

增強學習初探以及最新發展趨勢

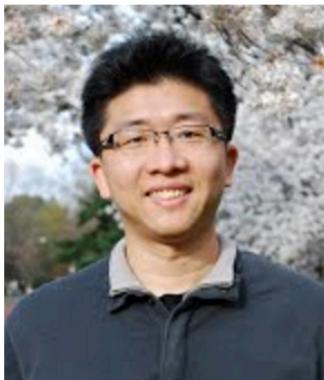
台灣清華大學
孫民教授



VSLab

Min Sun (孫民)

RESEARCH PUBLICATIONS LAB TEACHING ARCHIVE



Assistant Professor in Electrical Engineering
at National Tsing Hua University (Sept. 2014)
[\[CV\]](#),[\[Google Scholar\]](#)



Postdoctoral Researcher with Steve Seitz and Ali Farhadi
in CSE at University of Washington (Jan. 2013 - Aug. 2014)



Microsoft
Research
Cambridge



Sep 09-Dec 12

Jul 11-Sep 11

Jul 09-Sep 09

Sep 07-Jul 09

Sep 05-Sep

Sep 99-Jul 03

07

aliensunmin.github.io

2018年3月號

AI大師學2》師承華人兩大世界級名師

35歲助理教授孫民 打造下一個李飛飛、吳恩達

文 / 陳育晟 攝影 / 賴永祥 2018-03-01



<https://www.gvm.com.tw/article.html?id=43046>



Vision Science Lab@ NTHU Taiwan

PI: Min Sun

Web: aliensunmin.github.io

Office: Delta 962

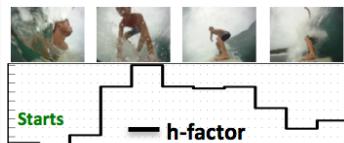
Lab: EECS Bldg 712

Tel: +886-3-5731058

Email: sunmin@ee.nthu.edu.tw

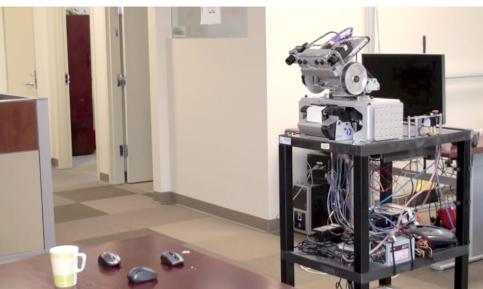
Research Topics

Understanding Personal Videos

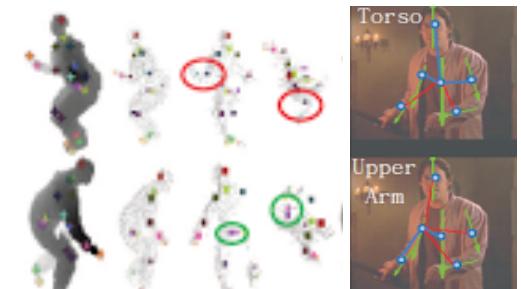


Make3D

3D & Robot Vision



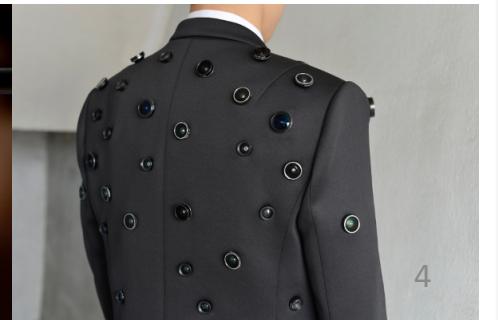
Human Sensing



Analyzing Street Views



Wearable Camera Applications



My Journey on Learning Reinforcement Learning (RL)

- Courses during my graduate study
 - CS229 Machine Learning:
<http://cs229.stanford.edu/syllabus.html>

Reinforcement learning and control (4 classes)

Lecture 16	11/15	1. MDPs. Bellman equations. 2. Value iteration and policy iteration.	Class Notes <ul style="list-style-type: none">• Reinforcement Learning and Control [ps] [pdf]• LQR, DDP and LQG [pdf]
Lecture 17	11/27	3. Linear quadratic regulation (LQR). LQG. 4. Q-learning. Value function approximation.	
Lecture 18	11/29		Problem Set 4 Out 11/15. Due 12/6.

- EE363 Linear Dynamical System:
<http://stanford.edu/class/ee363/courseinfo.html>
 1. [Linear quadratic regulator: Discrete-time finite horizon](#)
 2. [LQR via Lagrange multipliers](#)
 3. [Infinite horizon LQR](#)
 4. [Continuous-time LQR](#)

