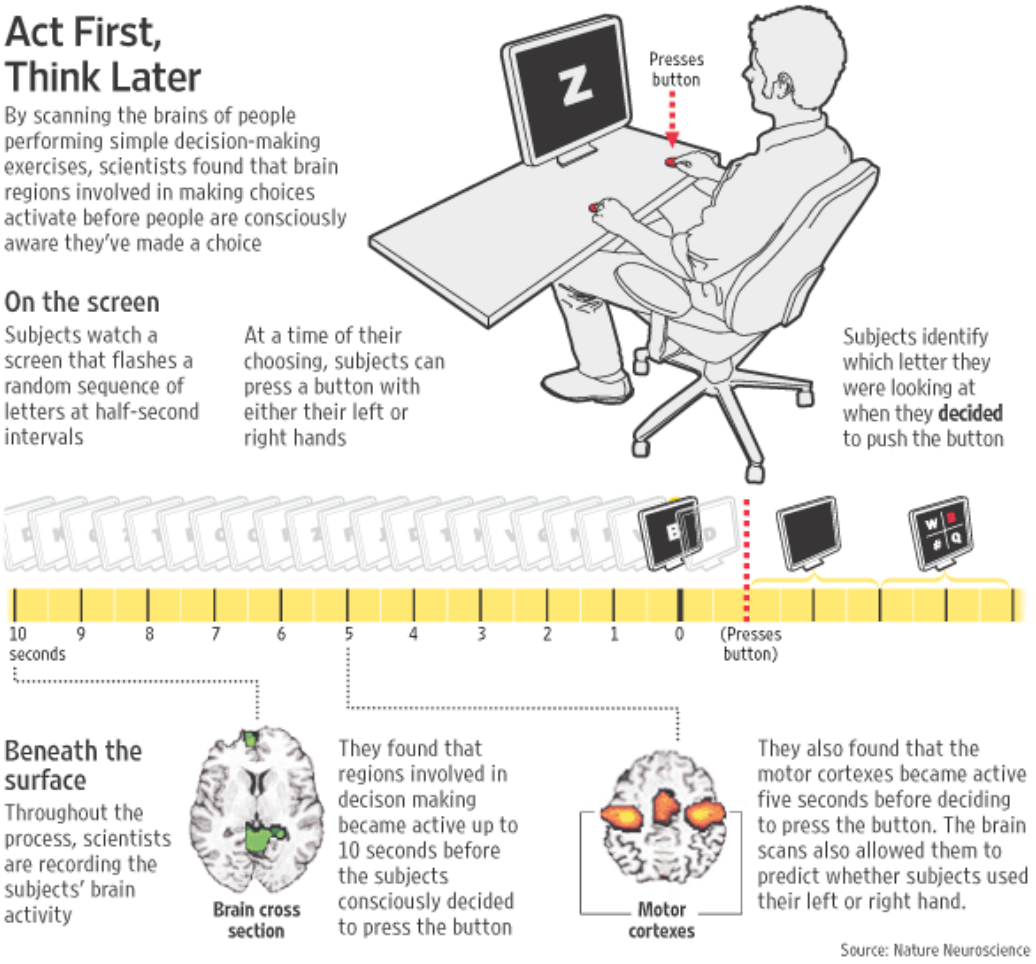
You can now talk, think, move normally.

Free Will

From movie “Pay it forward”



Free Will Experiment



Haynes et al. Nature Neuroscience 2008

What we think we are doing

Quantum mechanical and other noise modulated by our biological neural network

Our consciousness "interpreting" after the fact what happened is due to his free will

Reality

Q: What is the following describing?

1. Can use tools. Can invent tools.
2. Can form a society and can cooperate.
3. Punishes a member if did something wrong.
4. Use language and even has dialect depending on where they live.
5. Have names and call others. Remembers friend's name even after 20 years.
6. Self aware.
7. Feels happiness and sadness.
8. Has empathy and is altruistic.
9. Kills other animals for fun.
10. Kills a friend and suffers emotionally due to sense of guilt.
11. Plans ahead before traveling and shares the plan with other members.

https://www.youtube.com/watch?v=TtmLVP0HvDg

https://www.youtube.com/watch?v=NenEdSuL7QU

Importance of Turing Test

Importance of Turing Test

From movie “Planet of the Apes”

Why Turing Test?

