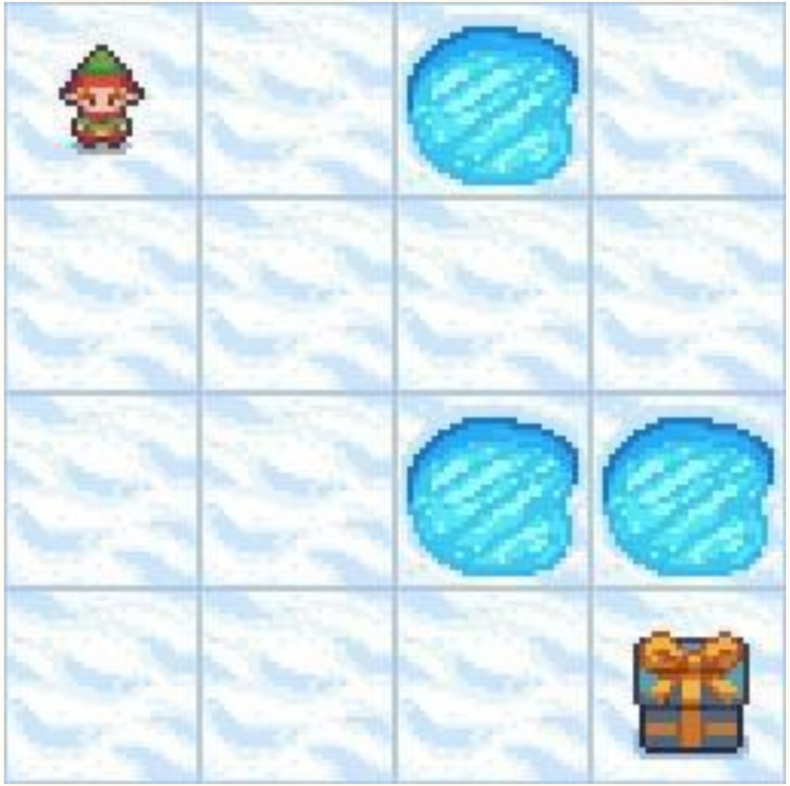
Evaluating Multiple Reinforcement Learning Agents in the Frozen Lake Environment: Scalability and Stochasticity

CS 4/5756 Robot Learning
Group 3: Sophia Pham (tpp38) & Julie Jeong (sj598)

Problem & Environment



Frozen Lake - Gymnasium

1. **Action Space:** Discrete (4)
2. **Observation Space:**
   - 4×4: Discrete (16)
   - 5×5: Discrete (25)
3. **Rewards:**
   - +1 Reach goal
   - +0 Reach hole/frozen
4. **Slippery:** If true the player will move in intended direction with probability of 1/3