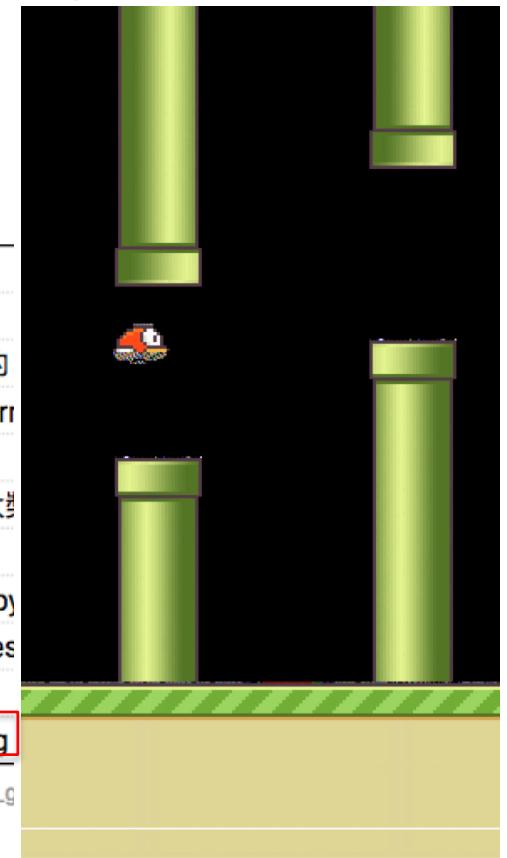
My Journey on Learning Reinforcement Learning (RL)

- In Early 2016, my student (Yen Chen Lin) tried RL to play video games (e.g., Flappy Bird)
 - Playing Atari with Deep Reinforcement Learning (Mnih et al. NIPS workshop 2013)

Top GitHub projects by stars for Taiwan (2016, special rank)

Total GitHub stars from Taiwan: 87,435

Rank	Stars	GitHub Project	Description
1	322	abalone0204/Clairvoyance	Make the Taiwan a Better Place to Work
2	294	kdchang/reactjs101	從零開始學 React (ReactJS 101) 是一本給初學者的
3	231	hackmdio/hackmd	Realtime collaborative markdown notes on all platforms.
4	142	detailyang/awesome-cheatsheet	useful cheat sheets as reference.
5	137	doggy8088/Learn-Git-in-30-days	這是 Will 保哥在 2013 第 6 屆 iT 邦幫忙鐵人賽年度大賽
6	122	shiningjason1989/react-quick-tutorial	
7	110	hexojs/hexo	A fast, simple & powerful blog framework, powered by Node.js.
8	108	yahoo/AppDevKit	AppDevKit is an iOS development library that provides
9	106	jserv/facebookoc	Yet another Facebook clone written in C++
10	106	yenchenlin1994/DeepLearningFlappyBird	Flappy Bird hack using Deep Reinforcement Learning



My Journey on Learning Reinforcement Learning (RL)

- Watch a few online/offline courses (2016, 2017)
 - Reinforcement Learning from UCL by David Silver
 - <http://www0.cs.ucl.ac.uk/staff/d.silver/web/Teaching.html>
 - <https://www.youtube.com/playlist?list=PLbWDNovNB5mqFBgq7i3MY6Ui4zudcvNFJ>
 - Deep Reinforcement Learning from Berkeley by Sergey Levine
 - <http://rll.berkeley.edu/deeprlcourse/>
 - Deep Reinforcement Learning bootcamp
 - <https://sites.google.com/view/deep-rl-bootcamp/lectures>

My Journey on Learning Reinforcement Learning (RL)

- Many research results leverage RL
 - Interact through Language
 - Video Title Generation
 - Transferring Sentence Style
 - Interact through Actions
 - Deep 360-degree Pilot
 - Target Driven Navigation
 - Interact through Attacks
 - Adversarial Attacks

2017



Training a Deep Agent to See and Interact
孫民 / 清華大學電機系助理教授

<https://www.slideshare.net/aliensun/training-a-deep-agent-to-see-and-interact-83796512>

