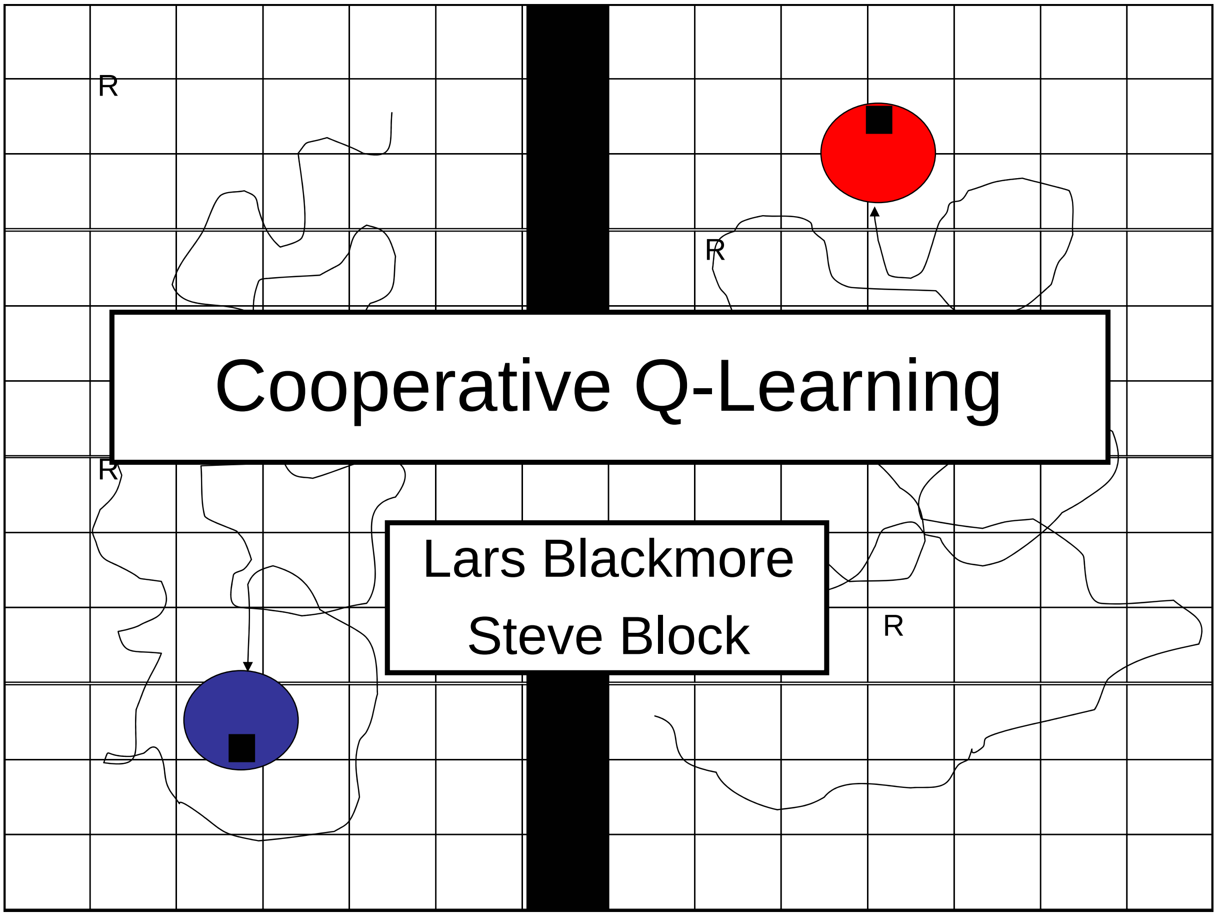
Cooperative Q-Learning

Lars Blackmore

Steve Block

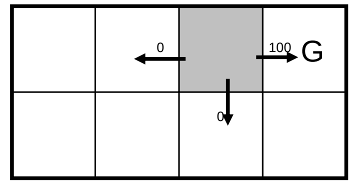


Overview

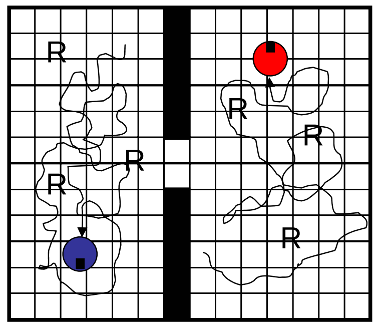
- Co-operative Q-Learning Reminder
- Repeating the results from Ahmadabadi paper
- Qualifying and extending Ahmadabadi's conclusions
- Dynamic environments
- Discounted expertness
- Conclusion

