

Could Active Perception Aid Navigation of Partially Observable Grid Worlds?

Paul A. Crook & Gillian Hayes

paulc@dai.ed.ac.uk

Institute of Perception, Action and Behaviour

School of Informatics

University of Edinburgh



Breakdown Of Talk



- Introduction & Background
 - Outline our area interest.
 - Partially observable worlds.
 - Active perception.
 - The questions this work addresses.
- Summary of Experiments & Results
- Conclusions

Area of Interest



Interested in agents that are:

- reactive;
- memoryless;
- model-free;
- embodied, embedded and situated in their environment.

Simple Reactive Agents

- **Reactive;** deterministic mapping between observations and actions.



Simple Reactive Agents

- **Reactive**; deterministic mapping between observations and actions.
- **Memoryless**; no memory of previous actions selected or observations previously obtained.



Simple Reactive Agents



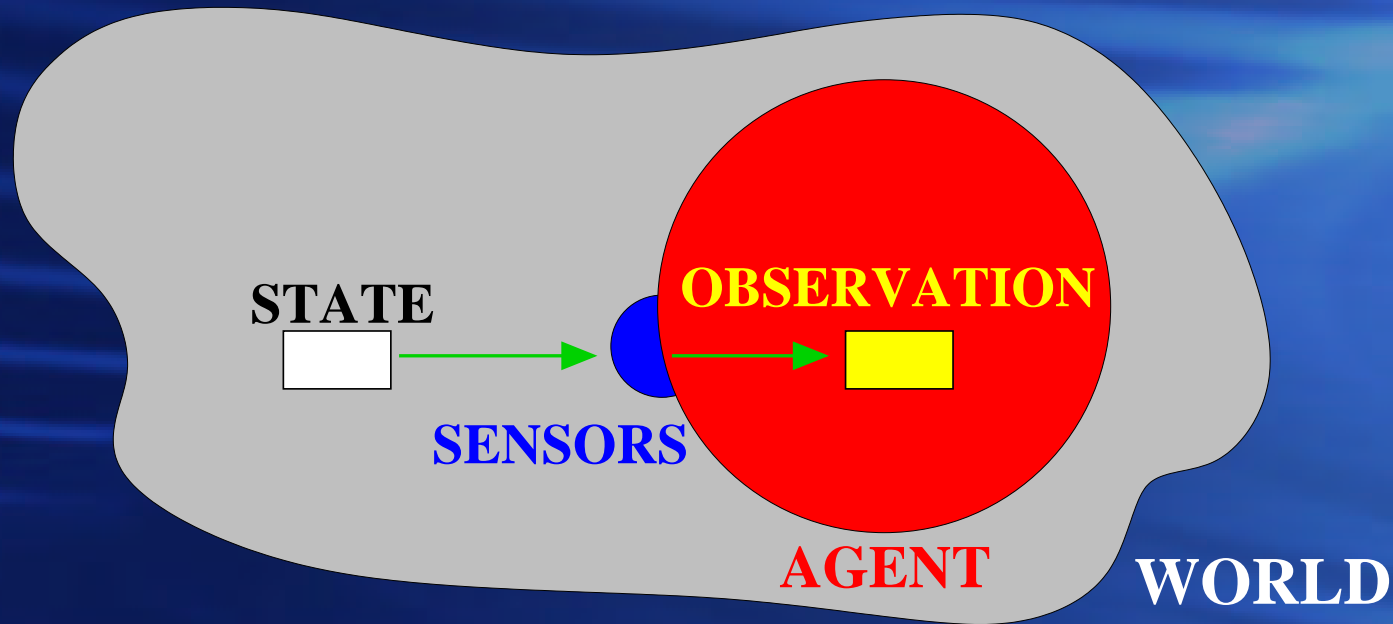
- **Reactive;** deterministic mapping between observations and actions.
- **Memoryless;** no memory of previous actions selected or observations previously obtained.
- **Model-free;** does not learn a model of the environment.