Acknowledgement

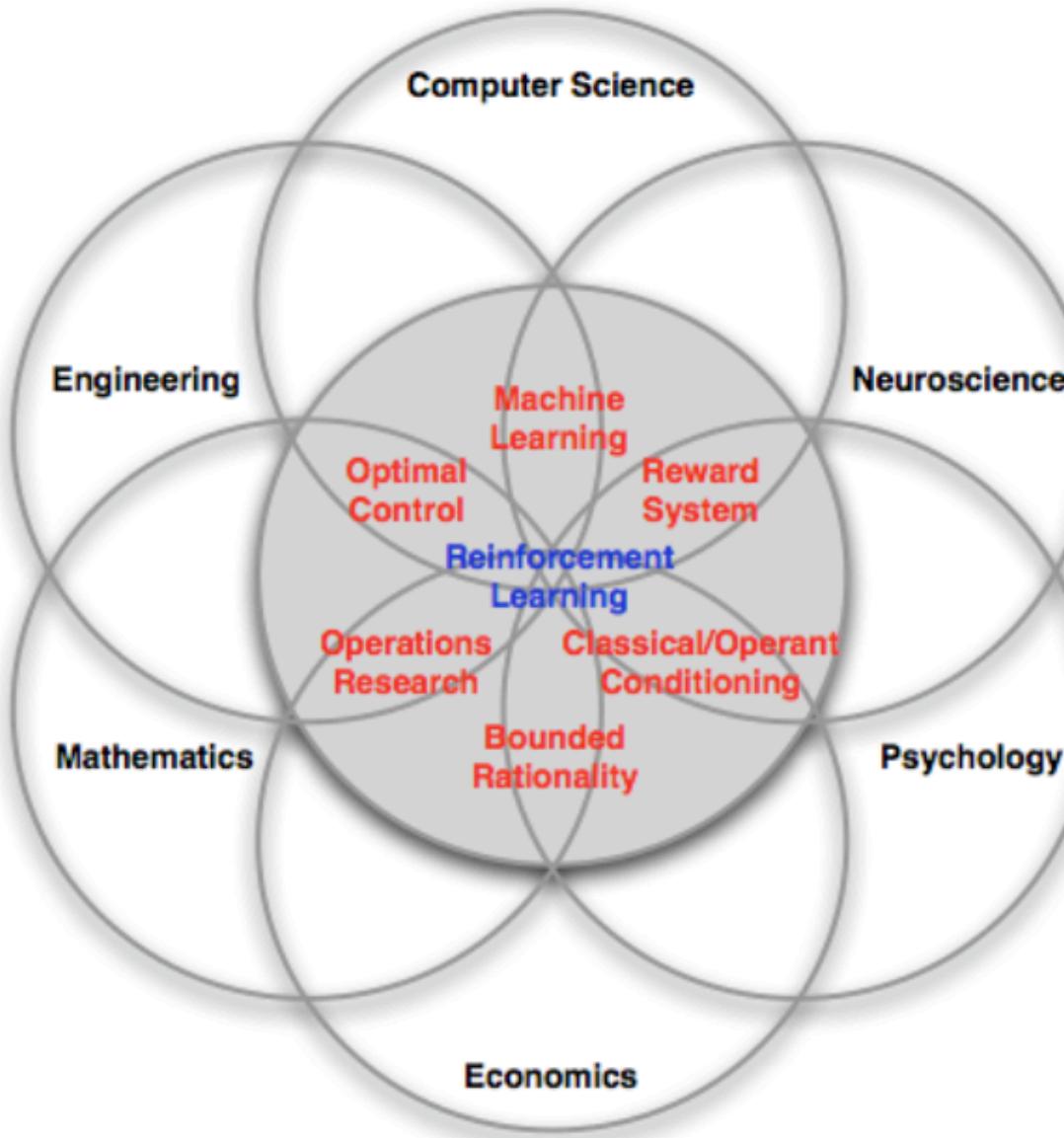
- 教材使用
 - David Silver RL slides
 - Deep Reinforcement Learning Bootcamp slides
 - Our Lab material contributed by Yen Chen Lin, Mark Dong, Gina Wu, Wan-Ting Hsu, and myself.
<http://aliensunmin.github.io/lab/info.html>

議程

- Introduction and overview
- Markov Decision Process 以及傳統增強學習的
算法
- 不需環境資訊 (Model-free) 的增強學習
- Deep RL: DQN
- 直接優化動作政策 (Action Policy) 的 (PG) 演算
法以及不需Gradient的Evolution Strategies(ES)演
算法
- 最新研究發展、應用範例以及使用增強學習的
小訣竅

