Explorations in Reinforcement Learning Curriculum Connections

Kathryn Meier
kmeier@sbcsc.k12.in.us
Riley HS, ***AP Biology***, Chemistry, Robotics Sponsor
John Gensic
john.gensic@gmail.com
Penn HS, ***Dual Credit Biology*** (Biology II)

Overview

- Task #1: Reinforcement Learning (RL) Basics...Biology Connection
- Task #2: RL and Codes
- Curriculum and Materials

What Is Our Project?

Current state of research with Mike’s group is using CNNs and DNNs for object recognition in images.

Problem: Reinforcement learning presents opportunities for the group to use different algorithms to increase efficacy and decrease training time and resources, but they don’t know much about RL or how it works.

Research Task #1: Learn about RL and communicate the basics.

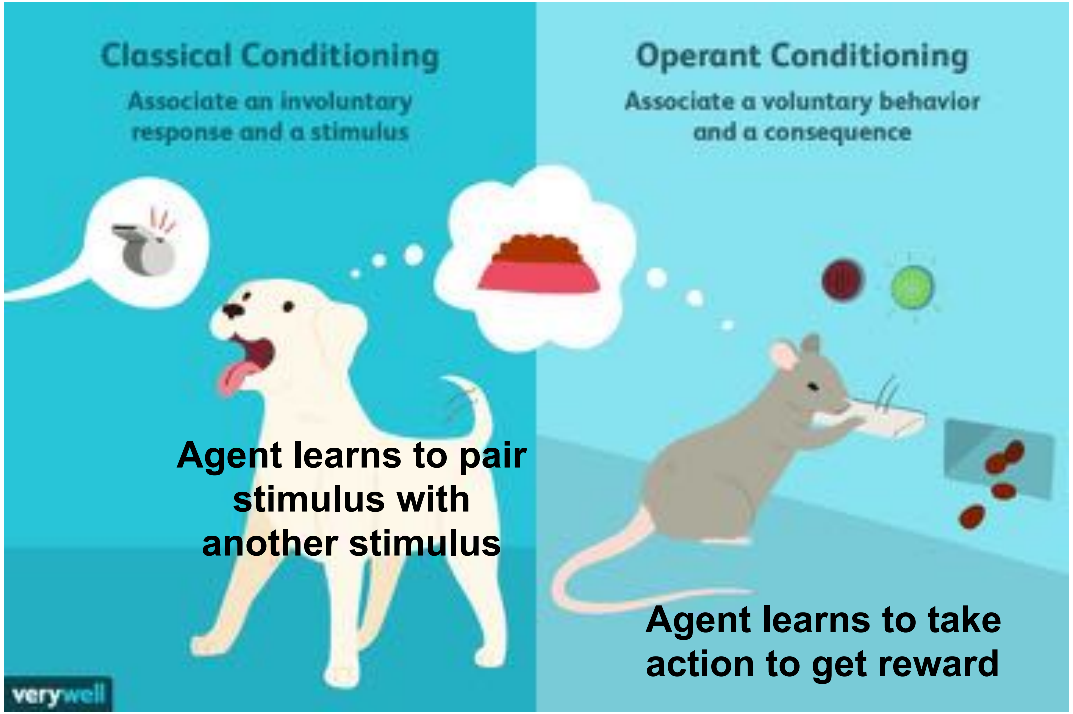
Biological Connections...

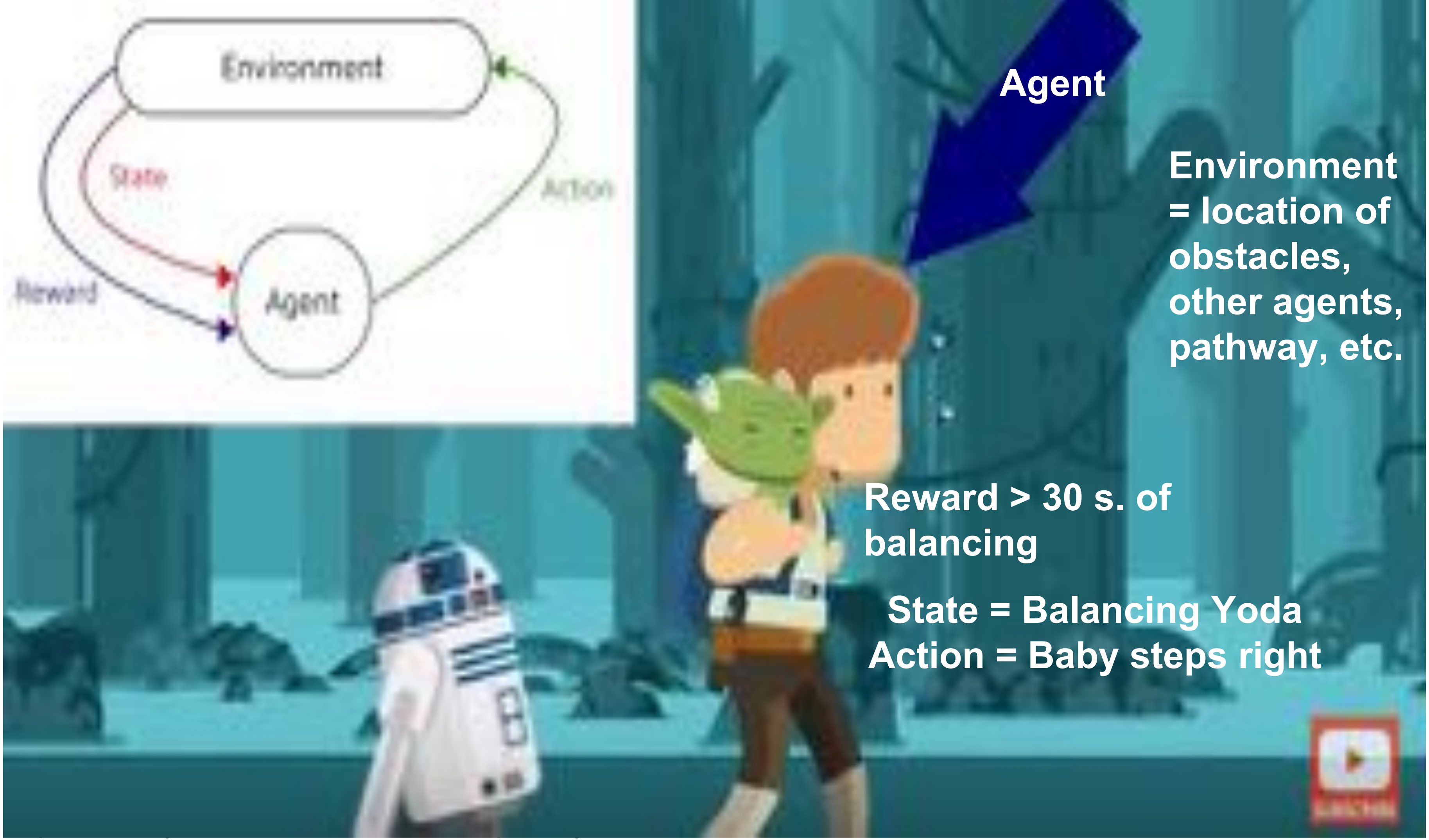
Traditional AI:classical conditioning as Reinforcement learning: operant conditioning

	Biology	RL
Rewards	**Dopamine, Glutamate, Epinephrine, etc.**	**Points**
Stimulus	**Environment**	**Environment**
Rewards lead to...	**Excitation of neurons**	**Change of state**

Like traditional ML

Like reinforcement learning




Environment
State
Action
Reward
Agent
Agent
Environment = location of obstacles, other agents, pathway, etc.
Reward > 30 s. of balancing
State = Balancing Yoda
Action = Baby steps right