Q-Learning Reminder

- Framework: Markov Decision Processes
  - States S, Actions A
  - Rewards R(s,a)
  - Transition function T(s,a,s')

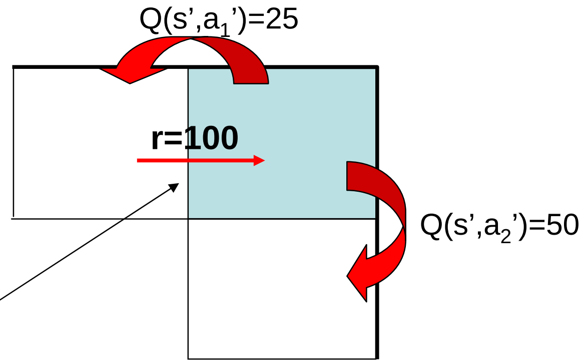


- Mobile Robots in a grid world



Q-Learning Reminder

- Q(s,a) values approximate the optimal lifetime reward
- Policy derived from Q values
- Q-Learning algorithm iteratively updates the table of Q values



$$Q(s,a) = 100 + 50\gamma$$
$$= 145$$

Co-operative Q-learning

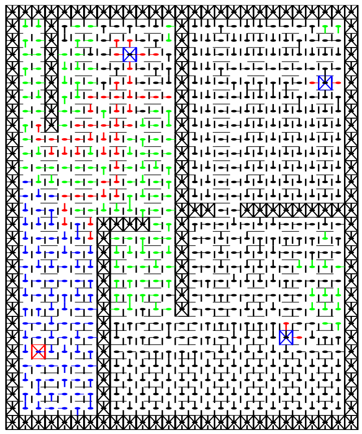
- How should independent agents share their Q-tables?
- Ahmadabadi: ‘Learn from Experts’
- Expertness value for each agent ⇨ e_i
  - Based on absolute value of reinforcement received $e_i^{ABS} = \sum_t |r_i(t)|$
- Specialised agents: expertness value for each zone
  - Global, Local, State

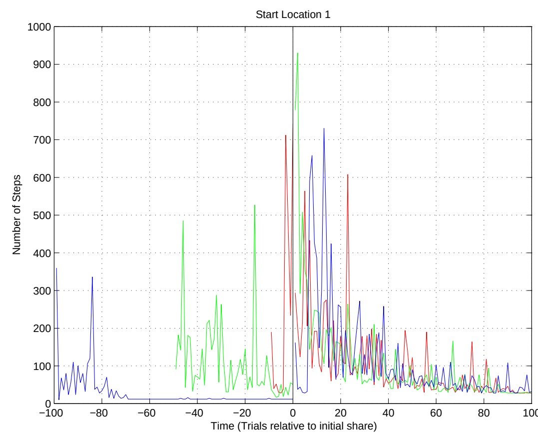
Our results

- We wanted to test the claim:
  - *‘Co-operative Q-Learning is better than individual Q-learning’*
- First in a segmented world (Ahmadabadi)
- Then in the general case

Setup, presentation of results

- Presentation of results:



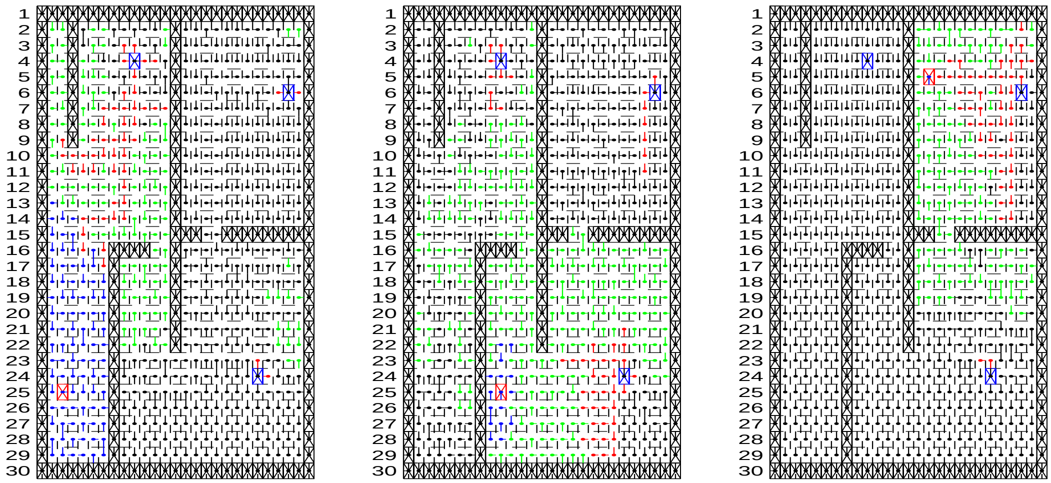


Our results

- We wanted to test the claim:
  - *'Co-operative Q-Learning is better than individual Q-learning'*
- **First in a segmented world (Ahmadabadi)**
- Then in the general case