Introduction and Overview

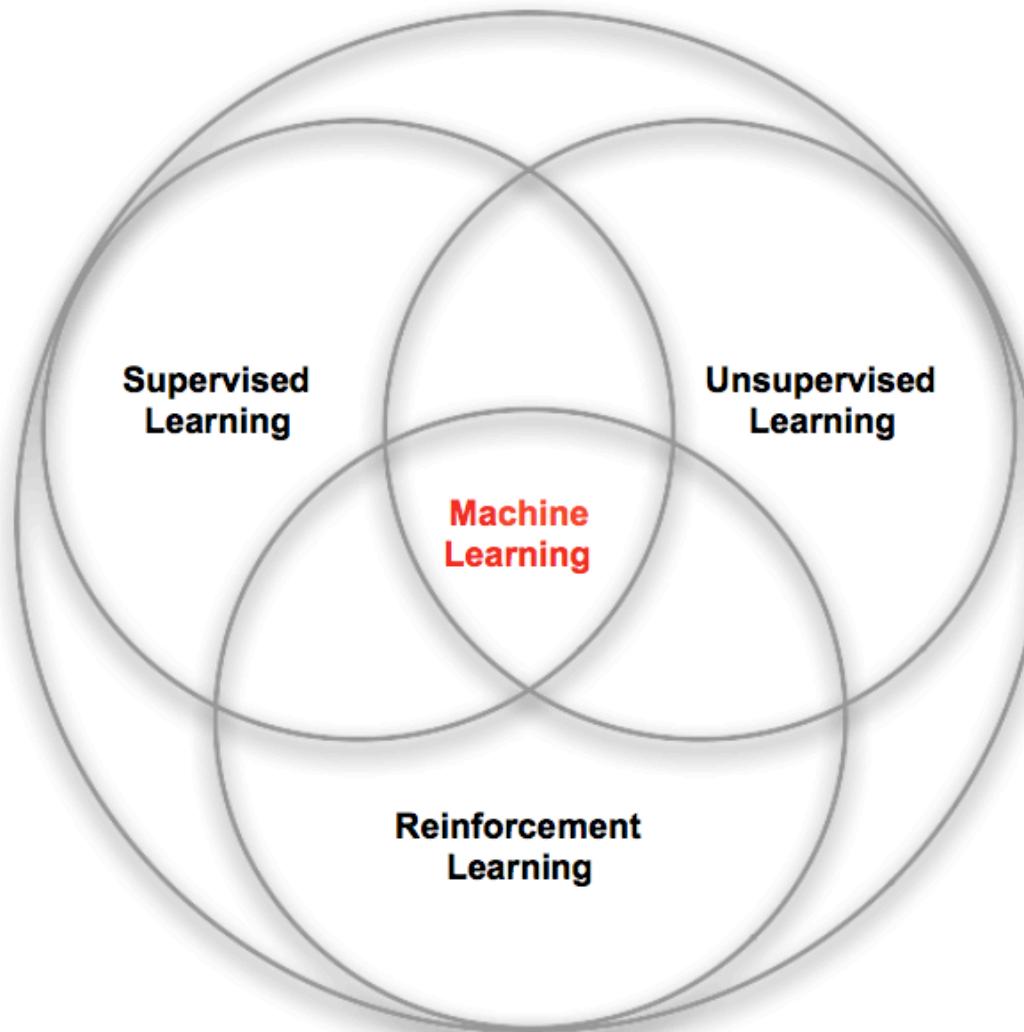
Many faces of Reinforcement Learning



What is Reinforcement Learning?

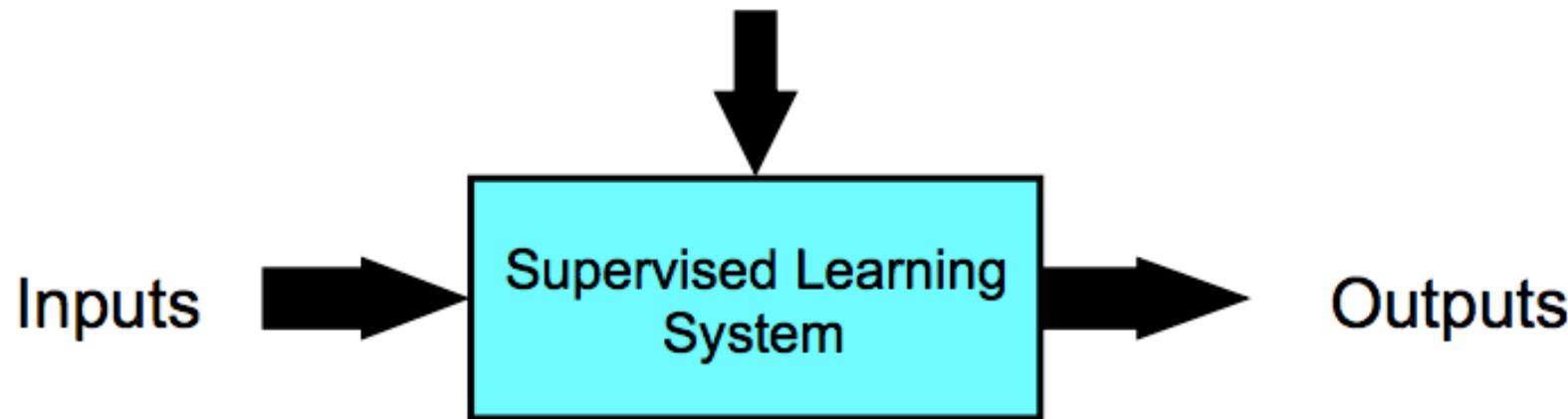
- Learning from interaction
- Goal-oriented learning 優化環境獲得的資訊來達到目標
- Learning about, from, and while interacting with an external environment 因為要跟環境互動，所以要先認識環境
policy是什麼？
- Learning what to do—how to map situations to actions—so as to maximize a numerical reward signal 跟環境互動來獲得回饋，才能學習

Branches of AI



Supervised Learning

Training Info = desired (target) outputs

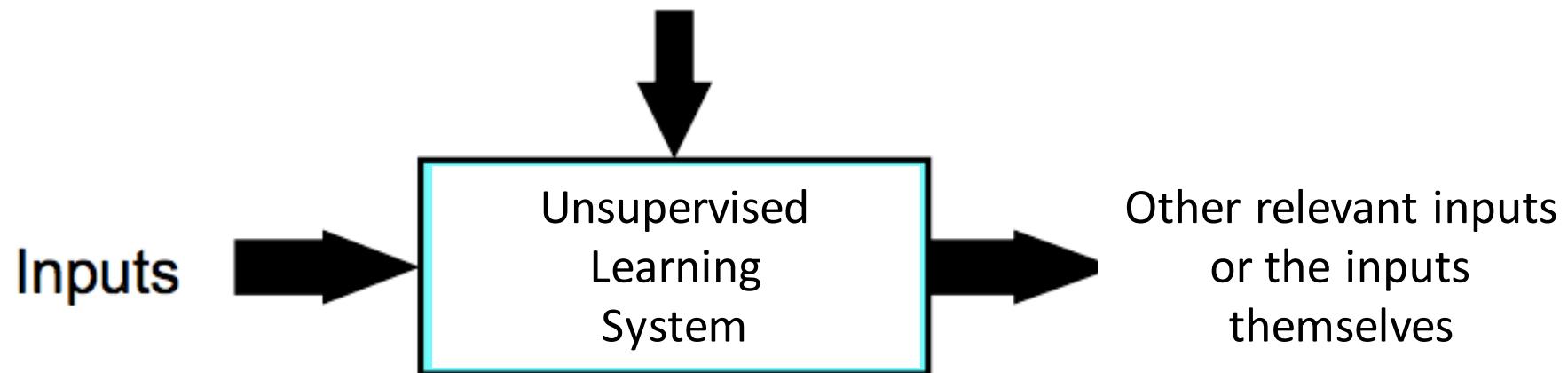


Error = (target output – actual output)