Research Hypothesis

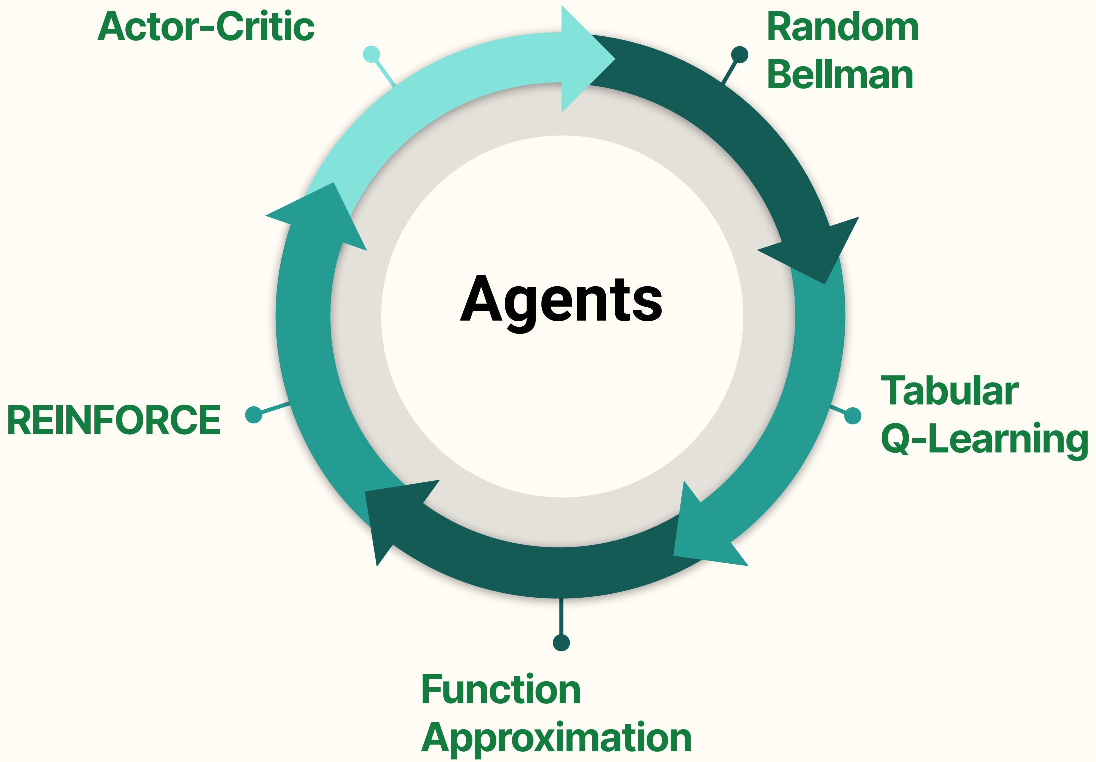
Hypothesis 1: Scalability

Function approximation methods and policy gradient algorithms (REINFORCE, Actor-Critic) outperform tabular Q-learning as grid size increases

Hypothesis 2: Stochasticity

Policy gradient methods are more robust to stochasticity (is_slippery=True) compared to tabular and function approximation methods

Approach



Experiments:

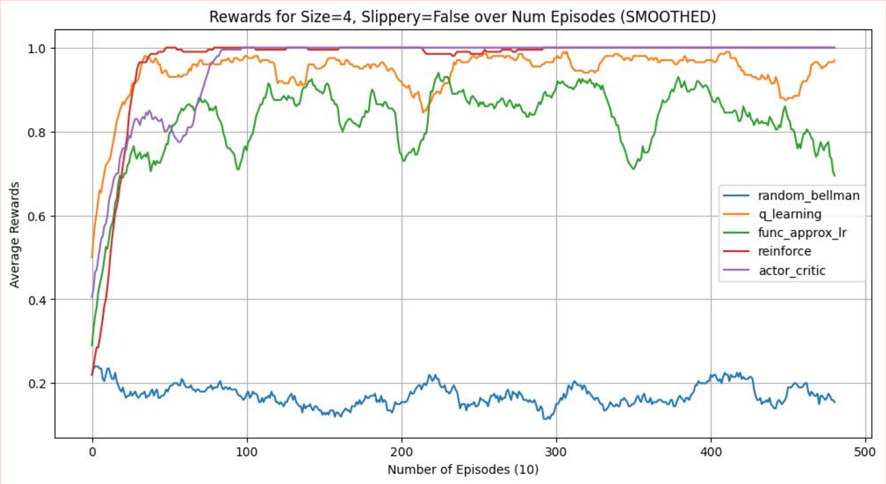
1. 4×4 and 5×5 grids
2. Deterministic and Stochastic conditions