- Why Turing test?
- Because we want to hear AI say

"I know that I know."
"I know that I know that I know."
"I think, therefore I am."
"I am as real as you."
And much more

- Because it doesn't matter what's inside.

Why Text?

Then, You Ask

- Q: Wait! Just because it talks like a human doesn't mean it can actually think and feel like a human, right? Humans have ***xyz***, but machines don't.
- But, we can ask the machine questions like the following

"What are you thinking now?"

"Do you have ***xyz***?"

- This is why Turing test is enough – we can know whether it thinks and feels like us even during its spare time and whether it thinks it has ***xyz***.
- Q: Perhaps it talks that way because it just thinks it has ***xyz***.
- So do we.

바둑 = 수담(手談)

- During game 1
  - 박정상 9단: 지금 알파고가 흑돌이 살아 있는지를 묻고 있는 거죠.
  - ...
  - 해설자: 상변의 사활을 의심한다는 것 자체가 저는 신기하네요.
  - 박정상 9단: 네 일단은 좀 간담이 서늘해지는 장면이긴 합니다. ...
  - 해설자: 과연 완생인지 미생인지를 묻고 있습니다.
- Tygem commentary for game 5
  - 여기서 알파고가 백48로 끊은 수가 재미있는 응수타진. 상대의 응수에 따라 다음 착점을 결정하겠다는 고수의 수법이다.
  - http://www.tygem.com/news/news/viewpage.asp?pagec=2&seq=20782&gubun=W013
- AlphaGo passed the Turing test of Go since it has beaten top human opponents, e.g., 3:0 against Ke Jie in 2017. (Assuming a relaxed definition of Turing test, i.e., super-human performance instead of indistinguishability from human)
- Not only that, it has reached a level that is far beyond our understanding. Top Go players are struggling to understand its moves.