Cons: mostly need a lot human annotation
for desired tasks

Unsupervised Learning

Training Info = desired (target) outputs



Error = (target output – actual output)

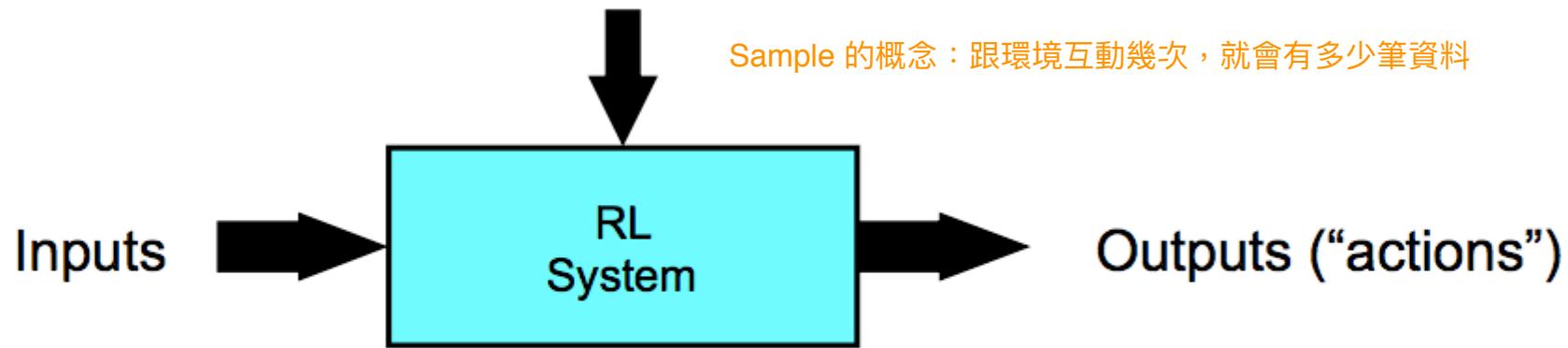
Cons: mostly learning a relevant representation but not directly for achieving desired tasks

Reinforcement Learning

Q : 偶爾給label那跟Semi Supervise有什麼差別？

A : Semi 沒有強調跟環境的互動

Training Info = evaluations (“rewards” / “penalties”)



解決兩個硬傷：
Supervise Learning中每個資料都要標籤的問題，RL只要偶爾告訴他做的好不好（顯然是跟他之前做的事有關）
第二個硬傷是資料搜集的問題，RL會自己去環境中找輸入。

Objective: get as much reward as possible

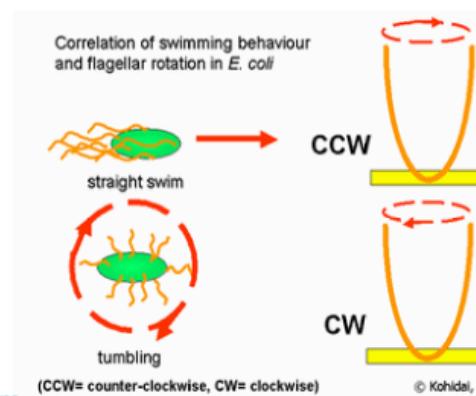
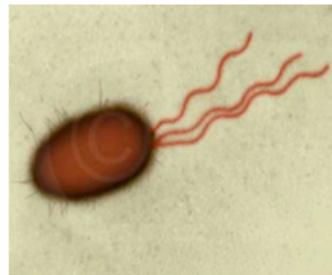
Pros: learning to achieve high reward by itself

Recipe for creative behavior: explore & exploit

- Creativity: finding a new approach /
- solution / ...
- – Exploration (random / systematic / ...) –
Evaluation (utility = expected rewards) –
Selection (ongoing behavior and learning)