Simple Reactive Agents



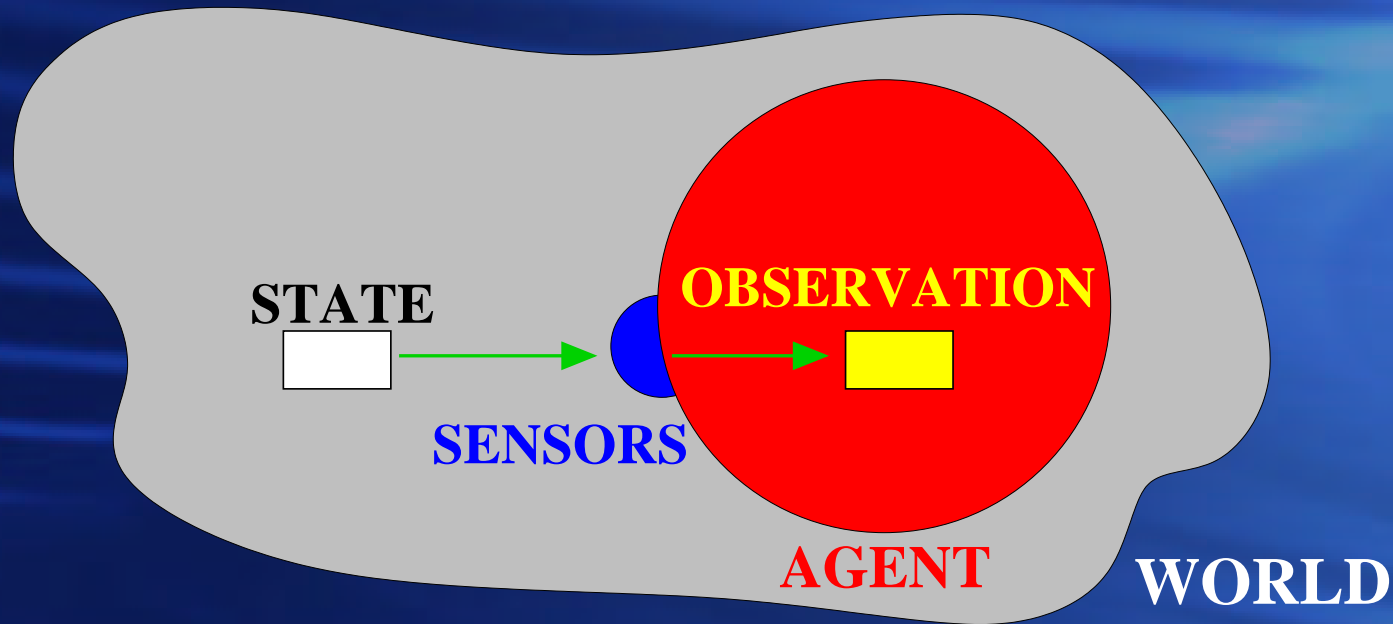
- **Reactive;** deterministic mapping between observations and actions.
- **Memoryless;** no memory of previous actions selected or observations previously obtained.
- **Model-free;** does not learn a model of the environment.
- **Embodied, Embedded & Situated;** physically located in and limited by its environment.

Partially Observable Worlds



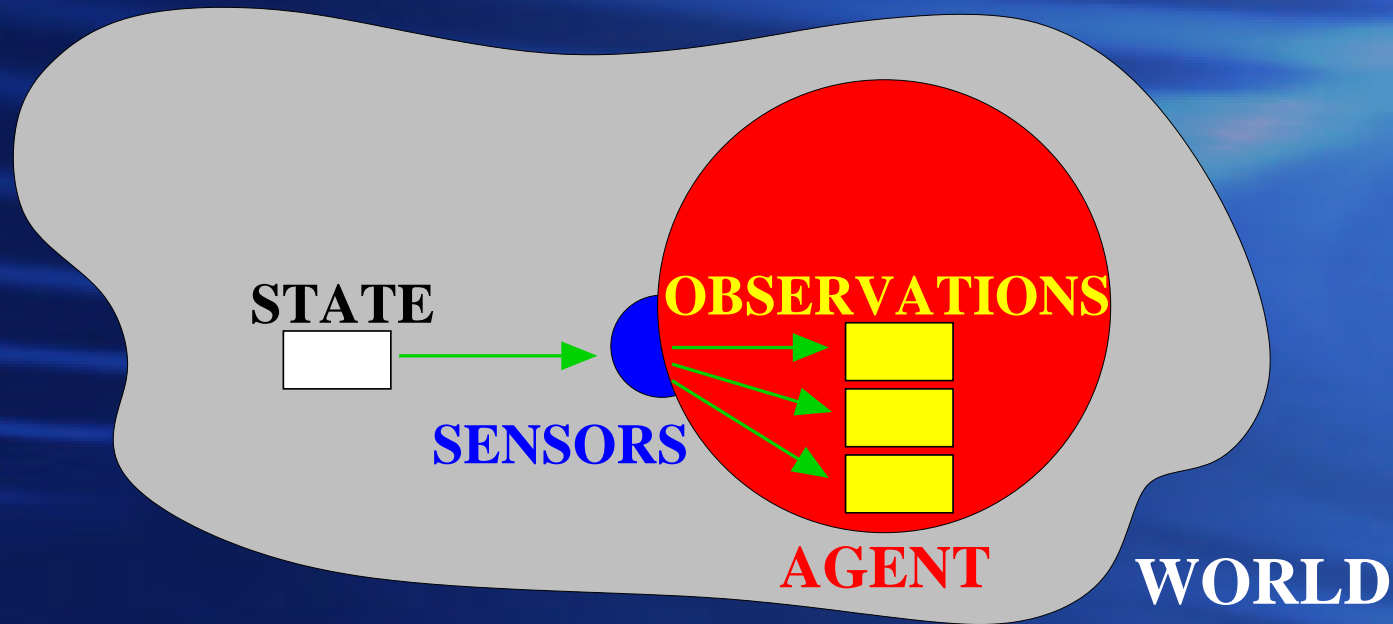
- With such an agent, its reactions are cued by its observation of the world.

