CNeuro2024 Lecture Abstracts

Nathaniel Daw

Abstract 1 – Basic Lecture:

Introduction to Soft Reinforcement Learning: Basic Algorithms, Brain and Behavior

In this lecture, we discuss the core formalism and algorithms for reinforcement learning (building on Bote's introduction). We focus in particular on the distinction between model-free and model-based methods. We discuss how these two types of approaches have been invoked in AI, systems neuroscience (linking to Uchida's lectures on dopamine), behavioral economics (linking to Zhu's lectures on games), behavioral psychology, and computational psychiatry.

Abstract 2 – Advanced Lecture:

Advanced Reinforcement Learning: Replay, Temporal Abstraction, and Function Approximation

In this lecture, we focus particularly on theoretical and empirical results about model-based evaluation in the brain. We first consider replay and the rational prioritization of replayed sequences as a candidate neural implementation of model-based credit assignment and planning. Second, we consider temporal abstraction, successor representations, and linear RL (related to Bote's soft RL) as a method for simplifying model search.