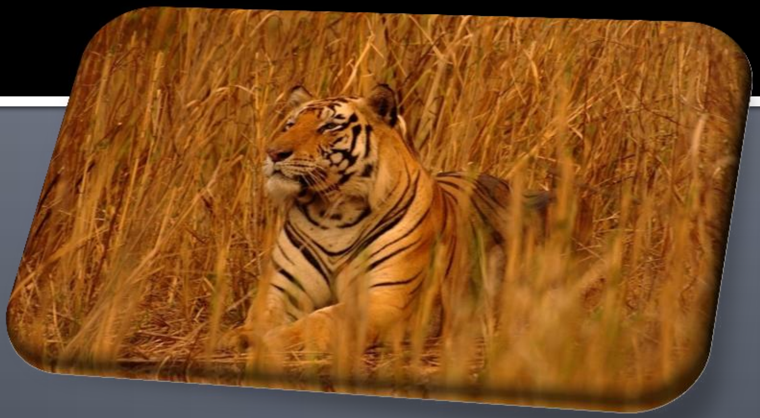
Clemens Lode, Urban Richter, Hartmut Schmeck
Karlsruhe Institute of Technology (Germany)
Institute AIFB
July, GECCO 2010

Adaption of XCS to Multi-Learner Predator/Prey Scenarios

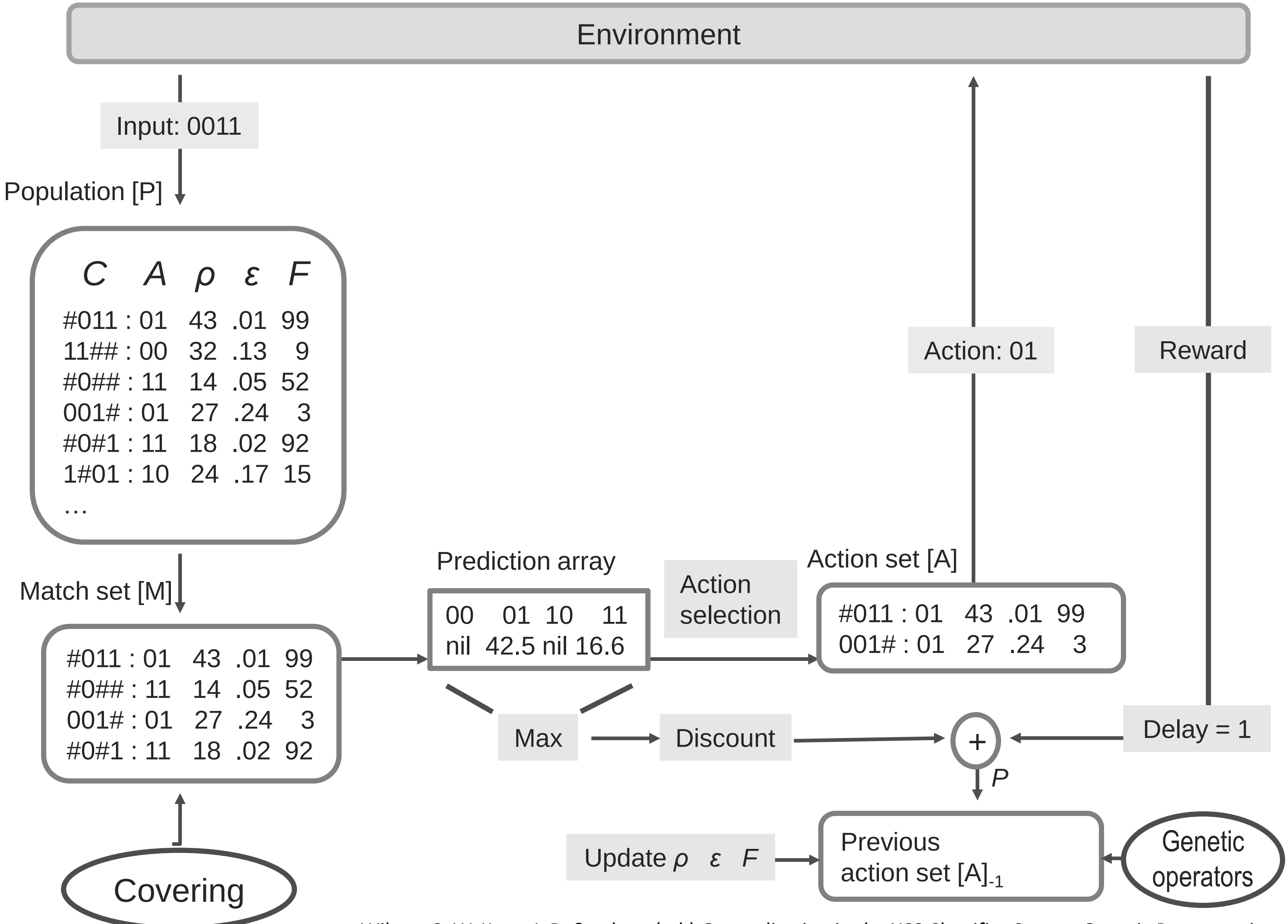


http://www.flickr.com/photos/shreeram

Outline

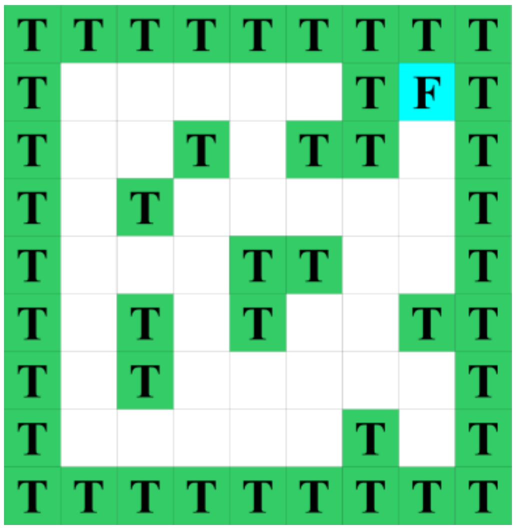
http://www.flickr.com/photos/yathin

- Learning Classifier Systems
- XCS in Predator/Prey Scenarios
- Adapting the Reward Function
- Experimental Results



Wilson, S. W. Koza, J. R. & others (ed.) Generalization in the XCS Classifier System Genetic Programming 1998: Proceedings of the Third Annual Conference, 1998, 665-674