Partially Observable Worlds



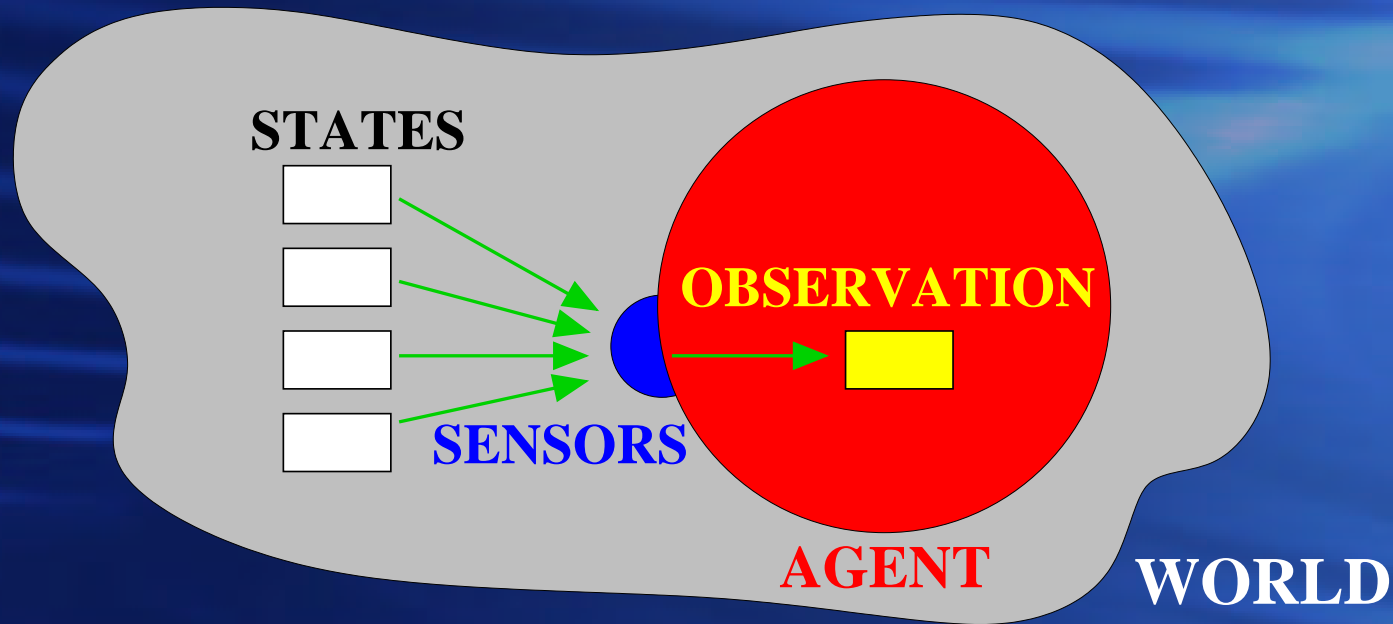
- With such an agent, its reactions are cued by its observation of the world.
- Its observation will however be limited by its physical location and by its sensors.

Partially Observable Worlds



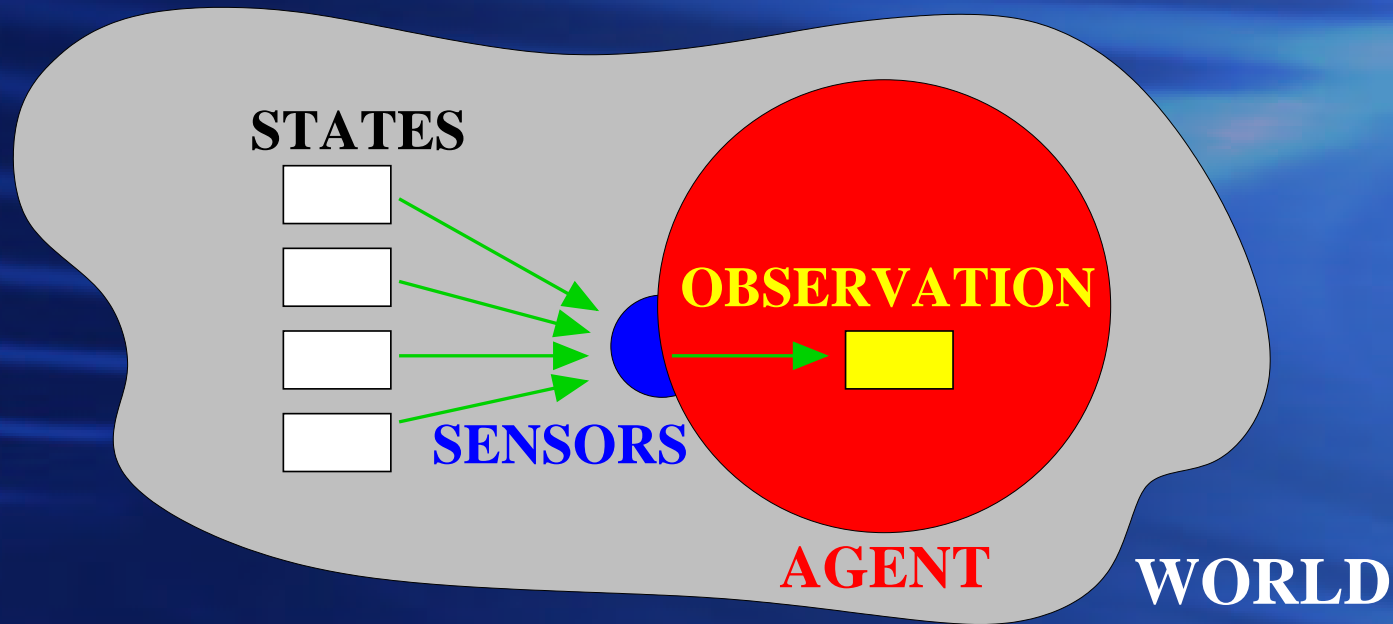
- With such an agent, its reactions are cued by its observation of the world.
- Its observation will however be limited by its physical location and by its sensors.

Partially Observable Worlds



- With such an agent, its reactions are cued by its observation of the world.
- Its observation will however be limited by its physical location and by its sensors.