Reinforcement Learning Example #2

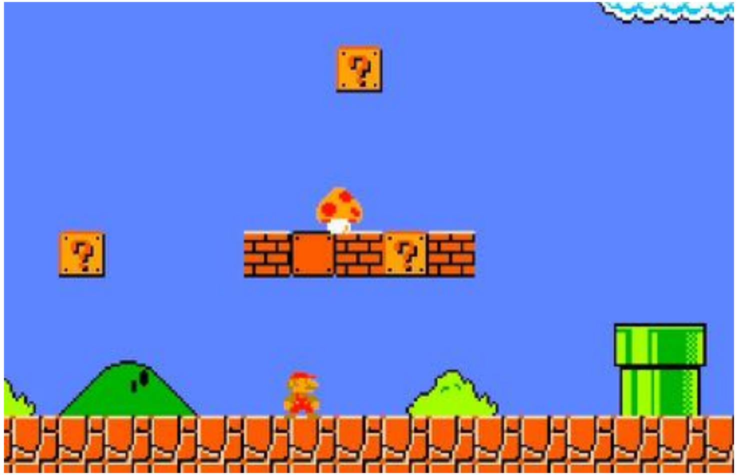
Agent: Mario

Environment: location of ? blocks, location of brick blocks, location of warp pipe, presence/location/direction of Koopa Troopa, any other obstacles, location of mushroom or other power ups, presence/location of coins

State: small Mario, location & movement of Mario, lack of power-ups, timestep, current score

Actions: move left, move right, jump, run left, run right

Reward: points accrued, not “dying”, power-ups



Overview of RL Algorithms

Algorithm	Model	Policy	Action Space	Observation Space	Operator
Q-Learning	Model-Free	Off-policy	Discrete	Discrete	Q-value
SARSA	Model-Free	On-policy	Discrete	Discrete	Q-value
DQN	Model-Free	Off-policy	Discrete	Continous	Q-value
DDPG	Model-Free	Off-policy	Continous	Continous	Q-value
TRPO	Model-Free	Off-policy	Continous	Continous	Advantage
PPO	Model-Free	Off-policy	Continous	Continous	Advantage

https://towardsdatascience.com/introduction-to-various-reinforcement-learning-algorithms-part-ii-trpo-ppo-87f2c5919bb9

Q-Learning

$$Q^{\pi}(s,a) = E\left[\sum_{t=0}^{\infty} \gamma^t \, r_{t+1} \, | s,a\right]$$

Uses the Bellman equation to estimate the Q-value

There are many variations on this equation. Some use V for value rather than Q.

Deep Q Network (DQN)

Pairs Q-learning with a DNN to define the state in an unknown situation

Inputs are frames from the environment

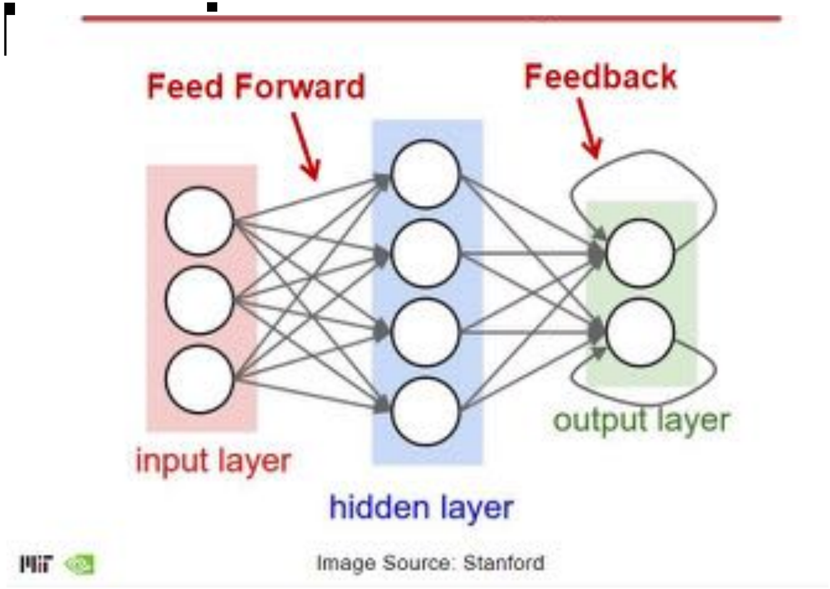
Hidden layers extract features

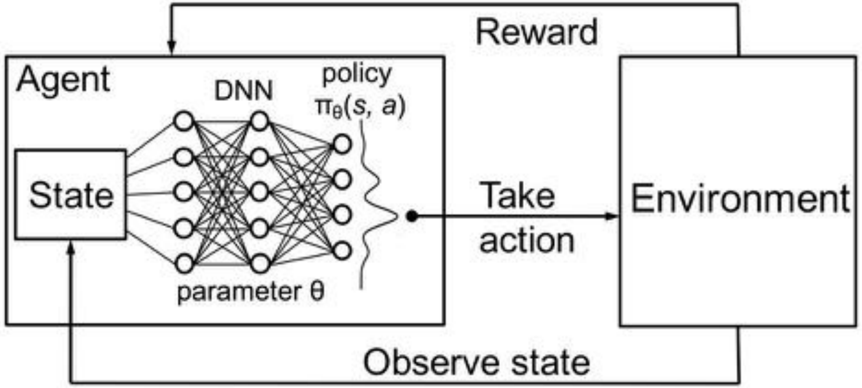
Outputs are predictions for reward earned by different actions taken

How Does RL in Machines Work?

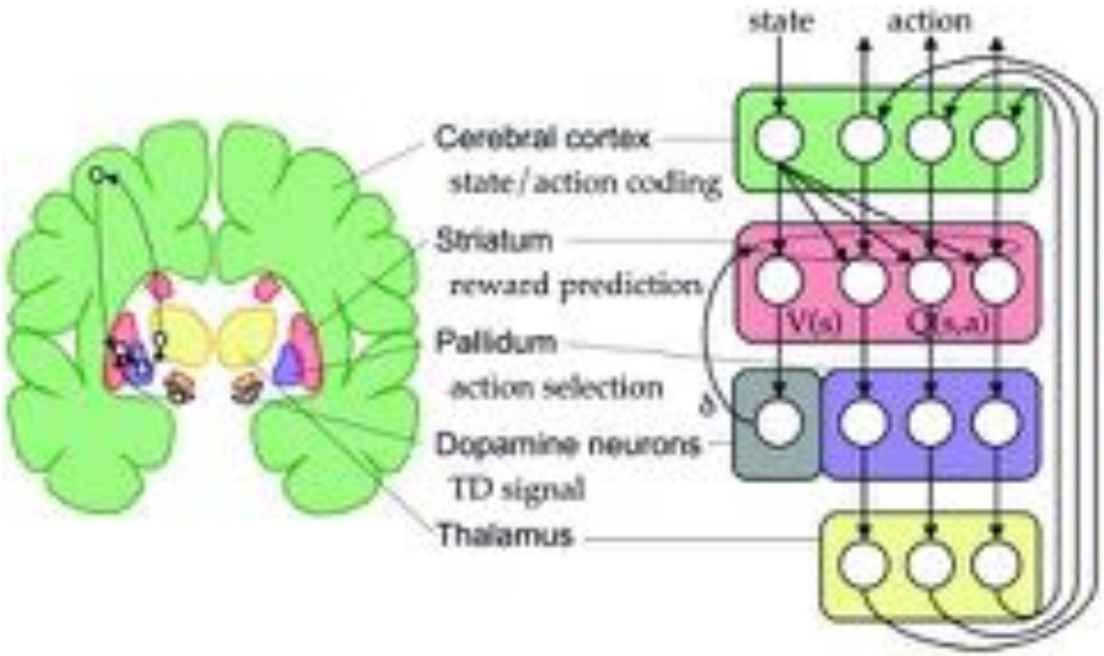
Visual comparison of...