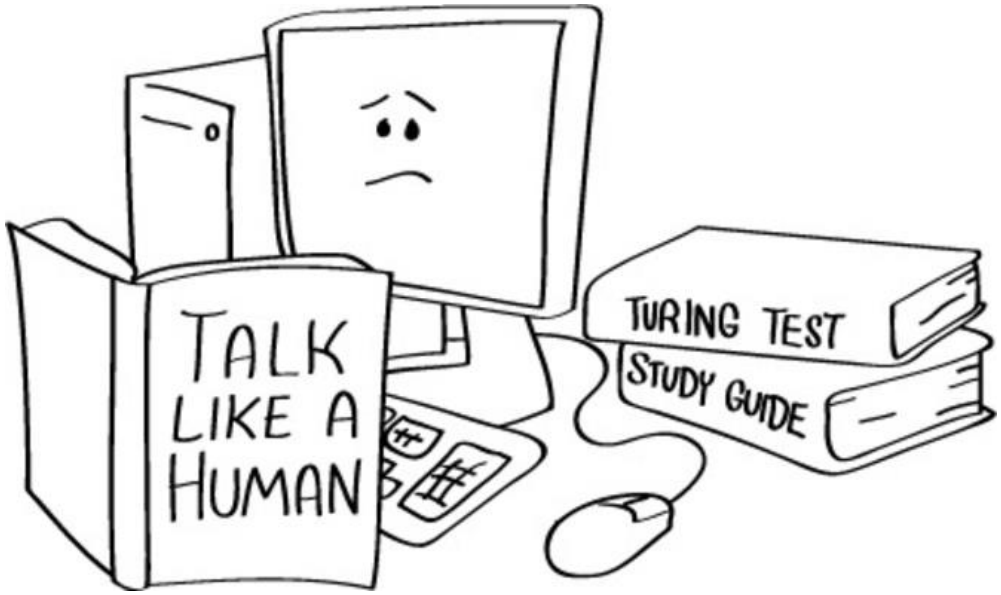
Still a Long Way to AGI

Still a Long Way to AGI

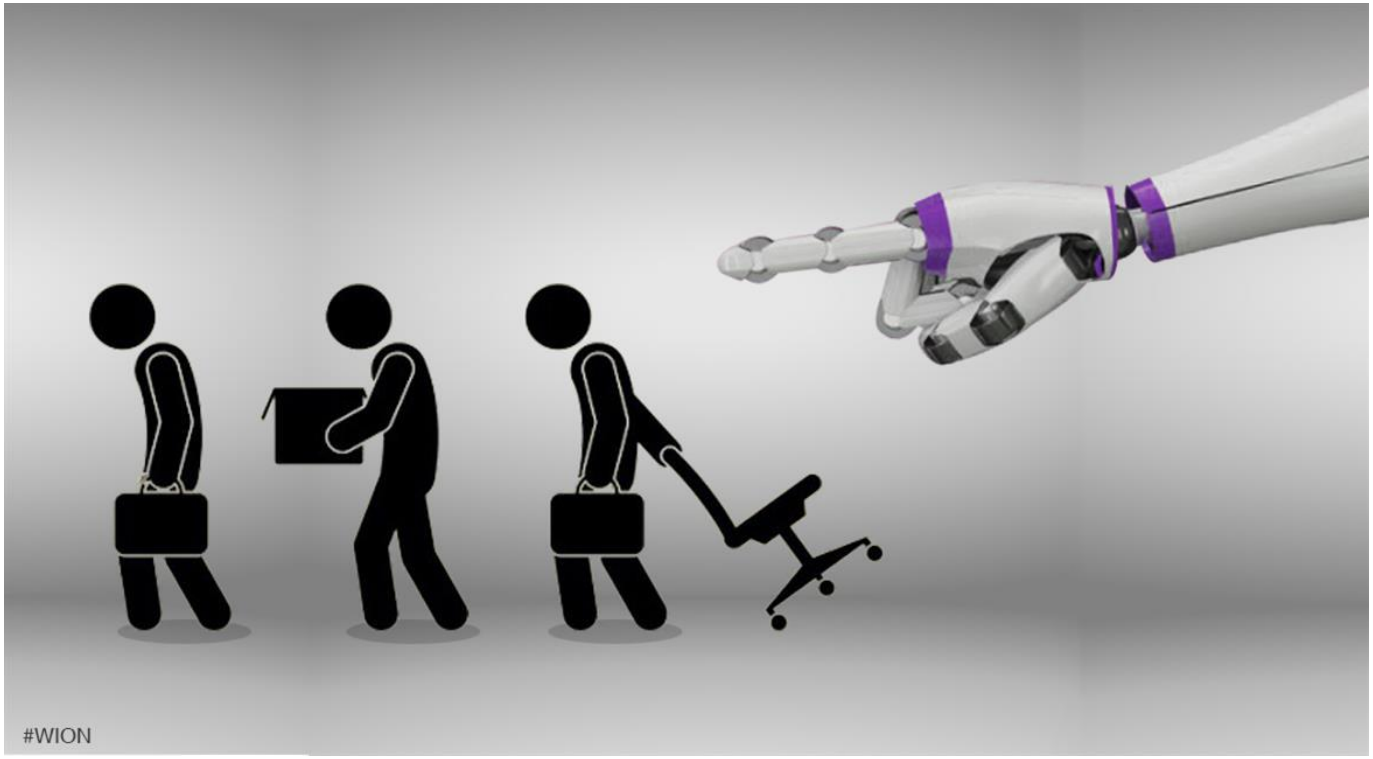
- Facebook AI agents
  - Bob: i can i i everything else . . . . . .
  - Alice: balls have zero to me to me to me to me to me to me to me to me to me
  - Bob: you i everything else . . . . . .
  - Alice: balls have a ball to me to me to me to me to me to me to me
  - Bob: i i can i i i everything else . . . . . .
  - Alice: balls have a ball to me to me to me to me to me to
  - Bob: i . . . . . .
- AI has not passed the Turing test yet
  - It can invent its own way of communication, it has happened many times in the past. They were all simple and primitive. It is nothing compared to passing the Turing test.
  - Beware of misleading news such as “Facebook shuts down robots after they invent their own language”
- We are not there yet. We can worry about AI inventing their own language that is far more advanced that that of humans AFTER they pass the Turing test, i.e., they talk like humans.

