How Can Birdsong Inform Learning Algorithms?

Introduction

Several people at Carnegie Mellon expressed interest in the potential applications of my research program to machine learning (ML) algorithms and artificial intelligence (AI). This is an informal document with some examples of potential insights into ML that might come from my research aims. I would appreciate any comments or feedback.

The songbird, with its highly tractable song circuit (Fig. 1), and naturally learned stereotyped motor output (song) is an excellent model system to address motor control. Many of the concepts used in the birdsong field have been borrowed from computer science and machine learning, such as reinforcement learning[1] (RL) and simulated annealing[2]. One of the goals of my research program is to contribute back to the ML literature with new learning algorithms or better implementations of known algorithms. I now briefly discuss three potential applications of my research aims to ML.

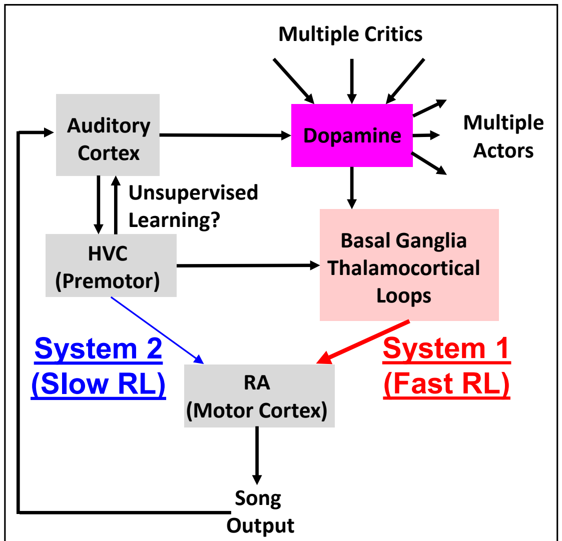

Fig 1. **(1)** Dopamine-basal ganglia-thalamocortical circuits in the songbird implement 2-step reinforcement learning (RL) for song learning. System 1 is fast and *biases* motor output over minutes and hours while system 2 is slow and *consolidates* learning over days. **(2)** The songbird brain must also solve a multi-objective reinforcement problem (song learning and place preference, for example). **(3)** Does reinforcement learning build on prior unsupervised learning?

Examples of potential applications

1. Two-Step Reinforcement Learning: Coupling Fast and Slow Systems for Efficient Learning

Motivation: Song learning is accomplished via a 2-step reinforcement mechanism, with what we may call System 1, which is fast and operates in minutes to hours, and System 2, which is slow and operates over day timescales (Fig. 1). System 1 is believed to implement RL via dopamine-mediated cortico-striatal plasticity[3] at the HVC-Medium Spiny Neuron (MSN) synapses in Area X (song-related basal ganglia). The output of basal ganglia circuitry biases motor output in RA (a motor cortical nucleus). System 2 consolidates this learning in HVC-RA synapses, typically overnight[3]. This fast learning followed by consolidation is reminiscent of hippocampus-mediated memory formation and cerebellum-mediated motor memory consolidation and might be a general feature of the brain. One hypothesis is that this 2-step RL is more efficient than single stage RL. There is now some evidence to support this idea from Larry Abbott's group[4] showing that song learning occurs too rapidly to be explained by single-stage RL. They propose that song learning occurs via quenched RL where vocal space is first explored and then motor output is consolidated in the direction of the best guess based on this exploration.

Approach and Outcomes: Following tutoring, juvenile songbirds begin to babble, and gradually learn their song over weeks. To test the role of DA and reinforcement during development, it is necessary to monitor DA activity over timescales of learning (hours to weeks). To this end, I built a fiber photometry system to chronically monitor DA activity in songbirds and developed a high channel (up to 64 channels) electrophysiology rig for stable, chronic recordings. The computational tools I have developed in collaboration with Adrienne Fairhall[4] will enable me to analyze the large data sets of natural fluctuations in song during babbling. The goal is to simultaneously acquire, for the first time, motor output (song) and the reward signal (dopamine activity) over the course of both fast learning (hours) and slow consolidation (days). This will allow us to test if birds implement algorithms such as quenched reinforcement learning and delineate the contributions of System 1 and System 2 to the learning rate. These experiments can also be carried out in adult birds that are modifying their songs over several days due to conditional auditory feedback.