Learning Classifier Systems



T: Tree
F: Food

- Standard (Multi-Step) Problem:
  - Maze6
- Goal:
  - Find the shortest path to from a random position to food

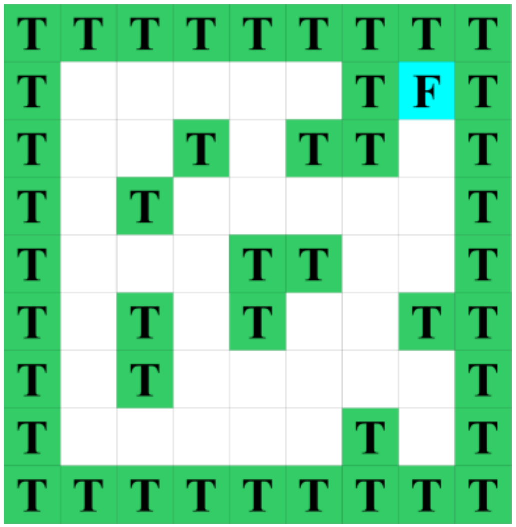
Learning Classifier Systems

T	T	T	T	T	T	T	T	T
T						T	F	T
T			T		T	T		T
T		T						T
T				T	T			T
T		T		T			T	T
T		T						T
T						T		T
T	T	T	T	T	T	T	T	T

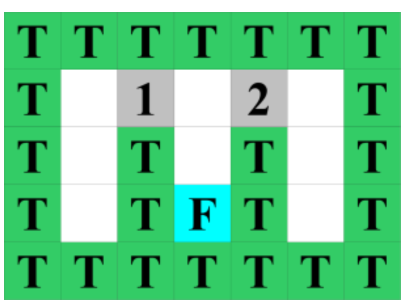
T: Tree
F: Food

- Problem:
  - Limited sensors, no global knowledge
    - Partially observable Markov decision process
- Solution:
  - Iterations, back-propagation of reward