DNN OR CNN Alone | **Reinforcement**





Biological Connections…(cont.)

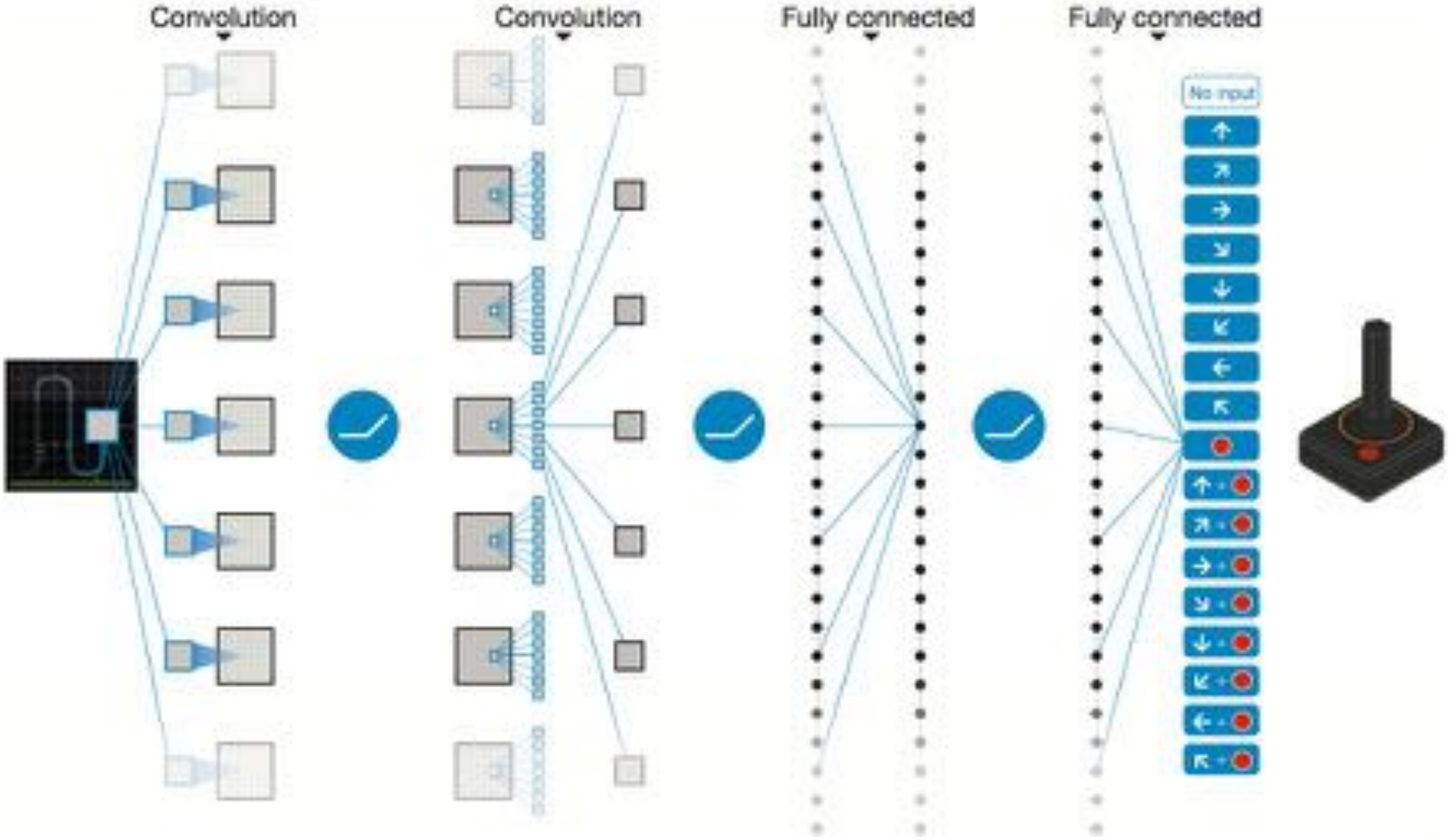


r = reward

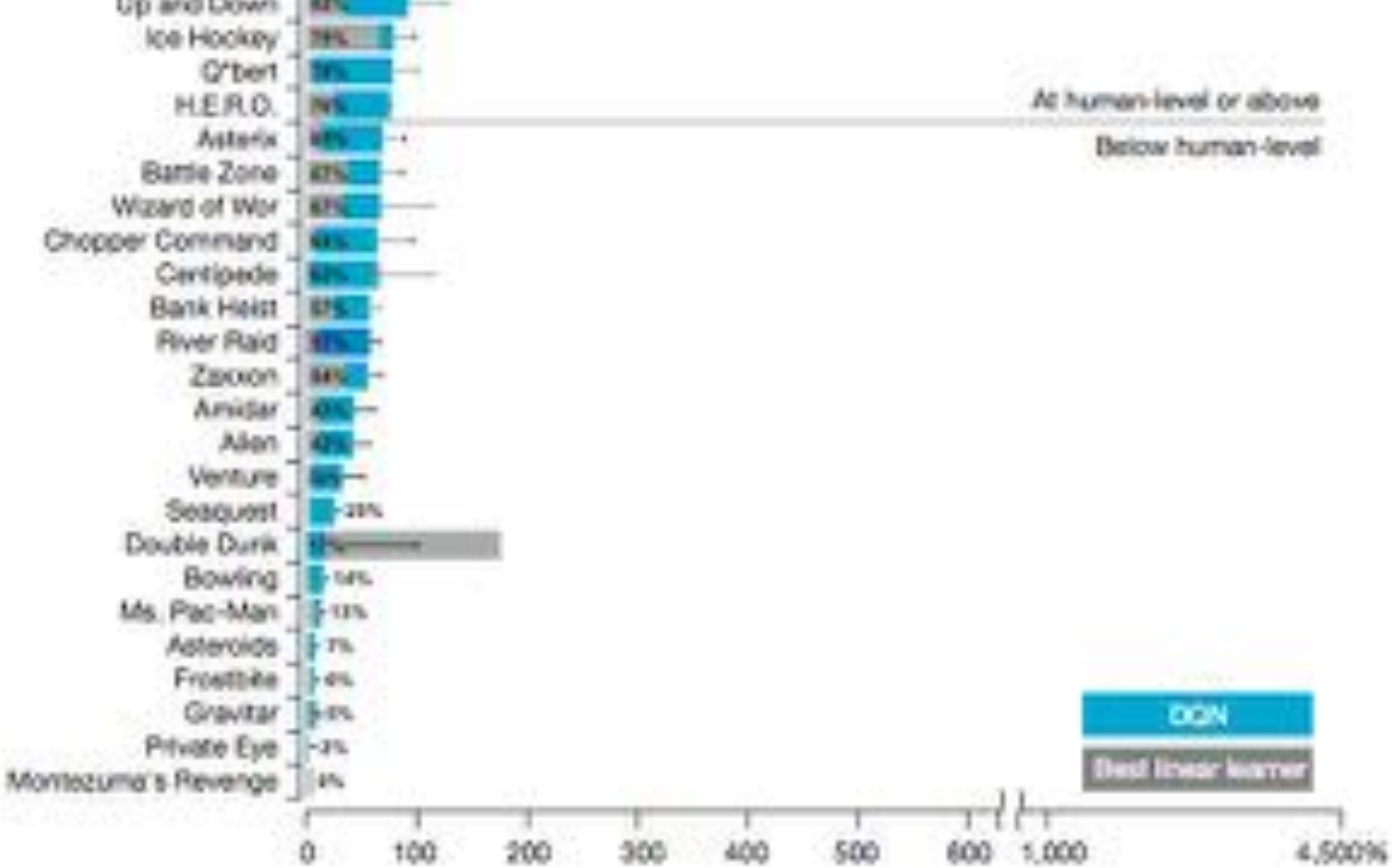
V = state, reward predicted

δ = expected behavior, action? Or Q? Dopamine release

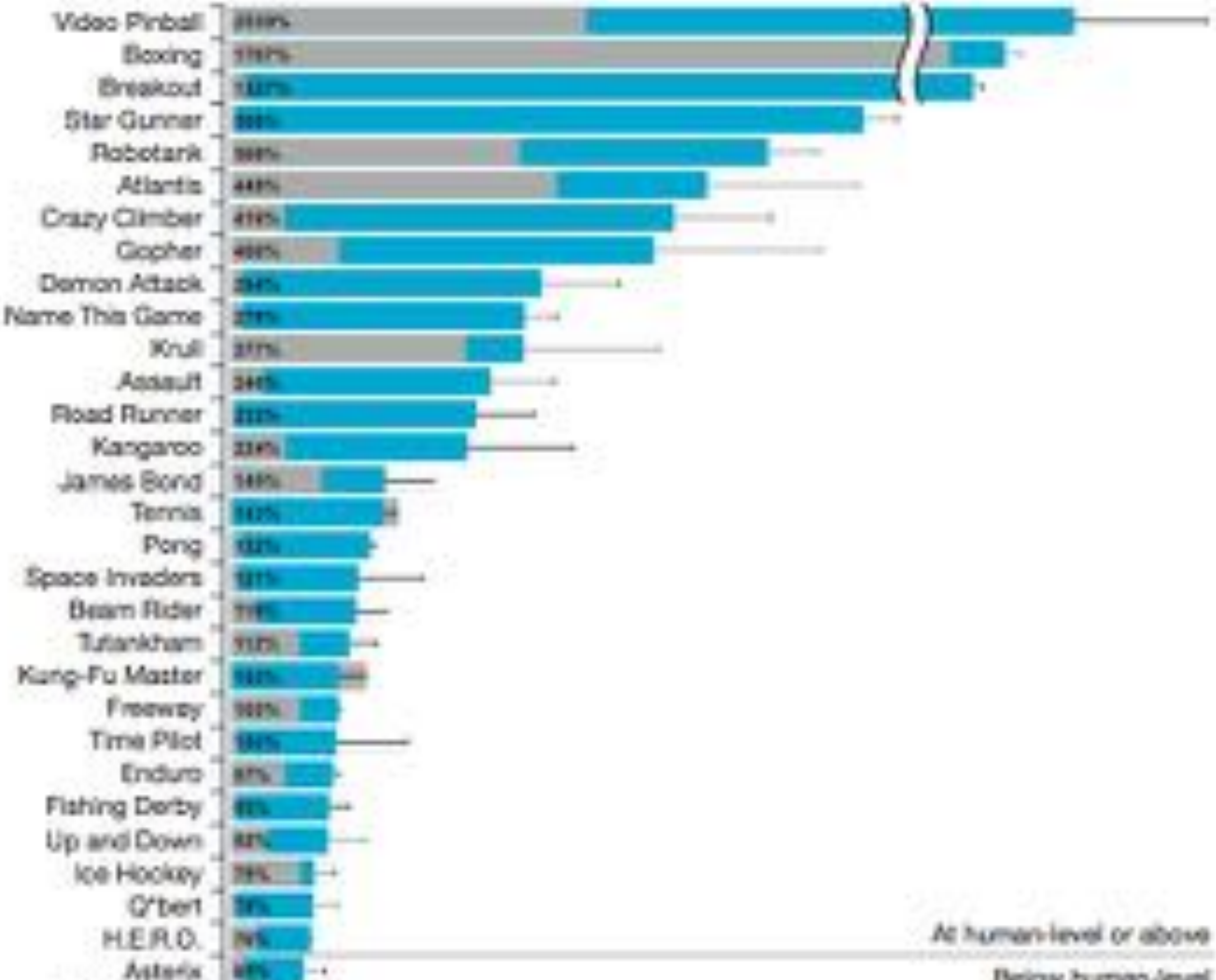
https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2645553/



Convolution
Convolution
Fully connected
Fully connected
No input

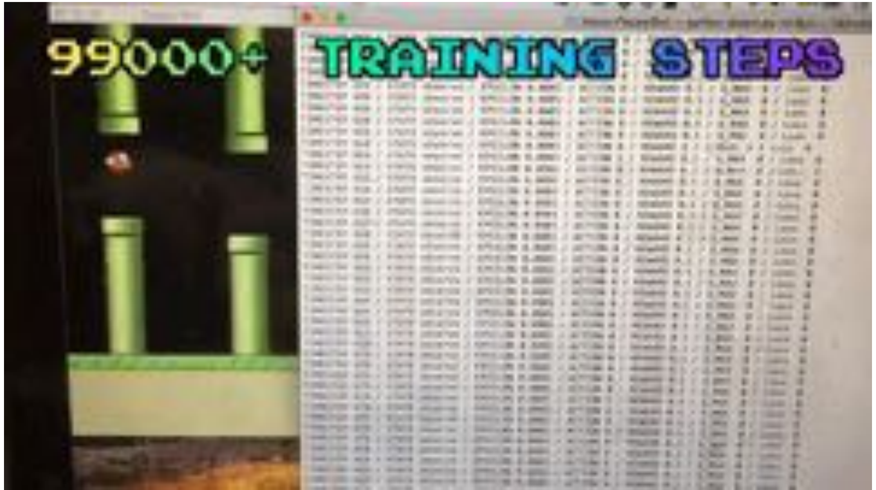


Ice Hockey
Q*bert
H.E.R.O.
At human-level or above
Below human-level
Asterix
Battle Zone
Wizard of Wor
Chopper Command
Centipede
Bank Heist
River Raid
Zaxxon
Amidar
Alien
Venture
Seaquest
Double Dunk
Bowling
Ms. Pac-Man
Asteroids
Frostbite
Gravitar
Private Eye
Montezuma's Revenge
DQN
Best linear learner
0
100
200
300
400
500
600
1,000
4,500%



Video Pinball
Boxing
Breakout
Star Gunner
Robotank
Atlantis
Crazy Climber
Gopher
Demon Attack
Name This Game
Krull
Assault
Road Runner
Kangaroo
James Bond
Tennis
Pong
Space Invaders
Beam Rider
Tutankham
Kung-Fu Master
Freeway
Time Pilot
Enduro
Fishing Derby
Up and Down
Ice Hockey
Q*bert
H.E.R.O.
Asterix
At human-level or above

An Example to DQN to Learn a Game - Flappy Bird



Research Task #2: Finding code for comparison...

Another group used RL on a chip they designed and showed large energy saving when training a micro-robot in 3D space with ultrasonic sensors.

Problem: How can Mike's group compare apples to apples and test their digital method to compare power use?