2. Multi-Objective Reinforcement Learning

Motivation: Robots often have to simultaneously optimize multiple objectives, such as staying upright and following a target. In ML, multi-objective RL is often achieved by dividing an agent into multiple sub-agents each with its own objective, but finding the most efficient agent architecture that best solves the credit assignment problem is an active field of research[5]. Animals also face the task of simultaneously maximizing multiple objectives, such as place preference and vocal learning in songbirds. Finding the agent architectures used by the brain can inform the construction of efficient ML algorithms.

Approach and Outcomes: To investigate the nature of reinforcement signaling in the face of multiple objectives, I have proposed experiments in both mice and songbirds. In mice, I am testing if dopamine neurons act as global reinforcers for multiple objectives. In other words, is the same dopamine neuron activated when a mouse finds food when hungry and water when thirsty or are there multiple actor-critic pairs that have their own specialized reinforcement signals? In songbirds, recent studies from the Goldberg lab showed that songbirds have distinct navigation and singing systems that can be differentially reinforced[6]: strobe flashes negatively reinforce place preference but not song syllables, while noise bursts negatively reinforce song syllables but positively reinforce place preference. However, it remains unknown how this agent architecture is implemented in the brain, specifically in the dopamine system. By recording dopamine neuron activity in such paradigms, we can gain insights into how the brain solves the credit assignment problem during multi-objective learning.

3. Relative Contributions of Unsupervised and Reinforcement Learning

Motivation: Reinforcement learning and its variants have emerged as some of the most successful artificial intelligence algorithms in recent years. For example, Google Deepmind's AlphaGo[7], the first computer program to defeat a Go world champion, is largely based on reinforcement learning from games of self-play. However, the program had to be initially trained with a dataset of human expert games on which reinforcement could build. More recently, Alpha Go Zero[8] achieved a long-standing and much more challenging goal of AI by learning to play Go, *tabula rasa*, entirely relying on reinforcement learning with no human data or guidance. Biological learning systems that use reinforcement face a similar challenge. Is reinforcement learning a valid paradigm for motor skill acquisition very early in training? Predictions are needed before the brain can generate a reinforcement learning signal based on prediction error. However, the brain might first need to learn the mapping between motor output and sensory feedback before a prediction can even be made.

Approach and Outcomes: Prediction error signaling requires a prediction to exist in the first place. But very early in learning, the babbling bird may not yet be able to know the auditory consequences of its vocal variations. Therefore, it has been proposed that vocal learning proceeds by constructing a map between auditory signals back to motor commands that caused them (an inverse model[9]) instead of by reinforcement mechanisms. Based on my results that dopamine neurons encode song prediction error[10], I propose that song learning is implemented by a combination of early inverse models on which RL can act. The relative contributions or fine-tuning of these mechanisms are likely to influence the efficiency or rate of learning. Inverse or Hebbian learning models make specific predictions about the time-structure of learned syllables[9]. By monitoring error signaling as well as the developmental trajectory of syllables, we can discover the precise contributions of Hebbian learning and RL that combine to make rapid and efficient song learning possible.

Summary

Song learning provides several opportunities to investigate the algorithms used by the brain to implement rapid and efficient motor learning. Insights gained from these studies have the potential to inform or fine-tune machine learning algorithms. I have presented three examples, (1) an efficient 2-step reinforcement learning algorithm, (2) the agent architecture for multi-objective reinforcement learning, and (3) the relative contributions of unsupervised and reinforcement learning for rapid and efficient learning.

---

[1] Sutton, R.S., and Barto, A.G. (1998). Reinforcement learning: an introduction (Cambridge, MA: MIT Press).
[2] Kirkpatrick S, Gelatt CD Jr., Vecchi MP (1983) *Science* **220**:671–680.
[3] Fee and Goldberg *Neuroscience* **198**, 152-170 (2011)
[4] Duffy and Gadagkar *et al. In Prep*
[5] Vamplew et al. *Machine Learning* **84**, 51-80 (2011)
[6] Murdoch et al. *Scientific Reports*, **8**, 6766 (2018)
[7] Silver *et al. Nature* **529**, 484-489 (2016)
[8] Silver *et al. Nature* **550**, 354-359 (2017)
[9] Hahnloser and Ganguli *Vocal learning with inverse models*. In Principles of Neural Coding. CRC Press: CRC Press, 547-564 (2013)
[10] Gadagkar et al. *Science* **354**, 1278-1282 (2016)