Prof. Kemp’s notes on Zackory’s presentation

- Common RL problems
  - Reward -> R = a_1 pos_error + a_2 success
    - Weights can be challenging to set
      - [0.5, 0.5] [0, 1], [0, 10], [0, 0.05]
      - Normalization if you know the min and max
- Computationally relevant terminology
  - Timestep: 1 simulation step = 0.1 sec of simulated time
  - Rollout: 200 timesteps = 20 sec
  - Update: PPO specific
    - 8 CPUs -> 8 Rollouts
  - Zackory recommendation: Do not change PPO’s parameters!
  - 10,000,000 timesteps used for training each policy
    - ~6 hours single machine on Amazon (itch scratching)
    - ~10 hours single machine on Amazon (bed bathing)
    - Drinking takes a long time due to large number of spheres representing water