Partially Observable Worlds



To such an agent the world will be partially observable, *i.e.* the observations it makes may not uniquely identify the world's state.

Problems...

- The best solution to a partially observable task when using a deterministic mapping from observations to actions, is arbitrarily worse than the optimal solution. (Singh et.al, ICML'94)



Problems...

- The best solution to a partially observable task when using a deterministic mapping from observations to actions, is arbitrarily worse than the optimal solution. (Singh et.al, ICML'94)
- 1-step back reinforcement learning algorithms, such as SARSA, suffer from both *local* and *global* impairment. (Whitehead, Phd Thesis)



Problems...



- The best solution to a partially observable task when using a deterministic mapping from observations to actions, is arbitrarily worse than the optimal solution. (Singh et.al, ICML'94)
- 1-step back reinforcement learning algorithms, such as SARSA, suffer from both *local* and *global* impairment. (Whitehead, Phd Thesis)

$$Q(s, a) \leftarrow Q(s, a) + \alpha[r + \gamma Q(s', a') - Q(s, a)]$$

However...

- Loch and Singh (ICML'98) showed empirically that using eligibility traces Reinforcement Learning algorithms can learn the best solution possible using a deterministic mapping, *i.e.* it appears to deal with global impairment.



So...

- If we can provide a deterministic agent with a way to differentiate between states that it would otherwise observe to be the same, it should be able to vary its actions in those states and thus possibly learn better solutions.



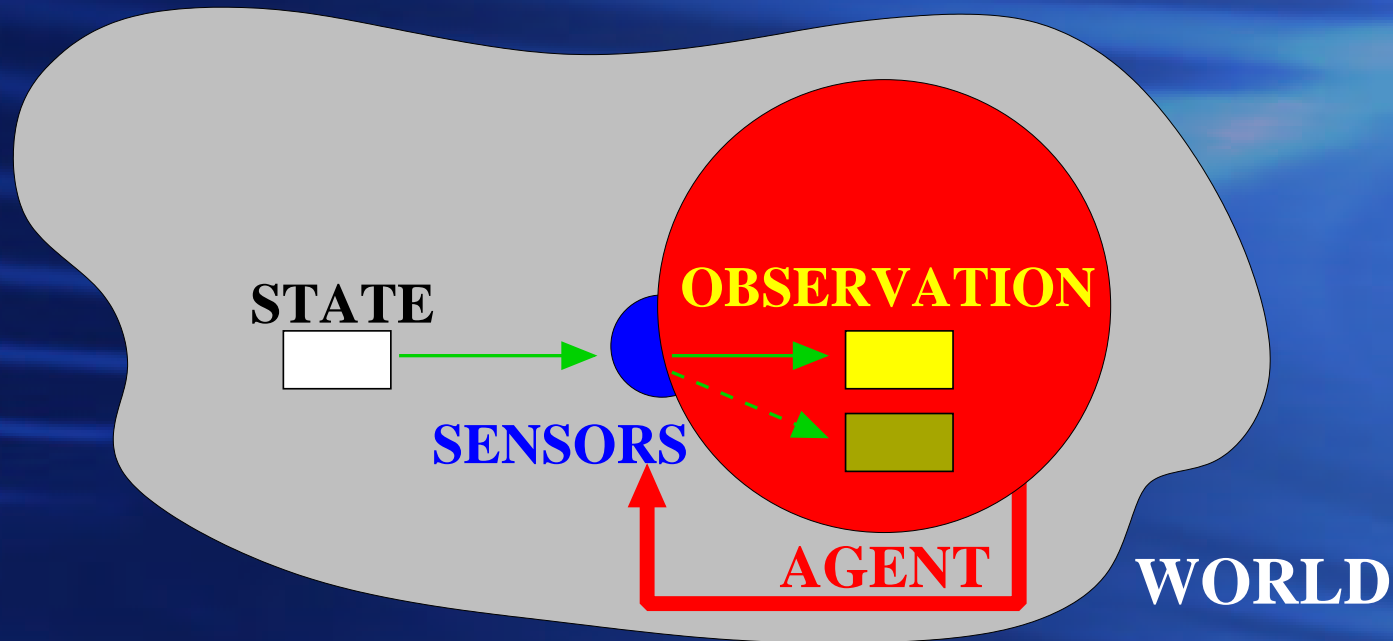
So...

- If we can provide a deterministic agent with a way to differentiate between states that it would otherwise observe to be the same, it should be able to vary its actions in those states and thus possibly learn better solutions.

Hence... Active Perception



Active Perception



Active perception is the ability of an agent to actively direct its own sensors and thus control the observations it receives from the environment.

Research Questions

- Our long term goal is to see if active perception can deliver benefits to a simple reactive agent.



Research Questions

- Our long term goal is to see if active perception can deliver benefits to a simple reactive agent.
- However the research questions we address in this work are:



Research Questions

- Our long term goal is to see if active perception can deliver benefits to a simple reactive agent.
- However the research questions we address in this work are:
 - Could reinforcement learning algorithms make use of additional resources (such as active perception) to achieve better solutions?



Research Questions

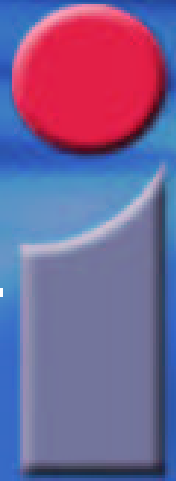


- Our long term goal is to see if active perception can deliver benefits to a simple reactive agent.
- However the research questions we address in this work are:
 - Could reinforcement learning algorithms make use of additional resources (such as active perception) to achieve better solutions?
 - What quality of information should these resources provide?

