Good Afternoon Colleagues
Good Afternoon Colleagues

- Are there any questions?
Logistics
Logistics

- Registering for the course
Logistics

• Registering for the course

• Nice responses!
Logistics

- Registering for the course
- Nice responses!
  - Length and content good
Logistics

- Registering for the course
- Nice responses!
  - Length and content good
  - Be clear and specific
Logistics

• Registering for the course

• Nice responses!
  – Length and content good
  – Be clear and specific
  – Look for programming assignment opportunities
Logistics

• Registering for the course

• Nice responses!
  – Length and content good
  – Be clear and specific
  – Look for programming assignment opportunities
  – I have author’s responses to exercises
Logistics

- Registering for the course

- Nice responses!
  - Length and content good
  - Be clear and specific
  - Look for programming assignment opportunities
  - I have author’s responses to exercises

- Programming language
Logistics

- Registering for the course

- Nice responses!
  - Length and content good
  - Be clear and specific
  - Look for programming assignment opportunities
  - I have author’s responses to exercises

- Programming language

- Today: self-introductions, discussion leader assignments
Reduced Formalism

**Knowns:**
- $S = \{\text{Blue, Red, Green, Black, \ldots}\}$
- Rewards in $\mathbb{R}$
- $A = \{\text{Wave, Clap, Stand}\}$

$S_0, a_0, r_0, s_1, a_1, r_1, s_2, \ldots$

**Unknowns:**
- $\mathcal{R} : S \times A \rightarrow \mathbb{R}$
- $\mathcal{T} : S \times A \rightarrow S$

\[
\begin{align*}
    r_i &= \mathcal{R}(s_i, a_i) \\
    s_{i+1} &= \mathcal{T}(s_i, a_i)
\end{align*}
\]
This course

- Agent’s perspective: only **policy** under control
  - State representation, reward function given
  - Focus on policy algorithms, theoretical analyses
This course

- Agent’s perspective: only policy under control
  - State representation, reward function given
  - Focus on policy algorithms, theoretical analyses
  - Appeal: program by just specifying goals
This course

- Agent’s perspective: only policy under control
  - State representation, reward function given
  - Focus on policy algorithms, theoretical analyses
  - Appeal: program by just specifying goals
  - Practice: need to pick the representation, reward
This course

- Agent’s perspective: only **policy** under control
  - State representation, reward function given
  - Focus on policy algorithms, theoretical analyses
  - Appeal: program by just specifying goals
  - Practice: need to pick the representation, reward
  - videos
This course

- Agent’s perspective: only **policy** under control
  - State representation, reward function given
  - Focus on policy algorithms, theoretical analyses
  - Appeal: program by just specifying goals
  - Practice: need to pick the representation, reward
  - videos

- Methodical approach
  - Solid foundation rather than comprehensive coverage
This course

- Agent’s perspective: only **policy** under control
  - State representation, reward function given
  - Focus on policy algorithms, theoretical analyses
  - Appeal: program by just specifying goals
  - Practice: need to pick the representation, reward
  - videos

- Methodical approach
  - Solid foundation rather than comprehensive coverage
  - RL reading group
Some Questions

- What’s a model?
Some Questions

- What’s a model?
- Does speed of learning matter?
Some Questions

• What’s a model?

• Does speed of learning matter?

• Distinguishing features (from supervised learning)?
Some Questions

- What’s a model?
- Does speed of learning matter?
- Distinguishing features (from supervised learning)?
  - trial-error search, delayed reward
  - exploration vs. exploitation (chapt. 2)
Some Questions

• What’s a model?

• Does speed of learning matter?

• Distinguishing features (from supervised learning)?
  – trial-error search, delayed reward
  – exploration vs. exploitation (chapt. 2)

• Learn just the policy, or also state representation?

• What about the reward function?
Some Questions

- Reward function vs. value function
Some Questions

- Reward function vs. value function
  - Tic-tac-toe example
Some Questions

• Reward function vs. value function
  – Tic-tac-toe example
  – Phil making breakfast example
Some Questions

- Reward function vs. value function
  - Tic-tac-toe example
  - Phil making breakfast example

- Distinction with evolutionary methods?
  - Tic-tac-toe example
Some Questions

- Reward function vs. value function
  - Tic-tac-toe example
  - Phil making breakfast example

- Distinction with evolutionary methods?
  - Tic-tac-toe example
  - Phil making breakfast example
Some Questions

- Reward function vs. value function
  - Tic-tac-toe example
  - Phil making breakfast example

- Distinction with evolutionary methods?
  - Tic-tac-toe example
  - Phil making breakfast example

- Is evolutionary learning ever better?
Some Questions

• Reward function vs. value function
  – Tic-tac-toe example
  – Phil making breakfast example

• Distinction with evolutionary methods?
  – Tic-tac-toe example
  – Phil making breakfast example

• Is evolutionary learning ever better?

• Tic-tac-toe example: what are the converged values?
Some Questions

- Reward function vs. value function
  - Tic-tac-toe example
  - Phil making breakfast example

- Distinction with evolutionary methods?
  - Tic-tac-toe example
  - Phil making breakfast example

- Is evolutionary learning ever better?

- Tic-tac-toe example: what are the converged values?
  - on-policy, vs. off-policy updates