→ Agents trained over **5000** episodes

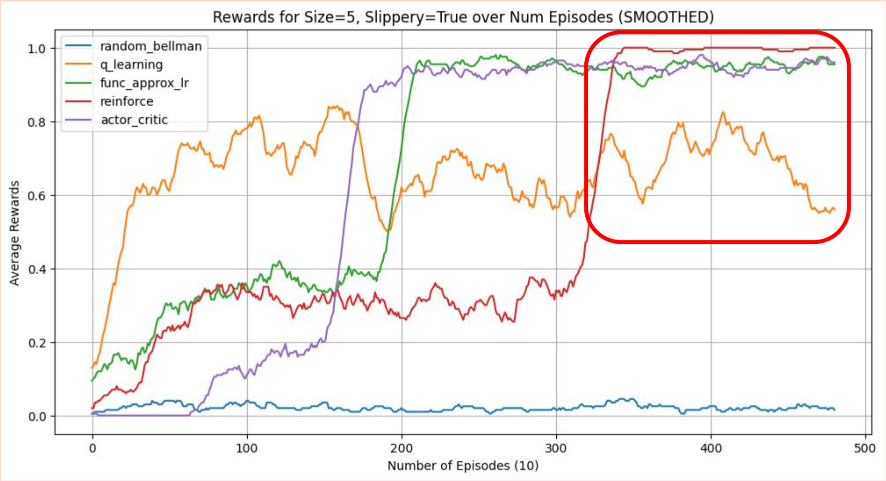
Key Takeaway 1

Policy gradient and function approximation methods are effective for scaling to larger environments

Training Rewards for Size = 4



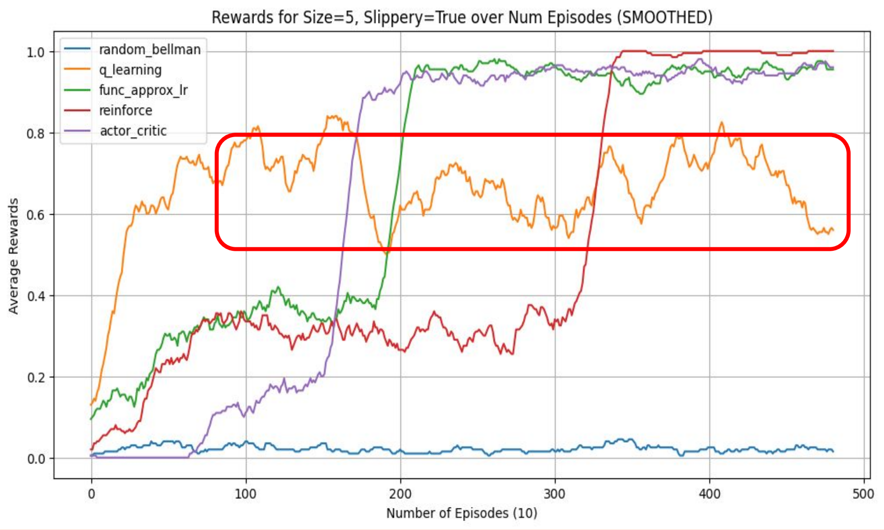
Training Rewards for Size = 5



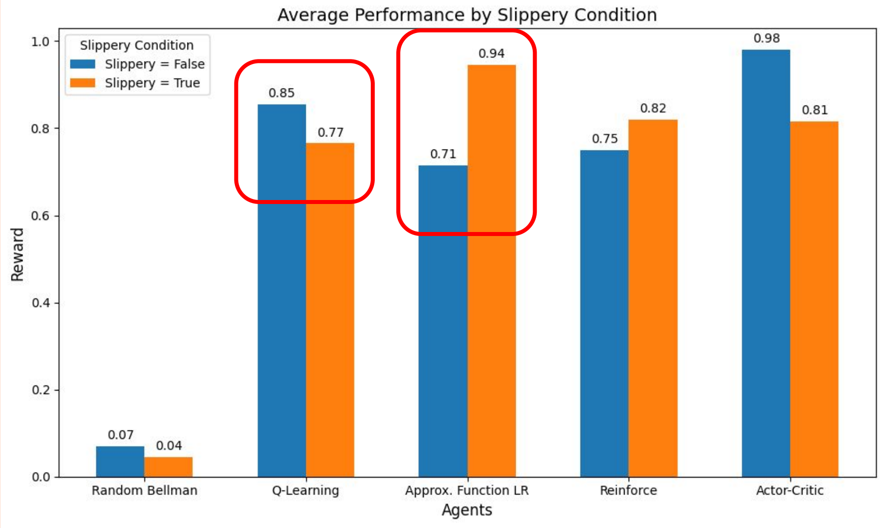
Key Takeaway 2

Stochastic environments challenge tabular methods, but function approximation shows surprising strength