Learning Classifier Systems



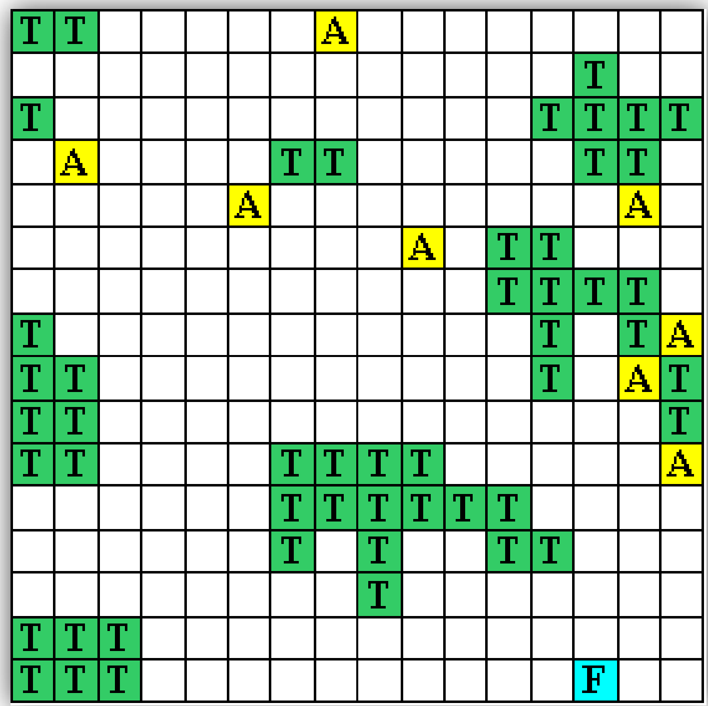
T: Tree
F: Food



- Problem:
  - Limited sensors, no global knowledge
    - Partially observable Markov decision process
- Solution:
  - Iterations, back-propagation of reward
- Aliasing positions:
  - Handle by using memory

Learning Classifier Systems

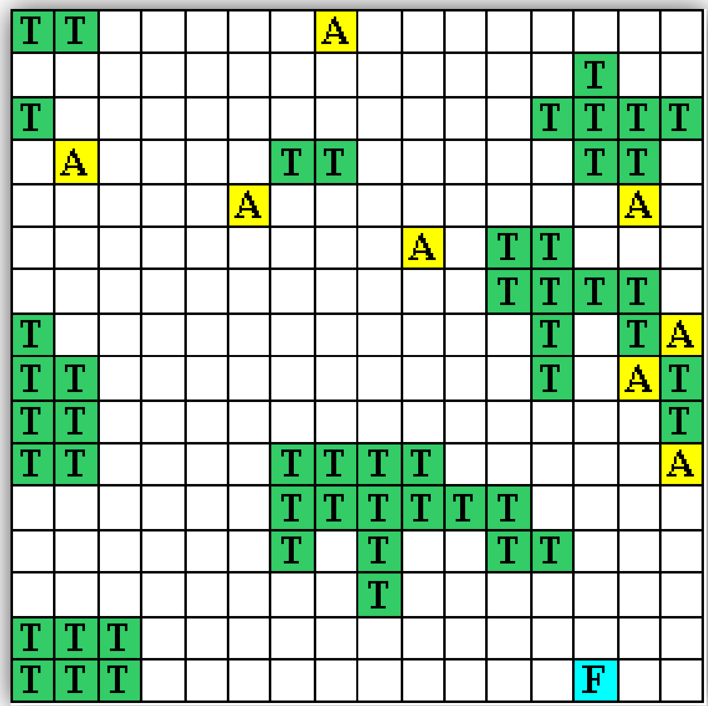
- Many aliasing positions
- Other agents present
- Dynamic world
  - food and other agents move
- Limited sensors



T: Trees
F: Food
A: Agent

Predator/Prey Scenarios

- Terminology:
  - Obstacles, prey, predator
- <u>Goal</u>: Try to stay near the prey
  - Global observation task
  - Runs continuously
  - Maximize average quality



T: Trees/Obstacles
F: Food/Prey
A: Agent/Predator

Classification of Predator/Prey Scenarios

(1) Access to local information only
(2) Open areas with some obstacles
(3) Internal state unknown to others

- No standard MDP
  - Limited sensors (1, 3)
  - Aliasing positions (2)

Classification of Predator/Prey Scenarios

(1) Access to local information only
(2) Open areas with some obstacles
(3) Internal state unknown to others
(4) Dynamic scenario

- No standard MDP
  - Limited sensors (1, 3)
  - Aliasing positions (2)
- No POMDP
  - Non-static scenario (4)

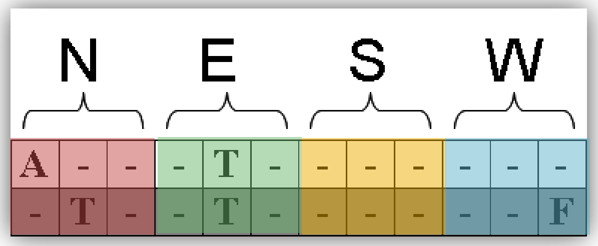
Classification of Predator/Prey Scenarios