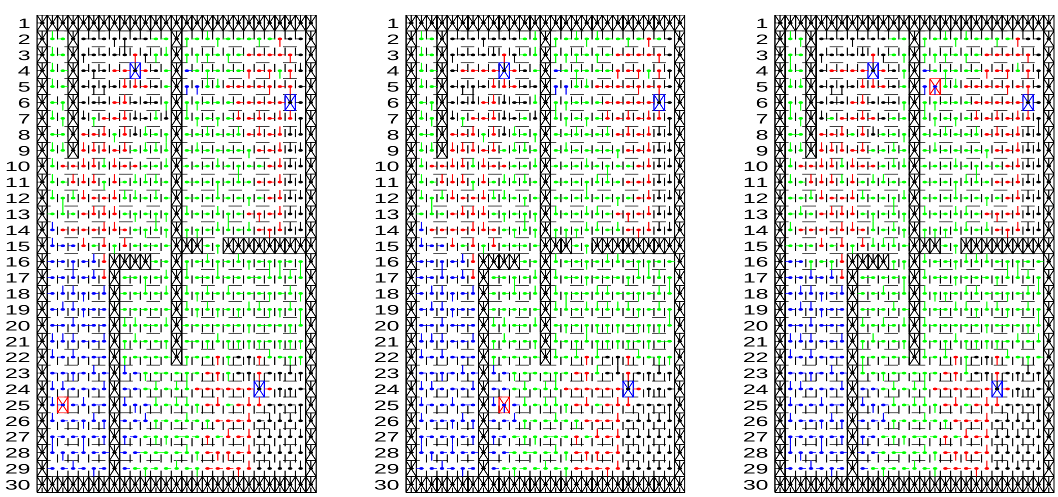
Segmented World

- Q Fields after individual learning



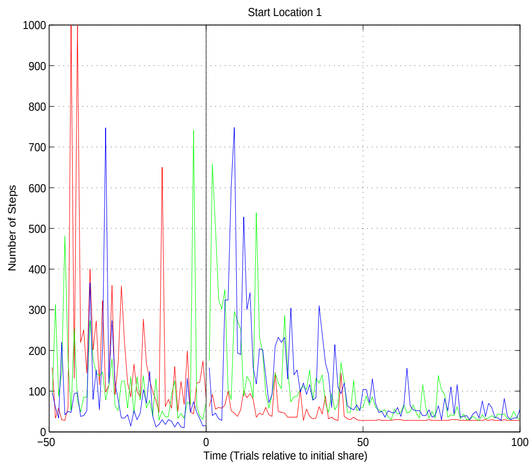
Segmented World

- Q Field after sharing

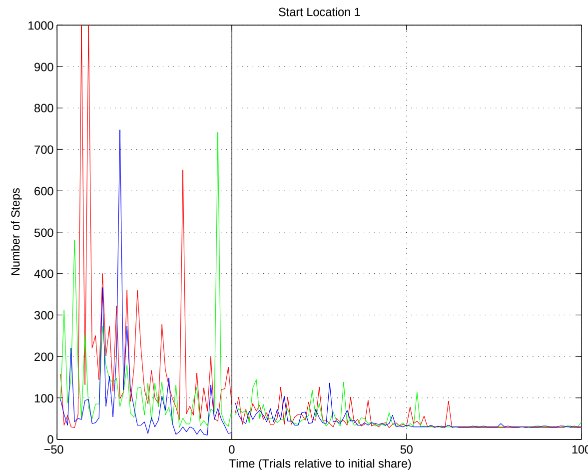


Segmented World

- Performance results



No Cooperation



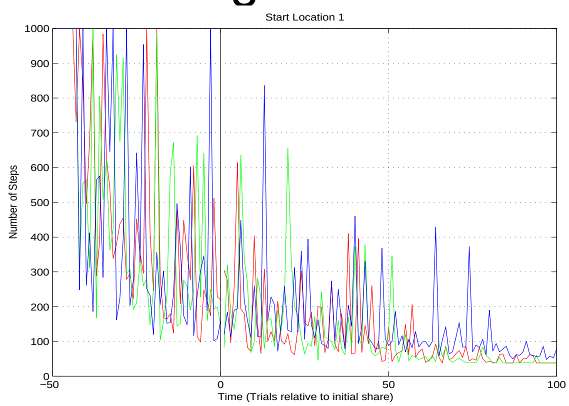
With Cooperation

Our results

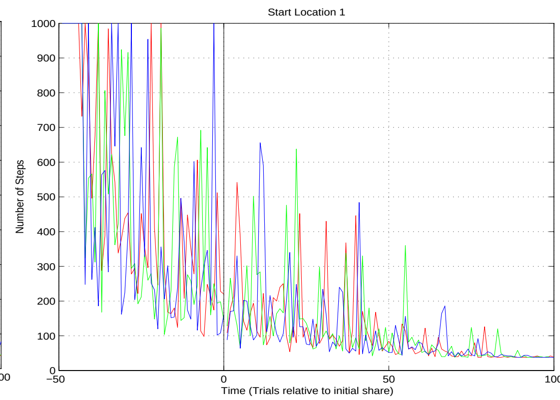
- We wanted to test the claim:
  - *‘Co-operative Q-Learning is better than individual Q-learning’*
- First in a segmented world (Ahmadabadi)☑
- **Then in the general case**
  - **Equal experience levels**
  - **Different experience levels**

General Maze World

- Equal experience – single start location
- Sharing DOES NOT improve performance



No Q sharing

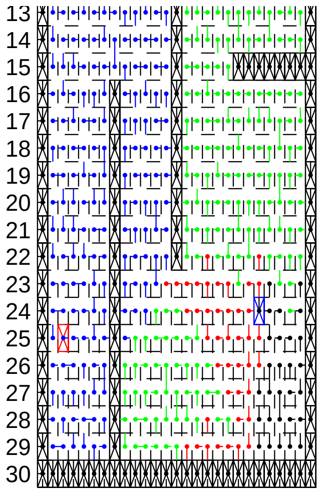


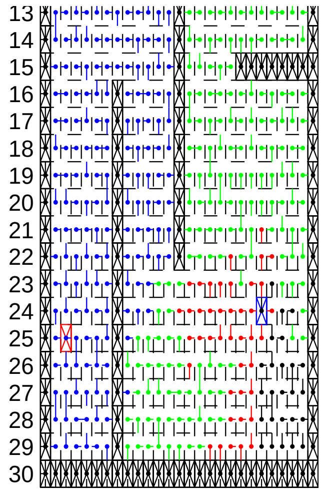
With Q sharing

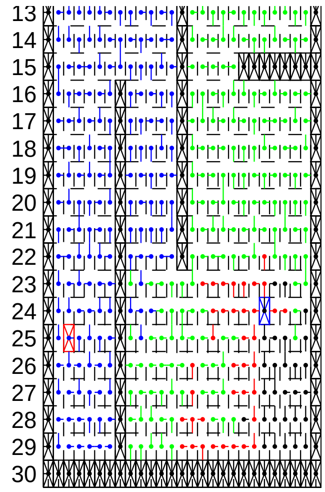
- Why?