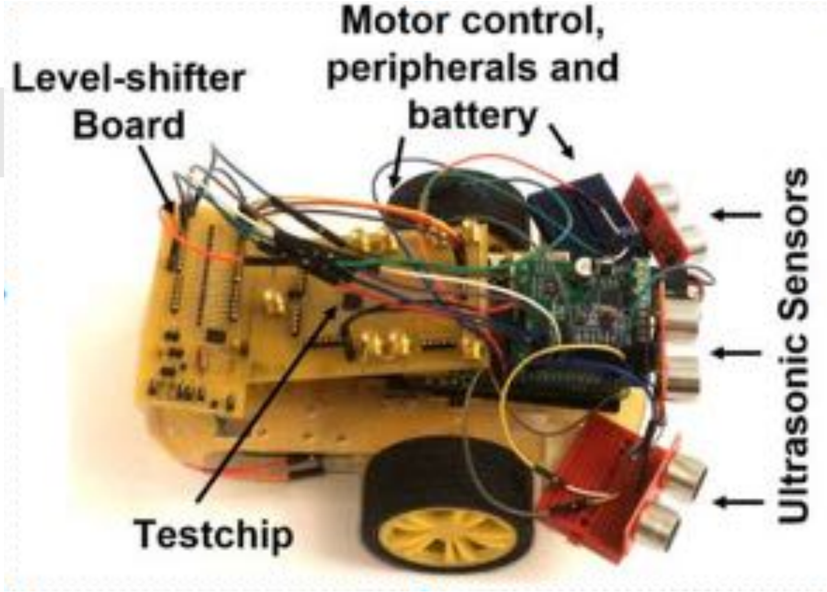
Task #2: Find code that might be useful in starting to integrate RL with the work already happening here.

7.4 A 55nm Time-Domain Mixed-Signal Neuromorphic Accelerator with Stochastic Synapses and Embedded Reinforcement Learning for Autonomous Micro-Robots

Anvesha Amravati, Saad Bin Nasir, Sivaram Thangadurai, Insik Yoon, Arijit Raychowdhury

Georgia Institute of Technology, Atlanta, GA

- 2018 IEEE International Solid-State Circuits Conference



What did they find?

	This work	[1]	[2]	[3]	[4]	[5]
ML System	Reinforcement Learning	Object Recognition	CNN-RNN	CNN	DNN	DNN
Technology	55nm	65nm	65nm	180nm	65nm	65nm
Circuit style	Time domain mixed-signal	Digital	Digital	Digital	Digital	Digital
Area	$3.4mm^2$	$4mm^2$	16mm2	$3.3mm^2$	$16mm^2$	$16mm^2$
Learning/Training	Online in real time	Offline	Offline	Offline	Offline	Offline
Stochasticity	Present	Absent	Absent	Absent	Absent	Absent
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Power	690 uW at peak performance	121mW	63mW	7.5-300mW	45mW	278mW
Supply voltage	0.4-1V	1.2V	0.77-1.2V	Unavilable	1.2V	0.82-1.17V
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Performance/Watt	3.12TOPS/W	1.24TOPS/W	2.1TOPS/W	0.26-10TOPS/W	1.42TOPS/W	0.21TOPS/W
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Min. Energy/inference	690pJ	Not Reported	Not Reported	Not Reported	Not Reported	Not Reported
Min. energy/training	1.5nJ	Not Reported	Not Reported	Not Reported	Not Reported	Not Reported

An RC Car that uses RL

Matt Harvey Follow

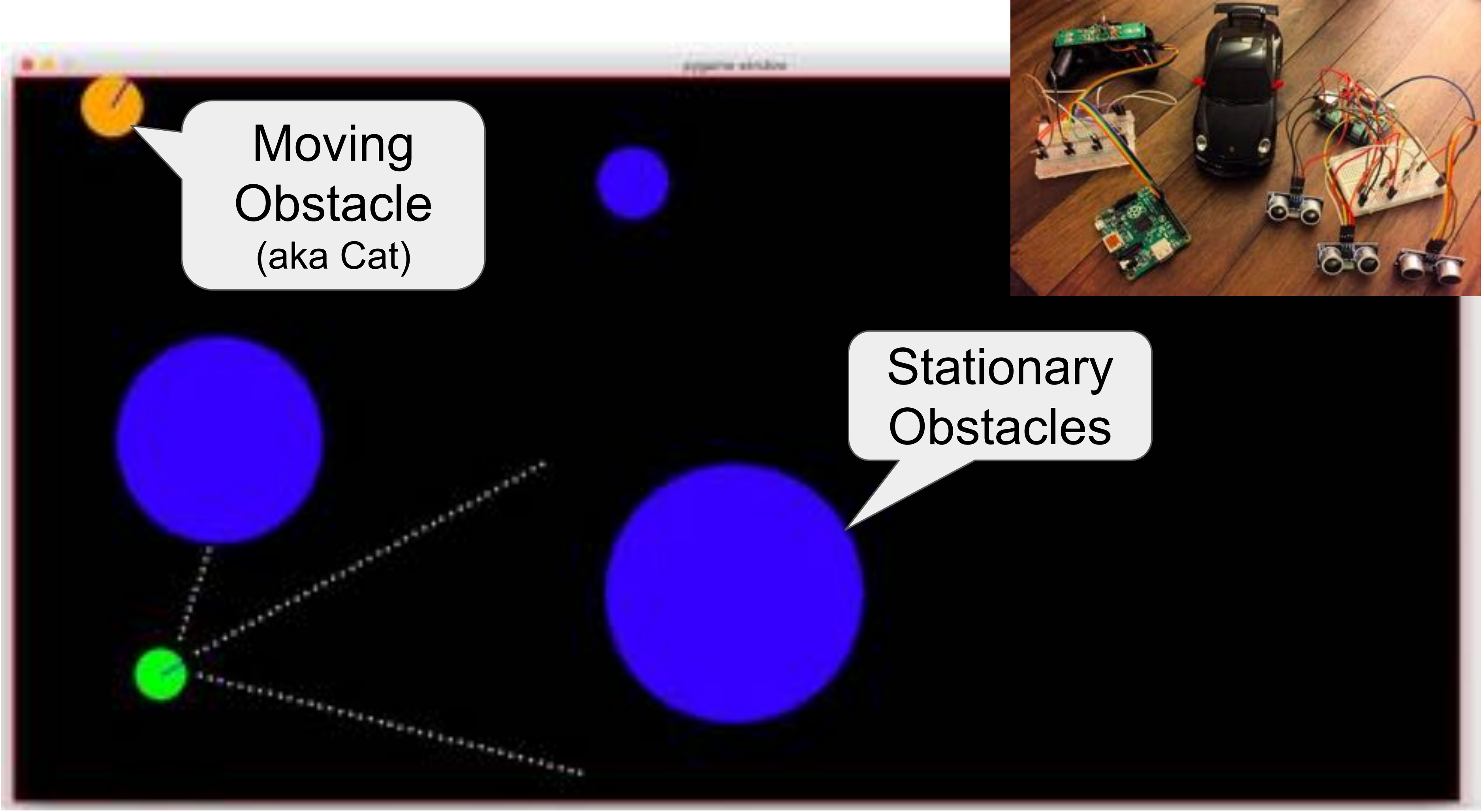
Founder of Coastline Automation, using AI to make every car crash-proof.