Oracles

To answer these questions we introduce Oracles. They are a useful initial test because compared to a real active perception systems:

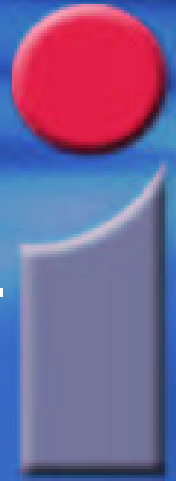
- The action required by the agent to gain information is much simpler.



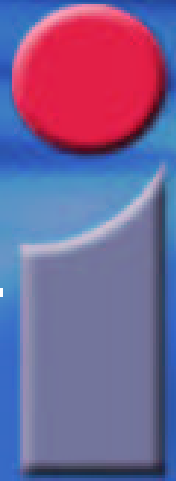
Oracles

To answer these questions we introduce Oracles. They are a useful initial test because compared to a real active perception systems:

- The action required by the agent to gain information is much simpler.
- The information gained is controllable and can be made unambiguous.



Oracles



To answer these questions we introduce Oracles. They are a useful initial test because compared to a real active perception systems:

- The action required by the agent to gain information is much simpler.
- The information gained is controllable and can be made unambiguous.

If the agent fails to learn when to consult an oracle it seems unlikely it can learn to coordinate its sensors in an active fashion to gain useful information.

Oracles

We devised oracles to test the possibility of using active perception. However, it is possible to think of them in a wider context. Two example robotic applications of “oracles” are:



Oracles

We devised oracles to test the possibility of using active perception. However, it is possible to think of them in a wider context. Two example robotic applications of “oracles” are:

- A central navigational resource with limited response rate, which is shared by millions of cheap robots with limited sensors.



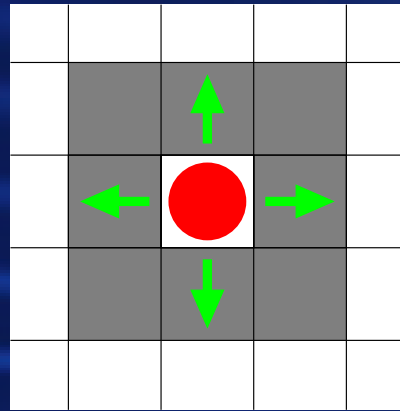
Oracles

We devised oracles to test the possibility of using active perception. However, it is possible to think of them in a wider context. Two example robotic applications of “oracles” are:

- A central navigational resource with limited response rate, which is shared by millions of cheap robots with limited sensors.
- An on-board highly accurate but power hungry sensor, where power available to the robot is limited, and less costly but coarser sensors are also available.



Grid World Experiments

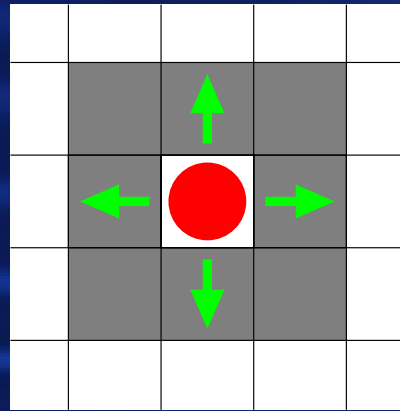


-  AGENT
-  FIXED PERCEPTIONS
-  PHYSICAL ACTIONS

Three different agents:



Grid World Experiments



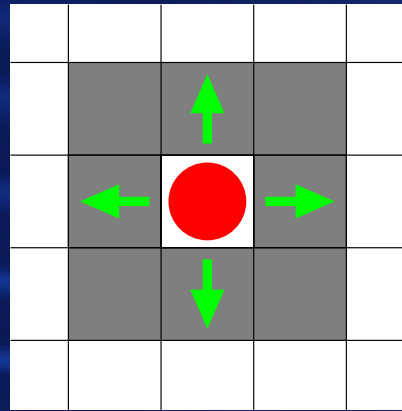
-  AGENT
-  FIXED PERCEPTIONS
-  PHYSICAL ACTIONS



Three different agents:

1. No access to an oracle,

Grid World Experiments



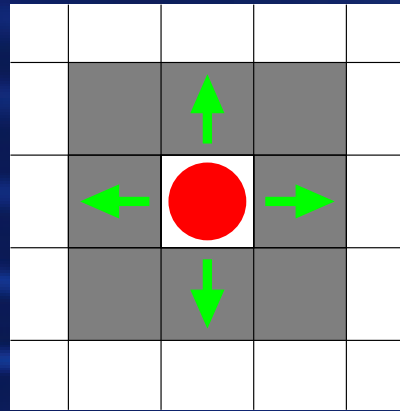
-  AGENT
-  FIXED PERCEPTIONS
-  PHYSICAL ACTIONS

Three different agents:

1. No access to an oracle,
2. Access to a State Oracle (absolute state)



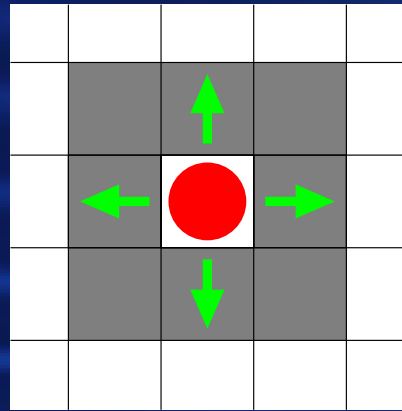
Grid World Experiments



Three different agents:

1. No access to an oracle,
2. Access to a State Oracle (absolute state)
3. Access to an Action Oracle (optimal action)

Grid World Experiments



Three different agents:

1. No access to an oracle,
2. Access to a State Oracle (absolute state)
3. Access to an Action Oracle (optimal action)

The latter two agents have one additional action which is to consult their respective oracles.