Equal Experience

- Before share

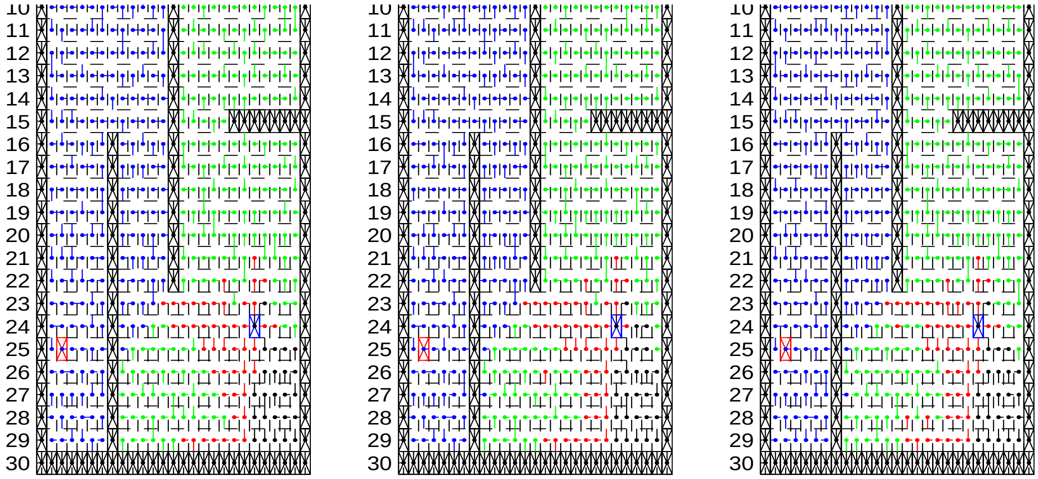






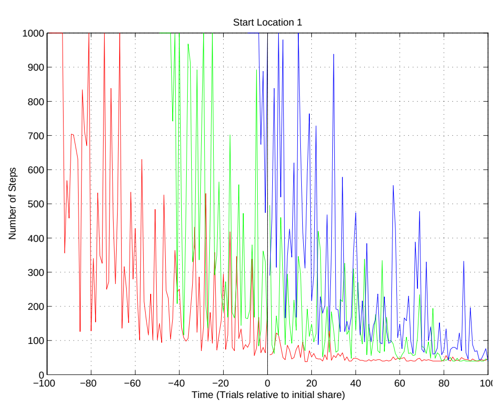
Equal Experience

- After share

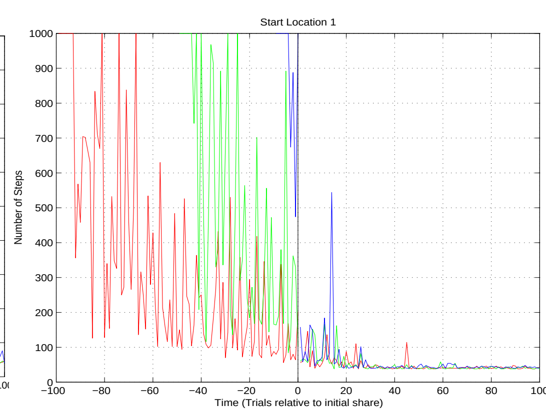


Different Experience

- Performance of less experienced agents improved



No Q sharing



With Q sharing

Our results

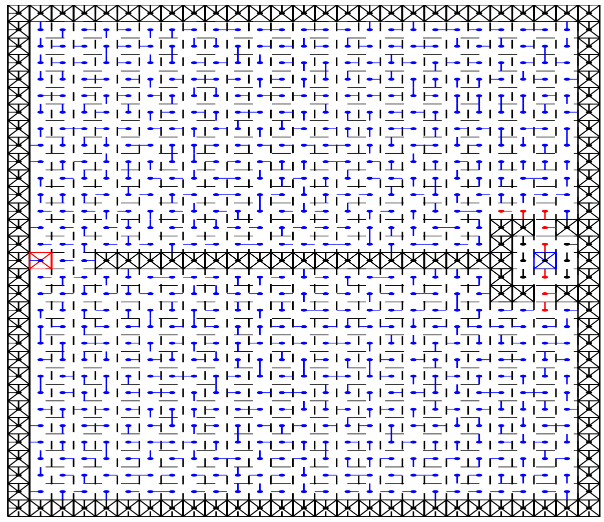
- We wanted to test the claim:
  - *'Co-operative Q-Learning is better than individual Q-learning'*
- First in a segmented world (Ahmadabadi)☑
- **Then in the general case**
  - **Same expertness** ☒
  - **Different expertness** ☑ but…

Dynamic Environments

- How does a dynamic environment affect the performance of co-operative Q-learning?
- 'General' case:
  - Simple maze, single goal
  - Obstacles disappear and reappear at random with a given probability
  - Very little change in performance
- Why?
  - Q-learning in general very good at local repair