Training Rewards for Slippery = True



Avg. Evaluation Reward By Slippery



Key Takeaway 3

Actor-Critic stands out as the most consistently reliable approach across all settings

Evaluation Rewards Averaged over 100 Iterations

	Size	Slippery	Random Bellman	Q-Learning	Approx. Function LR	Reinforce	Actor-Critic
Setting 1	4	false	0.14	1.0	0.89	1.0	**0.99**
Setting 2	4	true	0.08	1.0	0.92	0.64	**0.92**
Setting 3	5	false	0.0	0.71	0.54	0.5	**0.97**
Setting 4	5	true	0.01	0.53	0.97	1.0	**0.71**

Conclusion

Hypothesis 1:

As the grid size increases (from 4×4 to 5×5), function approximation methods and policy gradient algorithms (REINFORCE, Actor-Critic) will perform better than tabular Q-learning due to their ability to handle larger state-action spaces

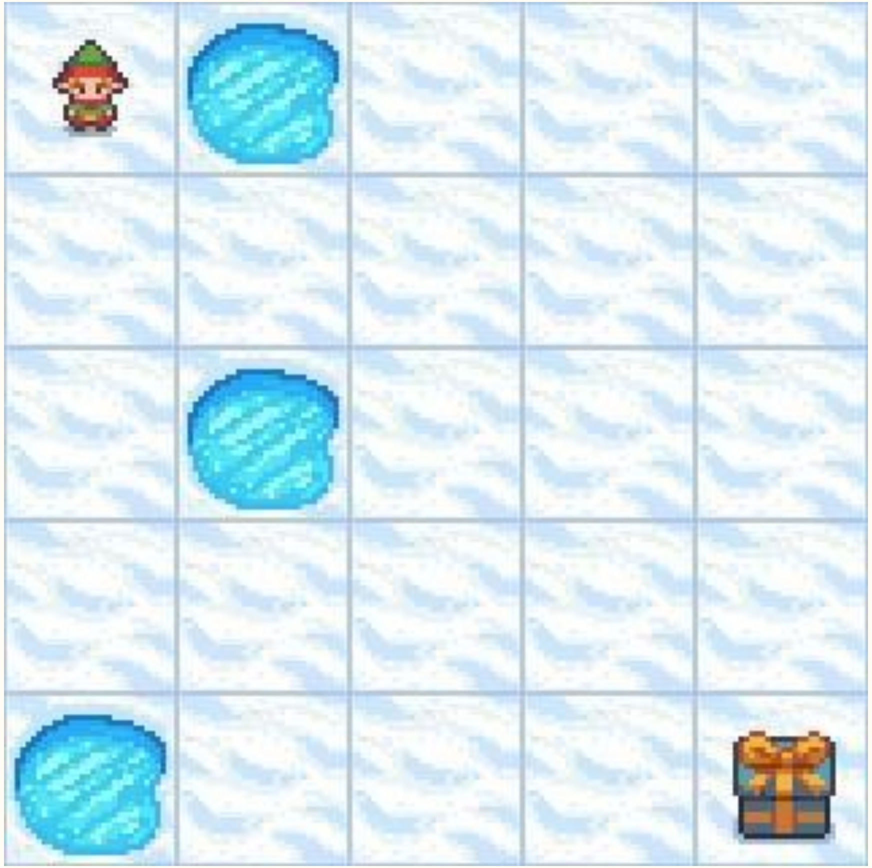
→ Proven to be ***TRUE***

Hypothesis 2:

When the environment is stochastic (is_slippery=True), policy gradient methods will show greater adaptability compared to tabular Q-learning and function approximation

→ Proven to be ***PARTIALLY TRUE***

5×5 Grid & Slippery
Actor-Critic



Thank you!