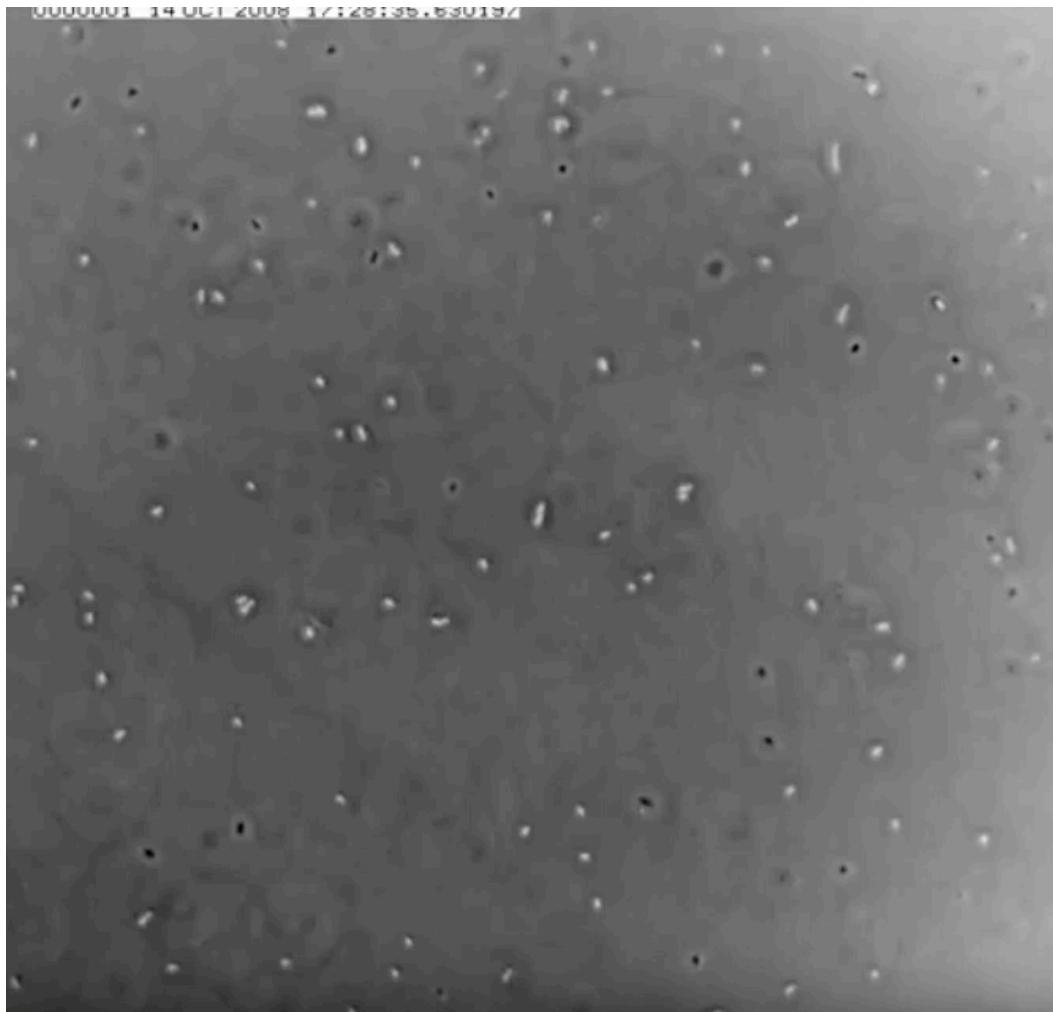
Coli bacteria (大腸桿菌) and creativity

- Escherichia Coli searches for food using trial and error:
 - Choose a random direction by tumbling and then start swimming straight
 - Evaluate progress
 - Continue longer or cancel earlier depending on progress



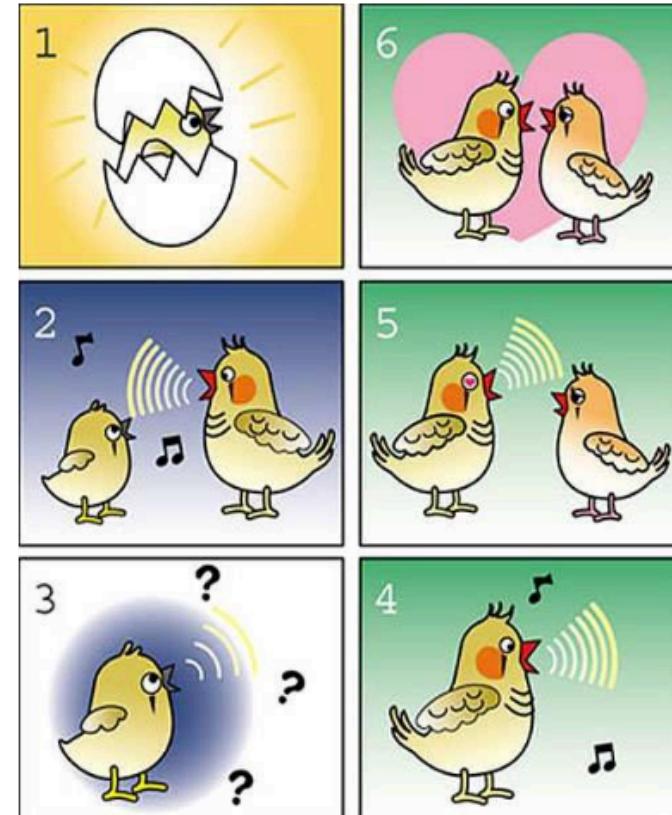
<http://biology.about.com/library/weekly/aa081299.htm>

Mobility of Coli bacteria



Zebra finch: from singing in the shower to performing artist

1. A newborn zebra finch can't sing
2. The baby bird listens to father's song
3. The baby starts to "babble" father's song as a target template
4. The song develops through trial and error – "singing in the shower"
5. No exploration when singing to a female