Still a Long Way to AGI

- Human language is a general tool that enables expressing complicated thoughts and ideas
- We don't have AGI yet not because computers cannot process human language but because ideas and common sense of humans are too rich for computers to handle yet

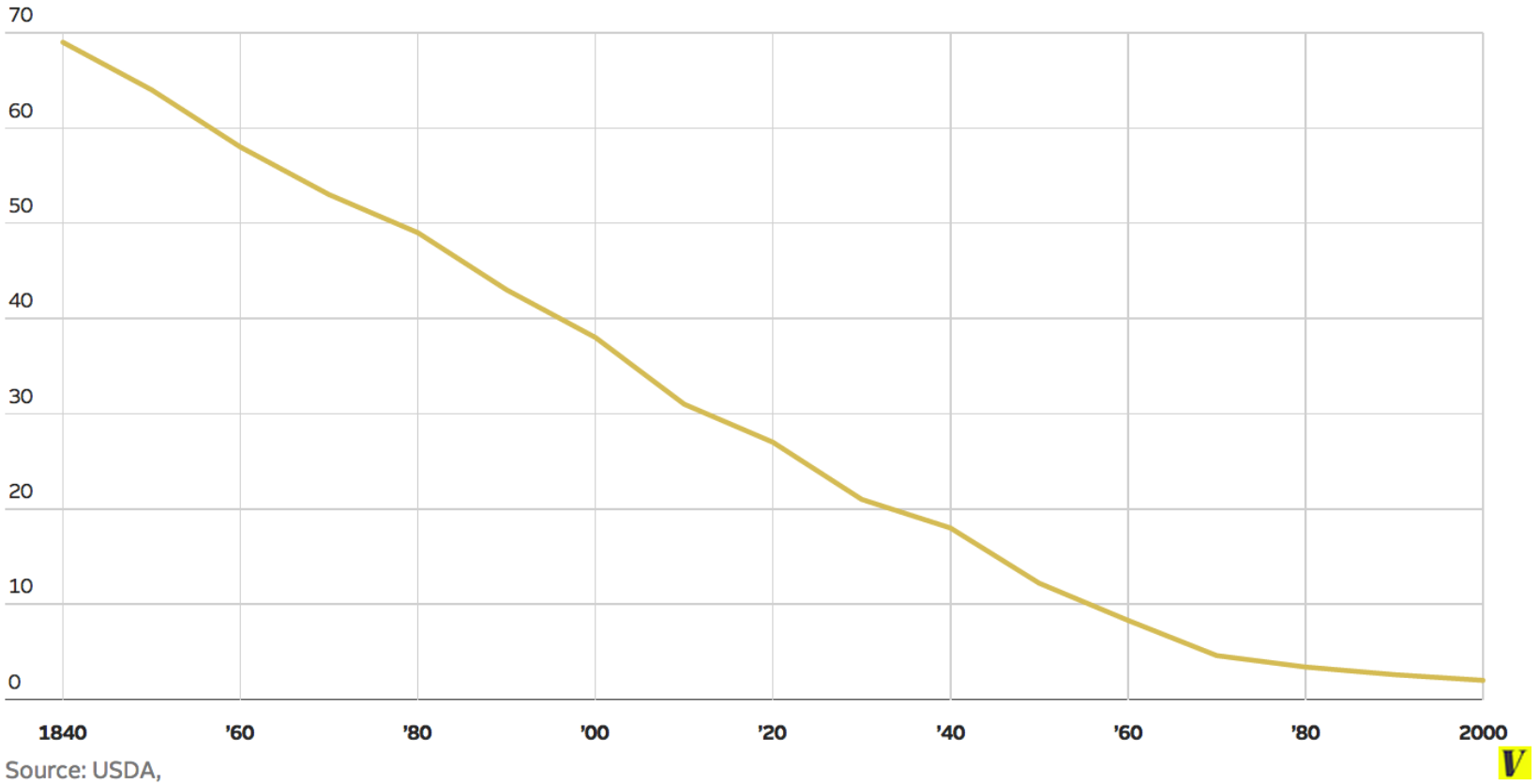


Are We Going to Lose Our Jobs?



The Luddite Movement in 1811 ~ 1812, London

% of American workforce in agriculture, 1840-2000



http://www.vox.com/a/explain-food-america

Possible Scenarios

- Year 1840
  - 70% of population in agriculture
  - Labor intensive, physical labor, mostly mundane tasks
  - Works 80 hours a week

- Year 2017
  - 70% of population in offices
  - Labor intensive, mental labor, mostly mundane tasks
  - Works 40 hours a week

- Year 2050
  - Optimistic scenario
    - 70% of population at home
    - Creativity, artistic skills, …
    - Works 20 hours a week
  - Pessimistic scenario
    - 70% of population on streets
    - Works 0 hour a week

KAIST

Reading Assignment

- Chapters 1, 2 and 3 of DL book