Sutton's Grid World



→ 47	→ 135	→ 71	→ 39	↓ 7	↓ 7	↓ 151		*
↑ 41	↑ 144		→ 40	↓ 0	↓ 0	↓ 148		↑ 189
↓ 41	↓ 148		→ 41	→ 0	↓ 0	↓ 20		↑ 157
→ 41	↓ 20		→ 9	→ 128	→ 64	→ 36	→ 2	↑ 149
→ 41	→ 4	→ 2	→ 1	↑ 16		→ 8	→ 0	↑ 148
→ 233	→ 224	→ 224	↑ 224	↑ 228	→ 226	→ 225	↑ 224	↑ 244

The objective is to reach the goal state (*) from any position.

Grid World Experiments

- Reward of zero for action directly reaching goal.



Grid World Experiments

- Reward of zero for action directly reaching goal.
- Reward of -1 for any other action.



Grid World Experiments

- Reward of zero for action directly reaching goal.
- Reward of -1 for any other action.
- Learning algorithms tried:



Grid World Experiments

- Reward of zero for action directly reaching goal.
- Reward of -1 for any other action.
- Learning algorithms tried:
 - Q-learning;



Grid World Experiments



- Reward of zero for action directly reaching goal.
- Reward of -1 for any other action.
- Learning algorithms tried:
 - Q-learning;
 - SARSA;

Grid World Experiments



- Reward of zero for action directly reaching goal.
- Reward of -1 for any other action.
- Learning algorithms tried:
 - Q-learning;
 - SARSA;
 - SARSA(λ);

Grid World Experiments



- Reward of zero for action directly reaching goal.
- Reward of -1 for any other action.
- Learning algorithms tried:
 - Q-learning;
 - SARSA;
 - SARSA(λ);
 - Watkins's Q(λ).

Example Policies Learnt



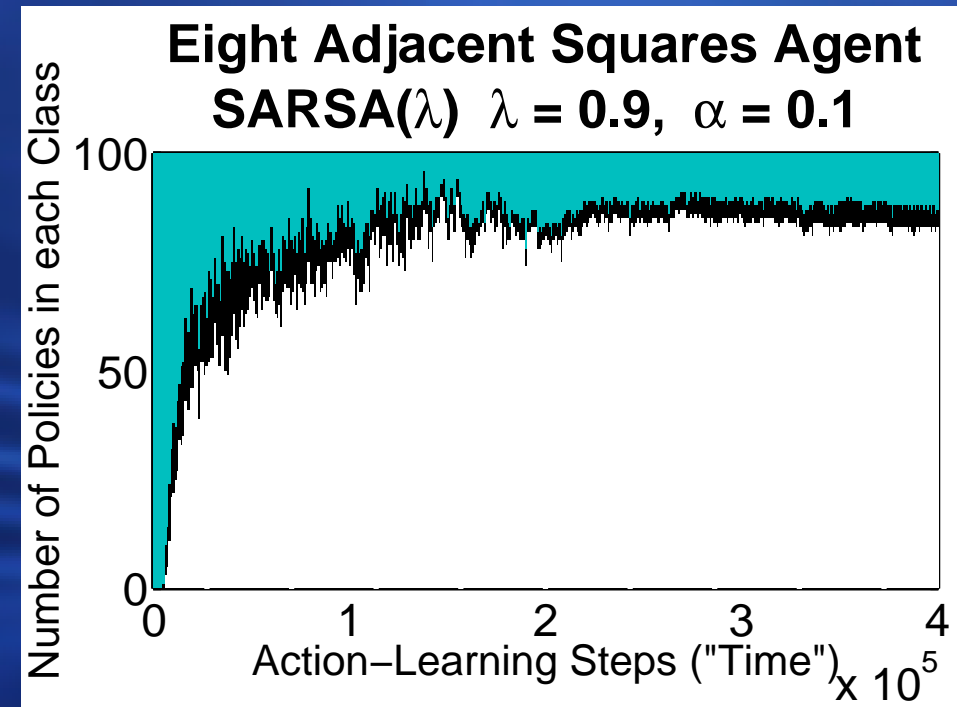
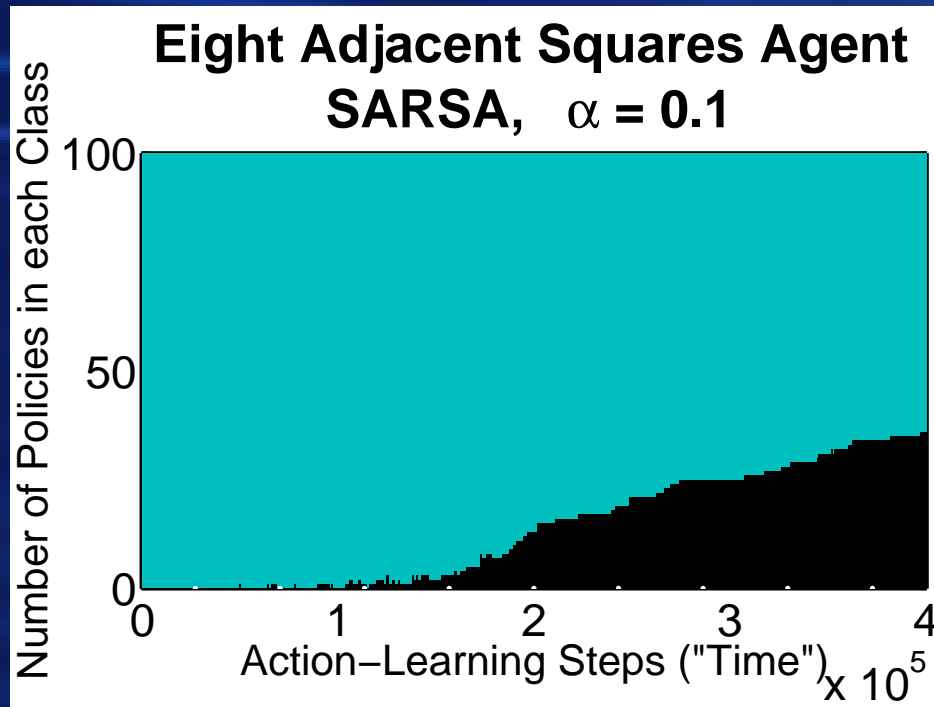
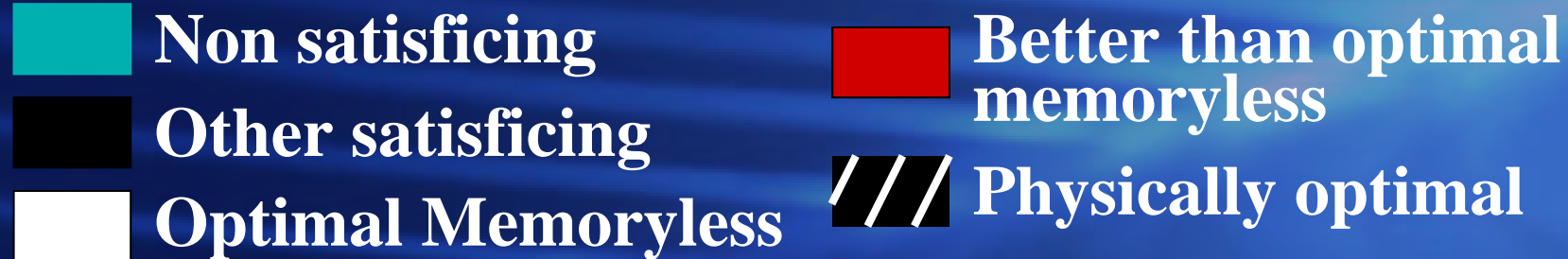
↓	↓	←	→	↓	↓	←	*
→	↓	█	→	→	→	↓	↑
→	↓	█	→	→	→	↓	↑
→	↓	█	→	→	→	→	↑
→	→	→	→	↑	█	↑	↑
→	↑	↑	↑	←	←	↑	←