<http://www.brain.riken.jp/bsi-news/bsinews34/no34/speciale.html>

Zebra finch: from singing in the shower to performing artist



Learning to Walk a Quadruped



Learning to Walk a Quadruped



https://www.youtube.com/watch?v=o_wQe8F9fUc

Learning to Fly A Helicopter



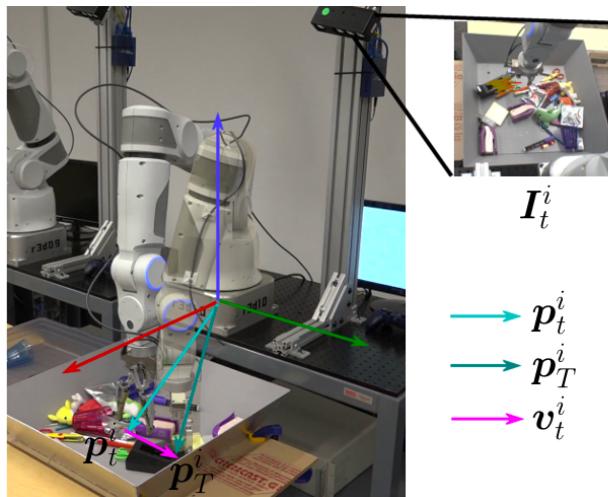
<https://www.youtube.com/watch?v=M-QUkgk3HyE>

More Robotic Applications

- Driving



- Grasping



End-effector

- Rotation
- Translation
- Grasp

Learning to Drive a Car

Performance

Trained for 196 million steps

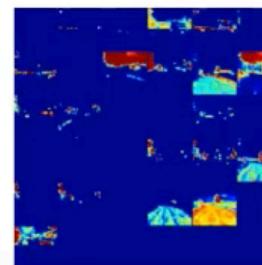


Game graphics

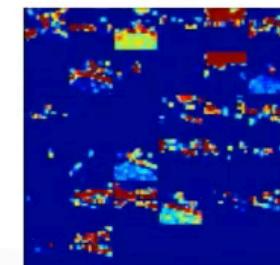
Test on training track



Network input and guided backpropagation



Layer 1



Layer 2

Snow (SE)



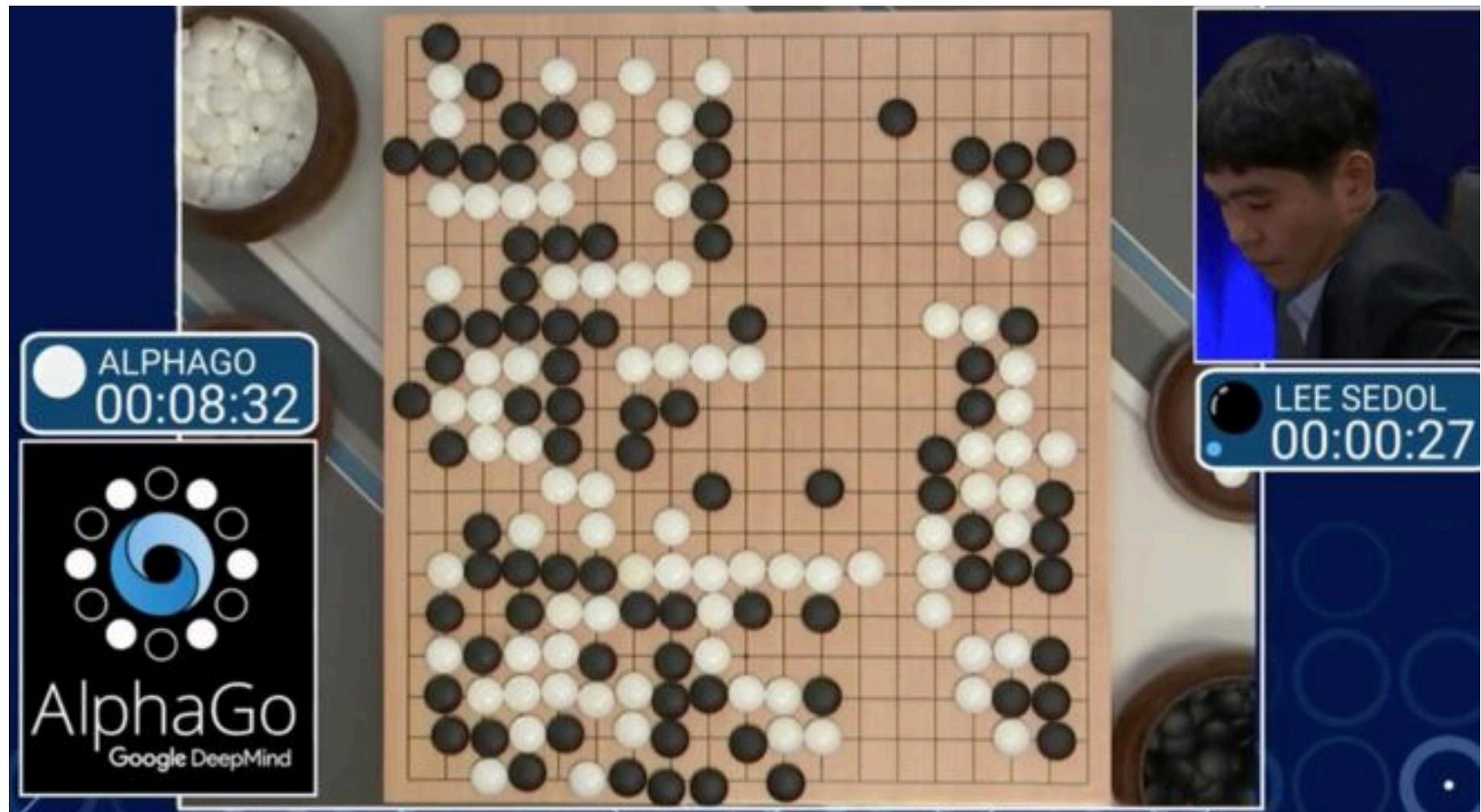
<https://www.youtube.com/watch?v=e9jk-lBWFlw>