Feb 6, 2016 · 9 min read

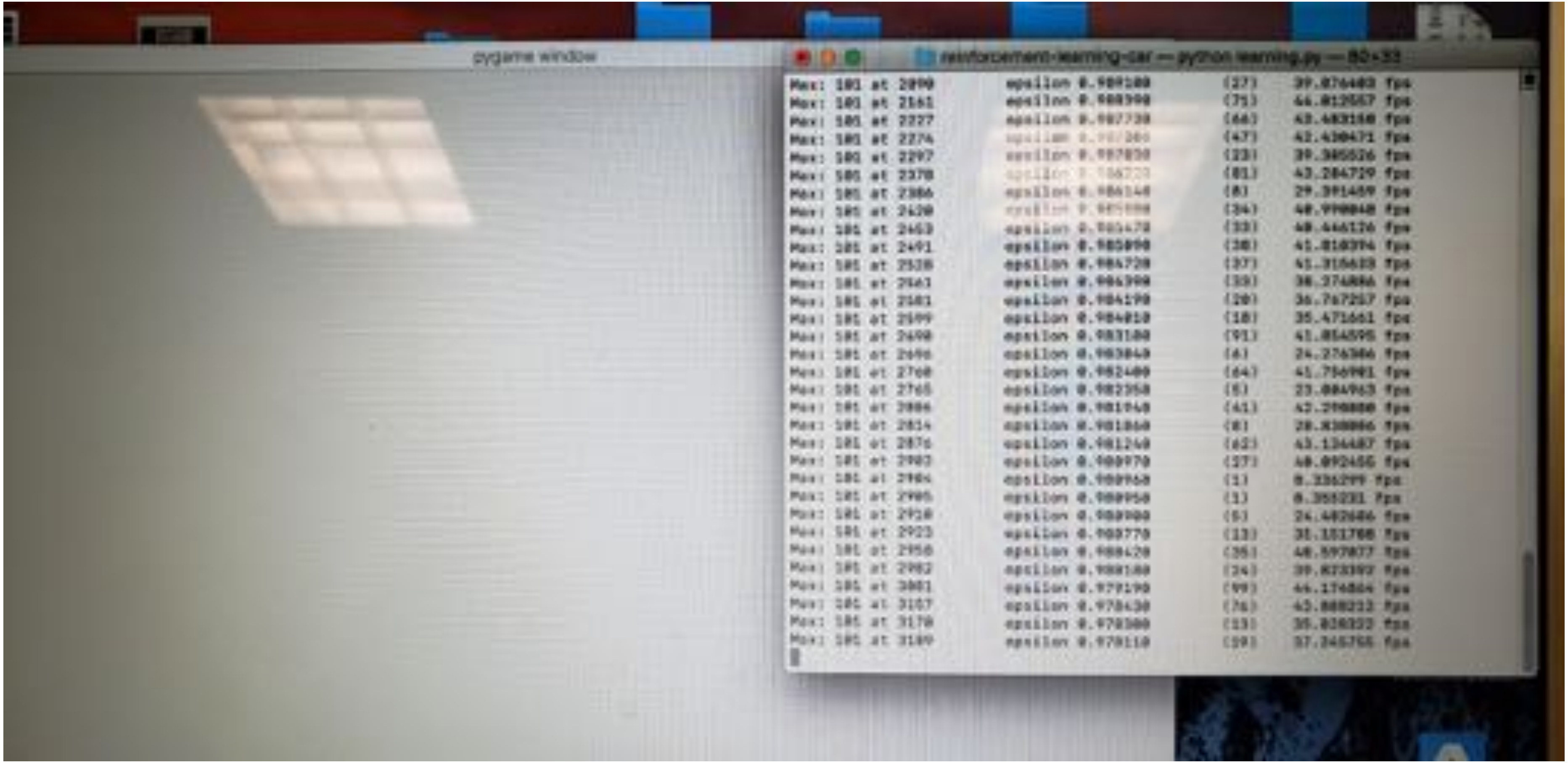
Using reinforcement learning in Python to teach a virtual car to avoid obstacles

An experiment in Q-learning, neural networks and Pygame.

I'd like to build a self-driving, *self-learning* RC car that can move around my apartment at top speed without running into anything—especially my cats.


Moving Obstacle
(aka Cat)
Stationary Obstacles

We Got the Code Running. Sort Of...



John...Curriculum Materials Developed...

- Self Driving Cars Debate with Evidence. *Theme: AI, Adversarial Agents*
- Collecting Phone Use data Checky App and Moment App. *Bio Reinforcement*
- Should Cell Phone Addiction be a Mental Illness Debate. *Theme: Reinforcement, Dopamine Rewards, Exploitation*
- What is being scored? *Theme: AI, Adversarial Agents*
- AIY Stations *Theme: Neural Networks, Biological Connections*
- RL Success after different numbers of training steps data collection (next slides). *Theme: Reinforcement Learning*
- Cell Part Classifier *Theme: CNN's*
- To still investigate…
  - Pavlok Negative Reinforcement Biting Nails, Checking email
  - Muse Brain Wave reader, testing waves in different situations, positive reinforcement for meditation, phone use.
  - *Themes: Reinforcement learning (biology and computer)*

Is cell phone/internet/gaming addiction an illness (to be recognized in DSM 5)?

Yes it is an illness and insurance should help cover treatment.

WHO Official

Parent of
"Addicted" Teen

Teacher

Mental Health Professional

No it is not an illness so insurance should not help cover treatment

Insurance Company
Executives

Parent of
"Unaddicted" Teen

Teacher

Video Game
Designer

Should self driving cars be allowed?

For self driving cars

Truck Driver

Business Owner

Car Engineer

Computer Scientist

Against self driving cars

Truck Driver

Family member of victim struck by self driving car

Driving Instructor

Neuroscientist

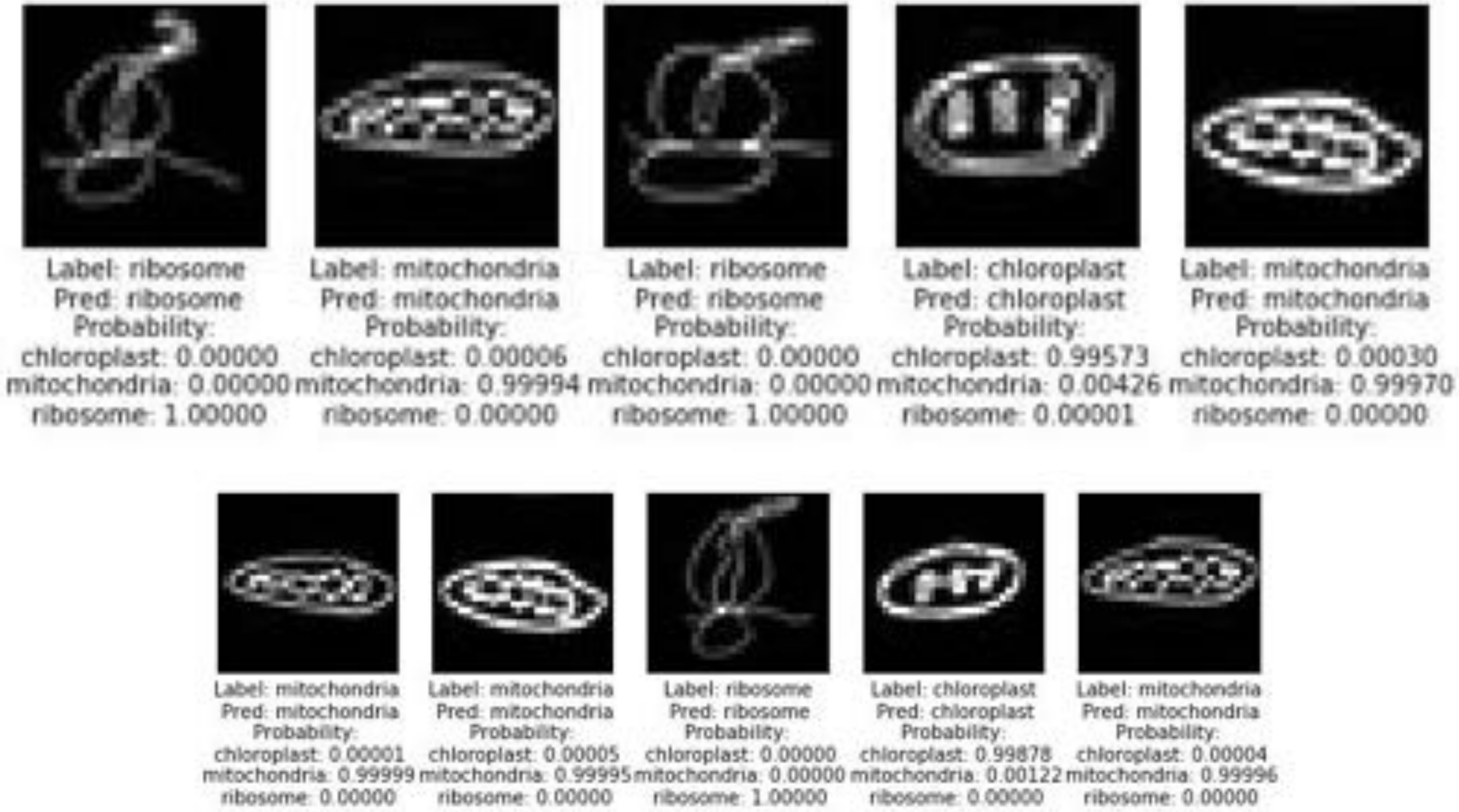
AIY Stations...Google AIY Joy Detection and Adversarial Examples