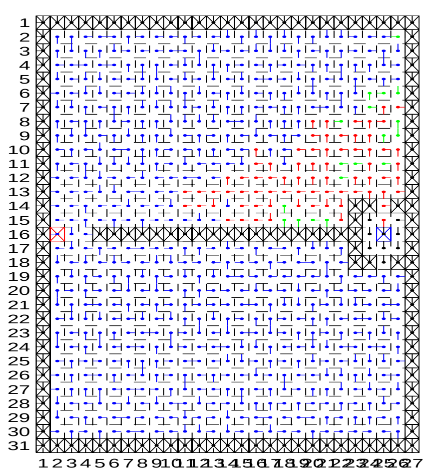
Contrived Case

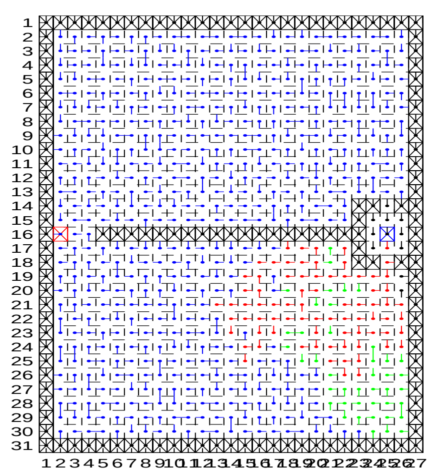
- Doorways alternate between agents



Contrived Case

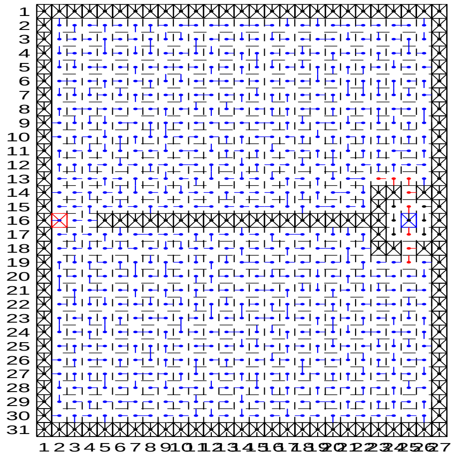
- Now co-operative Q learning fails:
  - After individual learning

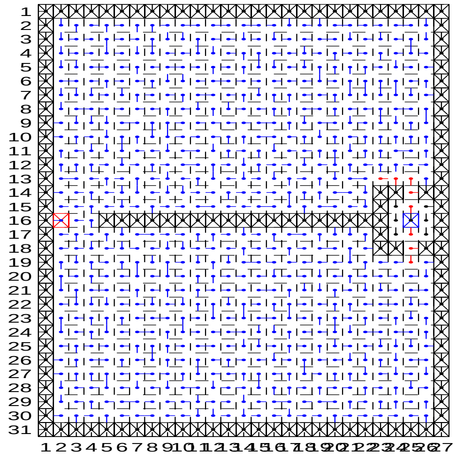




Contrived Case

- Now co-operative Q learning fails:
  - After sharing Q values

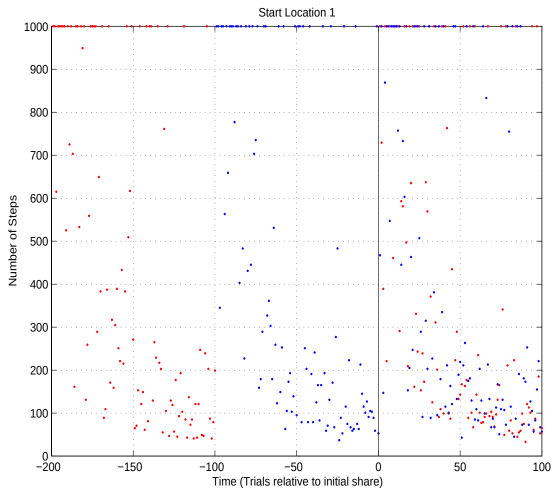


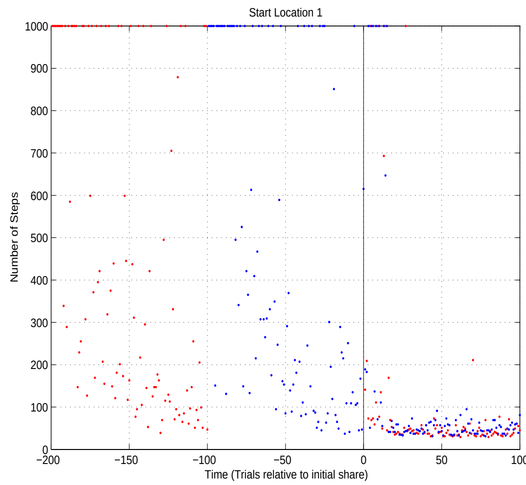


Discounted Expertness

- Idea: more recent experiences are most valuable
- *Discounted Expertness*: discount expertness as time progresses
- Expertness due to experiences in the past becomes less and less valuable

Expertness Discounting





Without Expertness Discounting

With Expertness Discounting

Conclusion

- Co-operative Q-learning is only useful for:
  - Highly segmented worlds
  - Very different experience levels
- Simpler approaches could be just as effective
- Q learning in general very effective in dynamic worlds
- Co-operation can be made to fail in artificially contrived dynamic worlds
  - Expertness discounting is useful in these cases
- Applicability of expertness to maze worlds?