Learning to Play Atari Game



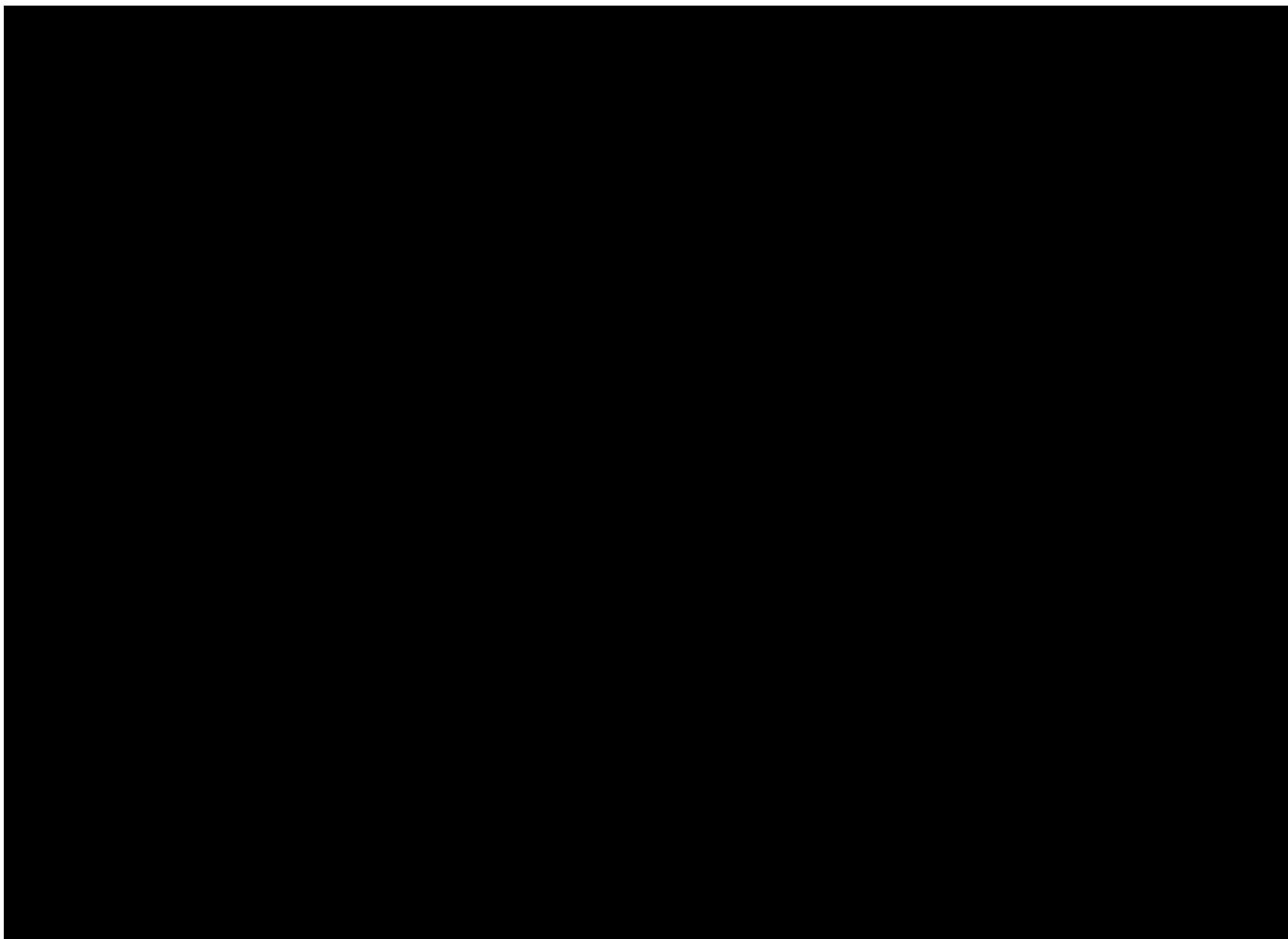
<https://www.youtube.com/watch?v=TnUYcTuZJpM>

AlphaGo = See and Place Stone



2016 by Google DeepMind

Learning to Parkour



Discussion