Action Oracle, SARSA, $\alpha=0.1$

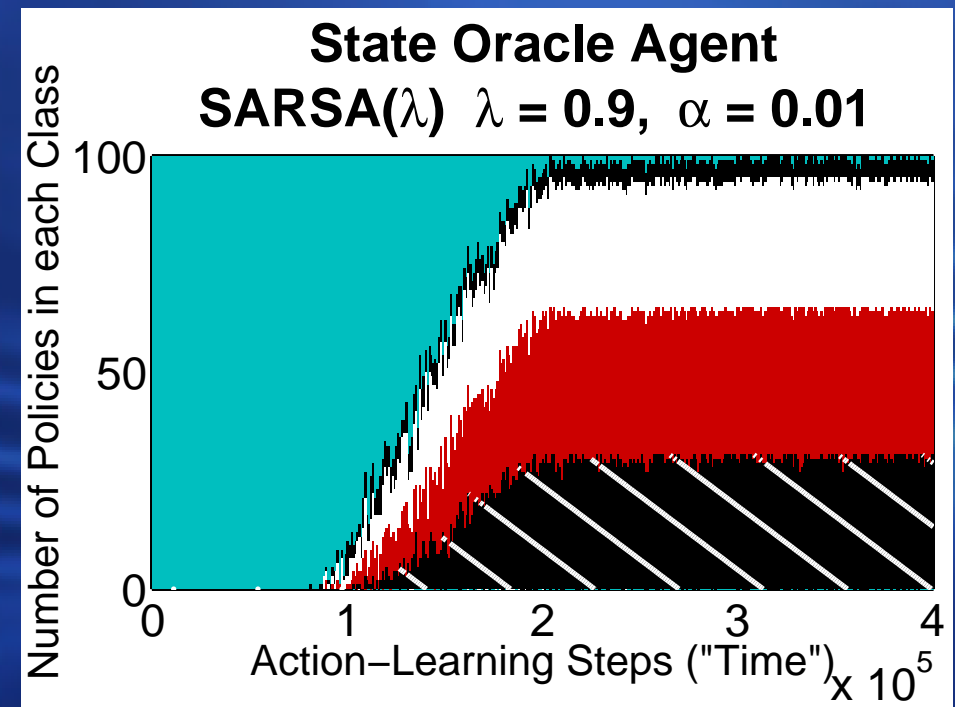
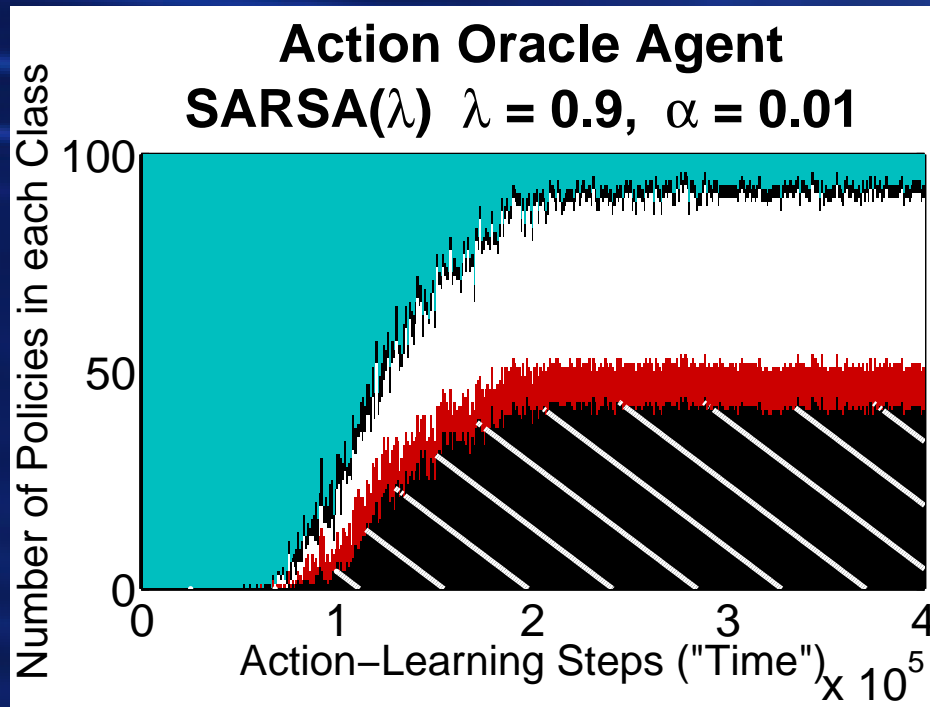
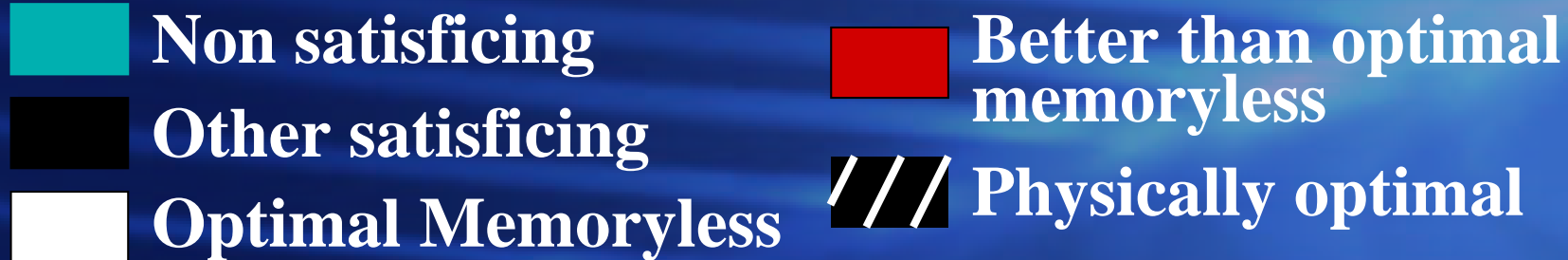
→	→	→	↓	↓	↓	↓	*
→	↓	█	↓	→	→	↓	↑
→	↓	█	→	→	→	↓	↑
→	↓	█	→	→	→	→	↑
→	→	→	→	↑	█	↑	↑
→	↑	↑	↑	↑	→	↑	↑

State Oracle, SARSA(0.9) $\alpha=0.01$

Quality of Policies



Quality of Policies



Categories based on number of physical actions.

Conclusions

- Basic reinforcement learning algorithms can learn when to make use of external resources to arrive at better solutions.



Conclusions

- Basic reinforcement learning algorithms can learn when to make use of external resources to arrive at better solutions.
- The relative success of the State Oracle Agent is encouraging as it has less information about the task than the Action Oracle.



Conclusions

- Based on the success of the State Oracle we conclude that provided an agent can use an active perception system to find unambiguous observations for each state, the agent should be able to learn physically optimal solutions.



Future Work

- Looking at what prevents the generation of greater number of physically optimal solutions.



Future Work

- Looking at what prevents the generation of greater number of physically optimal solutions.
- Generating comparative results for agents using memory or learning internal models of the world.



Future Work

- Looking at what prevents the generation of greater number of physically optimal solutions.
- Generating comparative results for agents using memory or learning internal models of the world.
- Extend results to other tasks.



Future Work



- Looking at what prevents the generation of greater number of physically optimal solutions.
- Generating comparative results for agents using memory or learning internal models of the world.
- Extend results to other tasks.
- Test results using some form of active perception system. (Early results in EWRL-6 2003).

Future Work



- Looking at what prevents the generation of greater number of physically optimal solutions.
- Generating comparative results for agents using memory or learning internal models of the world.
- Extend results to other tasks.
- Test results using some form of active perception system. (Early results in EWRL-6 2003).

<http://www.dai.ed.ac.uk/homes/paulc>