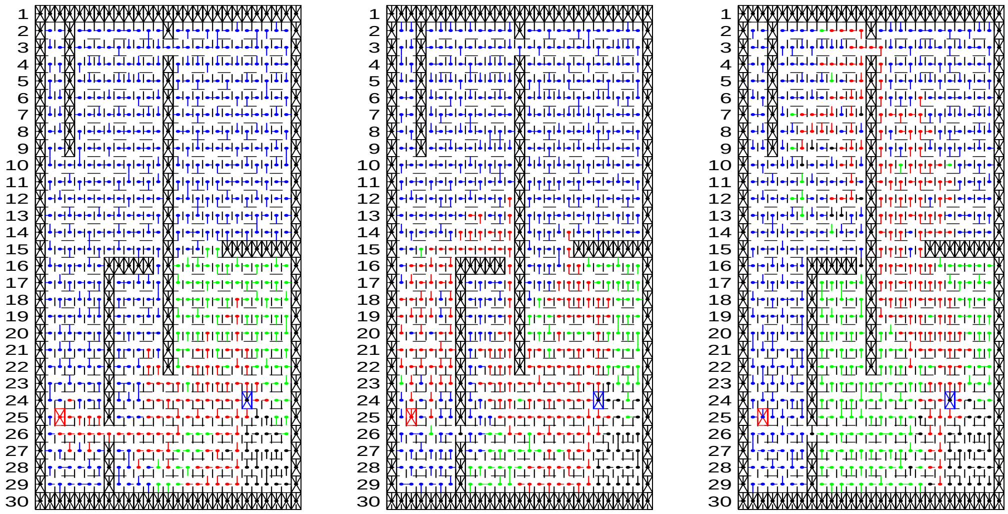
Appendix

Dynamic Environments

- ‘Doors’ scenario
  - 3 Agents, 1 Goal
  - Agents need to pass through doorways to get to goal
  - Each agents learns with a different doorway open
  - Agents 1 and 2 have found optimal paths which no longer exist
  - What happens when the Q-tables are shared?