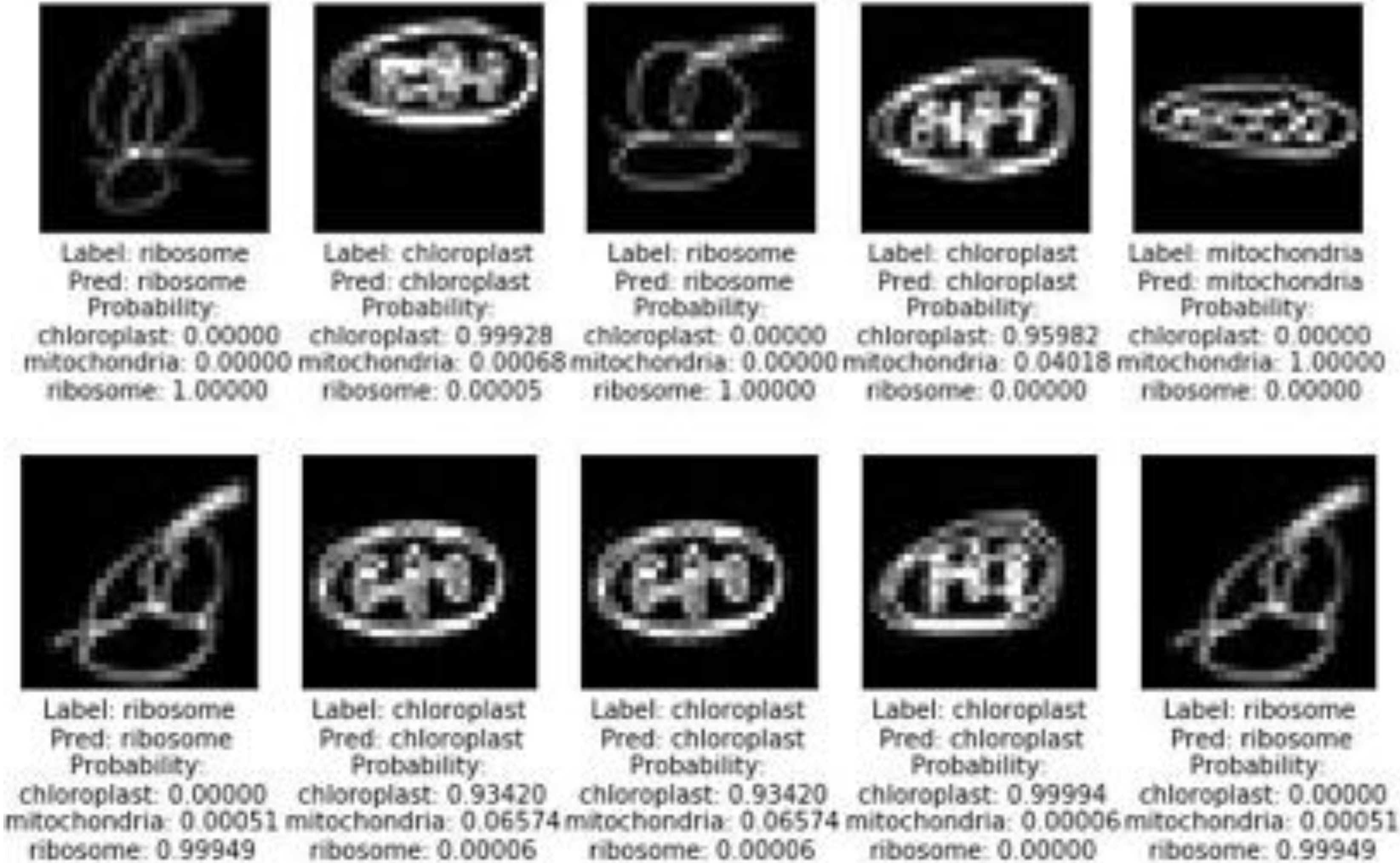
-used in evaluating advertisements

MNIST Classification
Cell Part Style...Mitochondria

Chloroplasts...

Ribosomes...


Label: ribosome
Pred: ribosome
Probability:
chloroplast: 0.00000
mitochondria: 0.00000
ribosome: 1.00000
Label: mitochondria
Pred: mitochondria
Probability:
chloroplast: 0.00006
mitochondria: 0.99994
ribosome: 0.00000
Label: ribosome
Pred: ribosome
Probability:
chloroplast: 0.00000
mitochondria: 0.00000
ribosome: 1.00000
Label: chloroplast
Pred: chloroplast
Probability:
chloroplast: 0.99573
mitochondria: 0.00426
ribosome: 0.00001
Label: mitochondria
Pred: mitochondria
Probability:
chloroplast: 0.00030
mitochondria: 0.99970
ribosome: 0.00000
Label: mitochondria
Pred: mitochondria
Probability:
chloroplast: 0.00001
mitochondria: 0.99999
ribosome: 0.00000
Label: mitochondria
Pred: mitochondria
Probability:
chloroplast: 0.00005
mitochondria: 0.99995
ribosome: 0.00000
Label: ribosome
Pred: ribosome
Probability:
chloroplast: 0.00000
mitochondria: 0.00000
ribosome: 1.00000
Label: chloroplast
Pred: chloroplast
Probability:
chloroplast: 0.99878
mitochondria: 0.00122
ribosome: 0.00000
Label: mitochondria
Pred: mitochondria
Probability:
chloroplast: 0.00004
mitochondria: 0.99996
ribosome: 0.00000


Label: ribosome
Pred: ribosome
Probability:
chloroplast: 0.00000
mitochondria: 0.00000
ribosome: 1.00000
Label: chloroplast
Pred: chloroplast
Probability:
chloroplast: 0.99928
mitochondria: 0.00068
ribosome: 0.00005
Label: ribosome
Pred: ribosome
Probability:
chloroplast: 0.00000
mitochondria: 0.00000
ribosome: 1.00000
Label: chloroplast
Pred: chloroplast
Probability:
chloroplast: 0.95982
mitochondria: 0.04018
ribosome: 0.00000
Label: mitochondria
Pred: mitochondria
Probability:
chloroplast: 0.00000
mitochondria: 1.00000
ribosome: 0.00000
Label: ribosome
Pred: ribosome
Probability:
chloroplast: 0.00000
mitochondria: 0.00051
ribosome: 0.99949
Label: chloroplast
Pred: chloroplast
Probability:
chloroplast: 0.93420
mitochondria: 0.06574
ribosome: 0.00006
Label: chloroplast
Pred: chloroplast
Probability:
chloroplast: 0.93420
mitochondria: 0.06574
ribosome: 0.00006
Label: chloroplast
Pred: chloroplast
Probability:
chloroplast: 0.99994
mitochondria: 0.00006
ribosome: 0.00000
Label: ribosome
Pred: ribosome
Probability:
chloroplast: 0.00000
mitochondria: 0.00051
ribosome: 0.99949