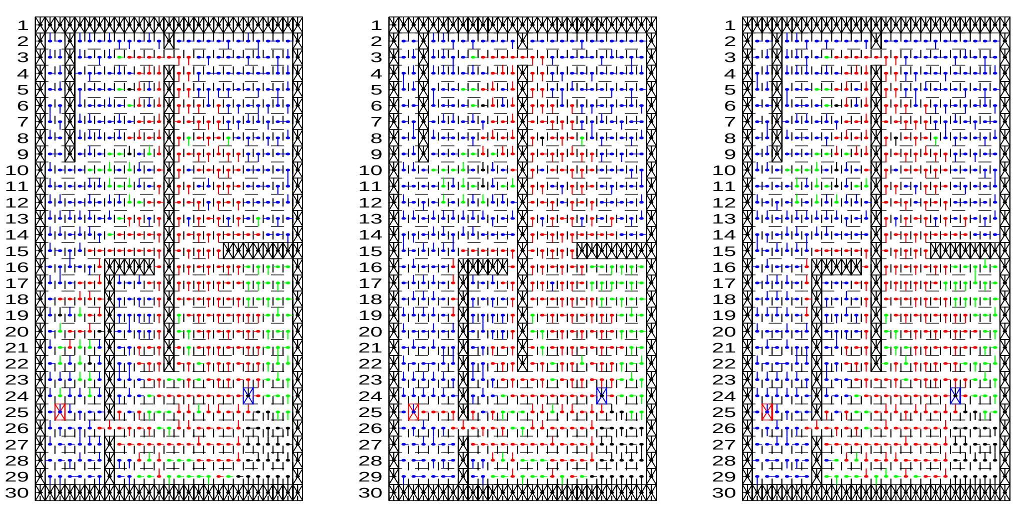
Doors Scenario

- At end of individual trials



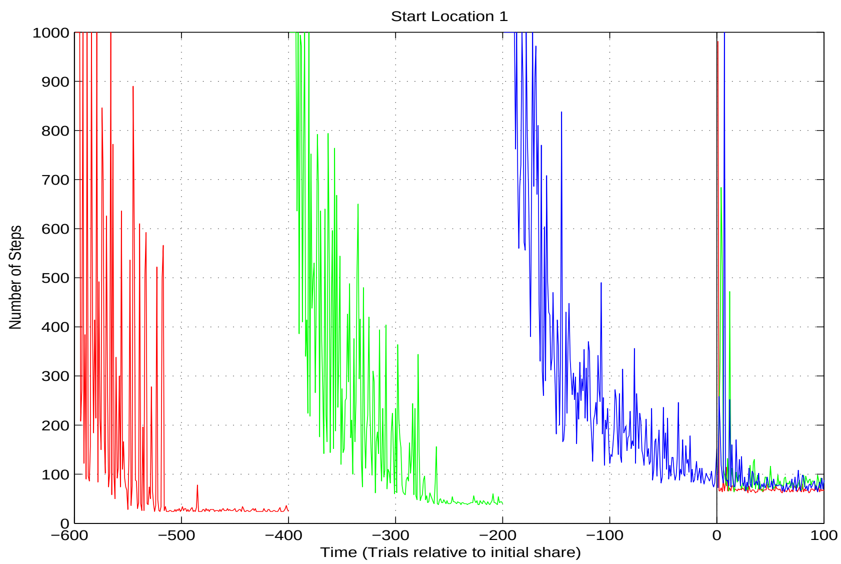
Doors Scenario

- After sharing Q-table



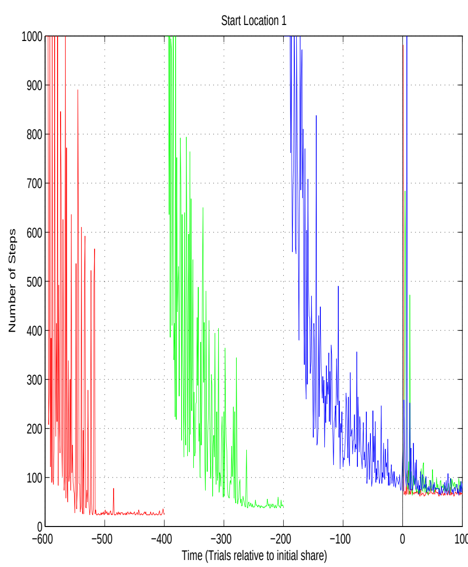
Dynamic Environments

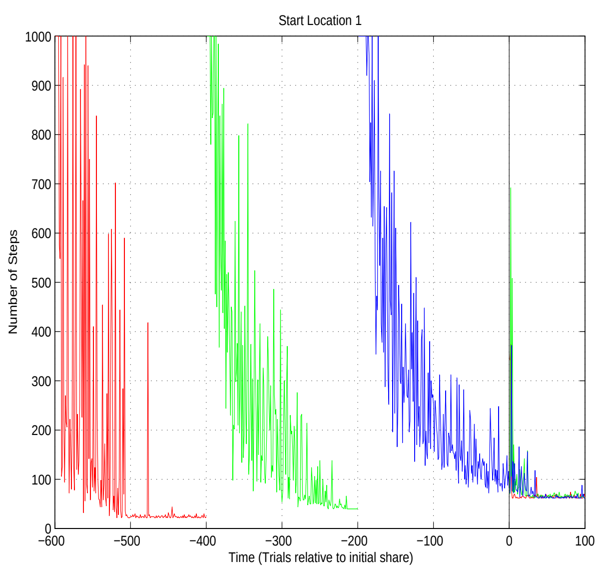
- Co-operative Q-learning works ok!



Dynamic Environments

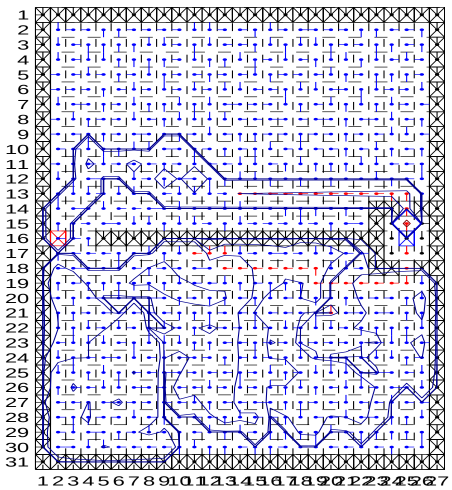
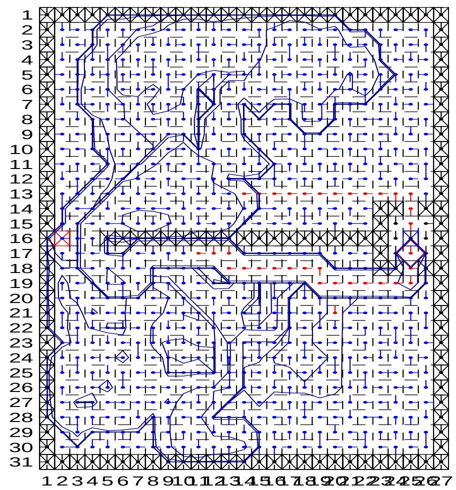
- Discounting helps a tiny bit





Expertness Discounting

- This solves the problem:
  - After sharing with expertness discounting

Segmented World

- Co-operation ***does*** significantly increase performance for state specialisation
  - Is this obvious?
- Why not just take the **most expert** agent's Q values?
  - Test showed this is just as good
- But now Q tables homogeneous
  - So continue with Ahmadabadi algorithm…