AP Biology Curriculum

Big Idea 1: Evolution

Big Idea 2: Energy

Big Idea 3: Information

Big Idea 4: Interactions

AP Biology Curriculum

Big Idea 1: Evolution

Big Idea 2: Energy

Big Idea 3: Information

- Nervous system

Big Idea 4: Interactions

AP Biology Curriculum

Big Idea 1: Evolution

Big Idea 2: Energy

Big Idea 3: Information

- Nervous system

Big Idea 4: Interactions

- Animal behavior
- Response to stimulus

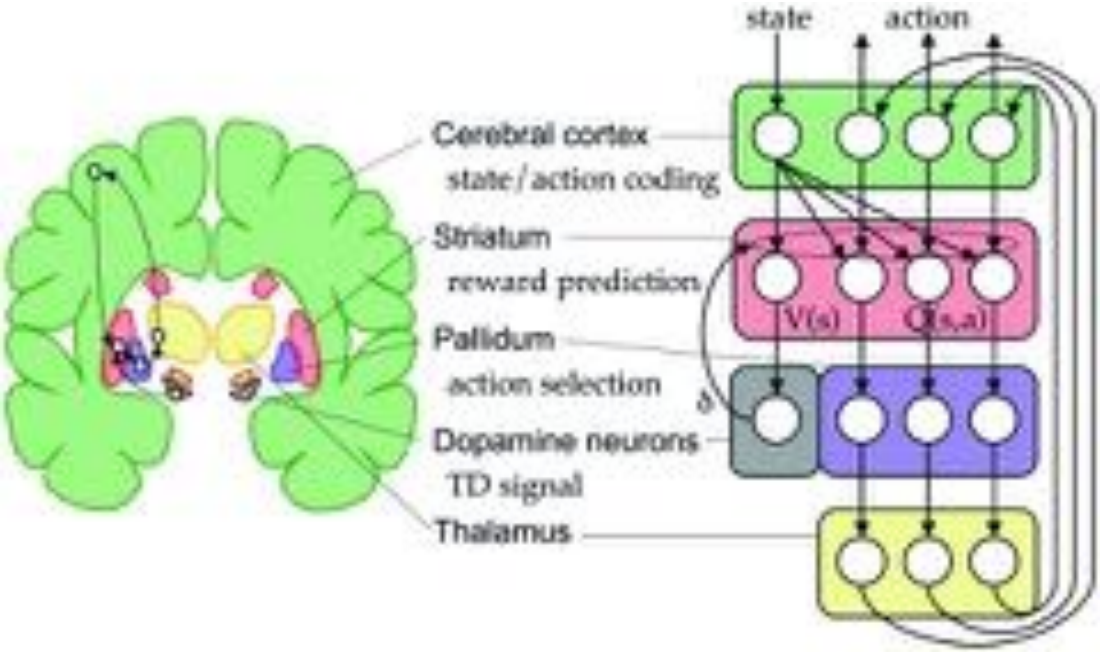
AP Biology Curriculum

Big Idea 1: Evolution
Big Idea 2: Energy
Big Idea 3: Information
- Nervous system

Big Idea 4: Interactions
- Animal behavior
- Response to stimulus

Science Practices:
- Use models
- Engage in scientific questioning
- Connect and relate knowledge across domains

Modeling the Nervous System



Nervous System

- Model the process of neurons detecting a stimulus, generating a response, and sending out a new signal

RET Technology in My Classroom

Students will use Google AIY Vision kits to investigate how neural networks detect features and identify objects.

- Day 1:
  - “Play” with the existing model
  - Discuss what the AI seems to be doing & how it might have been trained
- Day 2
  - design procedure to train the network to detect different objects Begin training protocol

A Meaningful Lab Activity for the Nervous System!

Students will use Google AIY Vision kits to investigate how neural networks detect features and identify objects.

- Day 3-4:
  - Train the network using their own design
- Day 4-5:
  - Test the model & collect data for analysis

Student Products for Assessment

Student groups will produce a “mini-poster” showing

- their analysis of the existing model of object recognition
- their thought process in developing a training regimen
- the metrics they used to evaluate the trained model
- their analysis their model’s ability to recognize the objects presented

Individual students will produce a reflection about the use of biological systems as inspiration for computer science.

Module Extension

Students traditionally complete an animal behavior lab activity using fruit flies in a choice chamber.

I will supplement this activity with activities involving an autonomous mini-robot that uses a neural network for object detection.

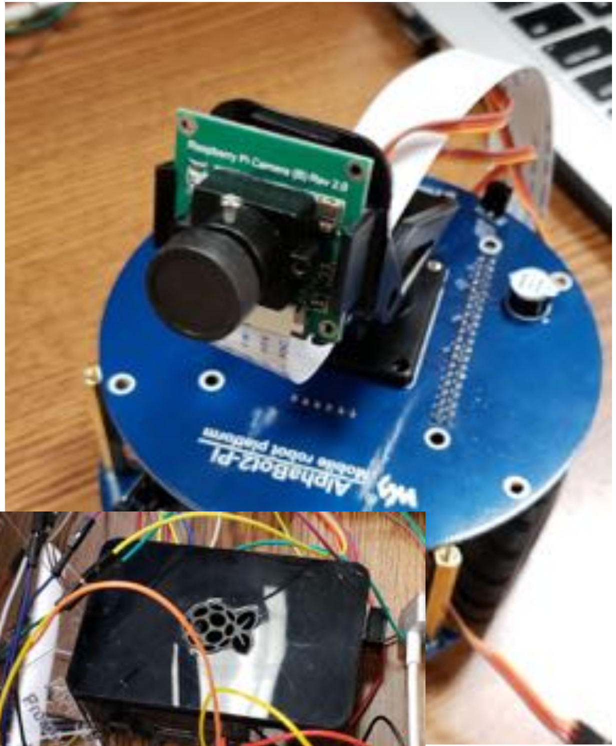
Module Extension

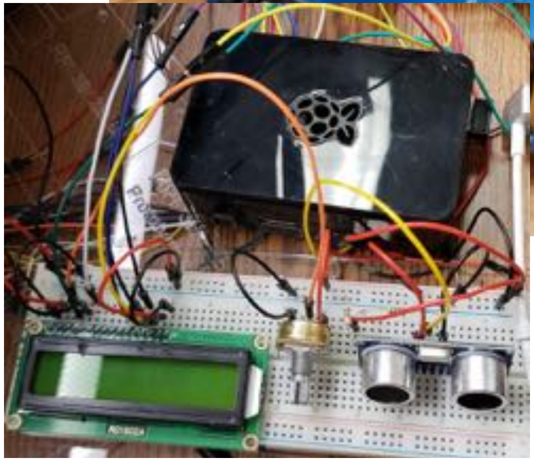
The robot runs on Raspberry pi and uses a camera to detect objects.

Module Extension

The robot runs on Raspberry pi and uses a camera to detect objects.

I am working on also integrating ultrasonic sensors into the robot.





Making Connections Across Domains

Students will

- Spend time observing the robot’s “behavior”
- Identify stimuli that cause changes in the robot’s actions
- Discuss feedback and reinforcement that are occurring
- Relate the concepts of reinforcement learning in machines to animal behavior including:
    - Predator warning
    - Pack behavior
    - Culture acquisition