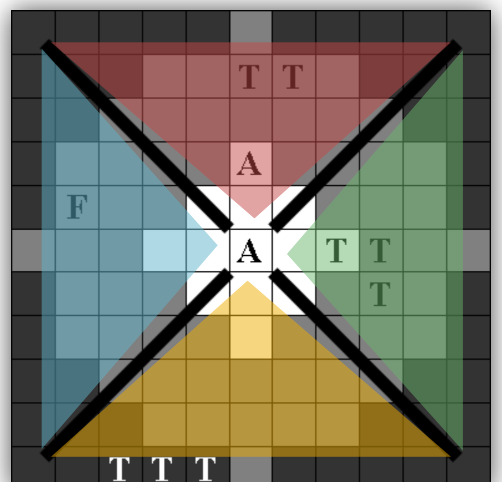
(1) Access to local information only
(2) Open areas with some obstacles
(3) Internal state unknown to others
(4) Dynamic scenario
(5) Predators share global observation task
(6) Runs continuously

- No standard MDP
  - Limited sensors (1, 3)
  - Aliasing positions (2)
- No POMDP
  - Non-static scenario (4)
- XCS has to be adapted
  - No "final" reward (5), no iterations (6)

Sensors

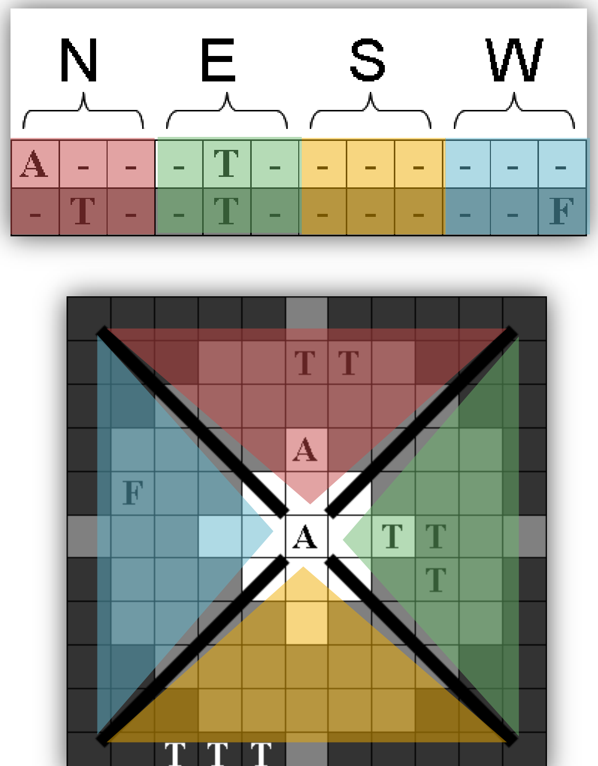
One sensor array for each direction



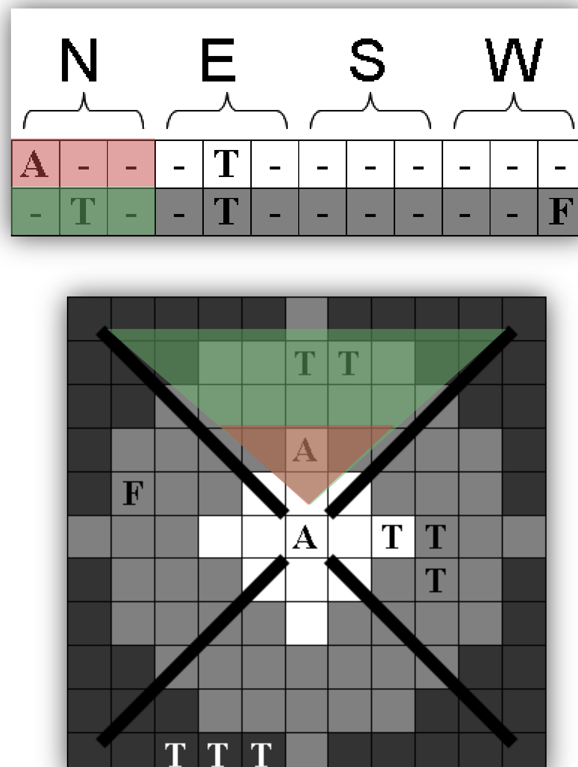


Sensors

One sensor array for each direction

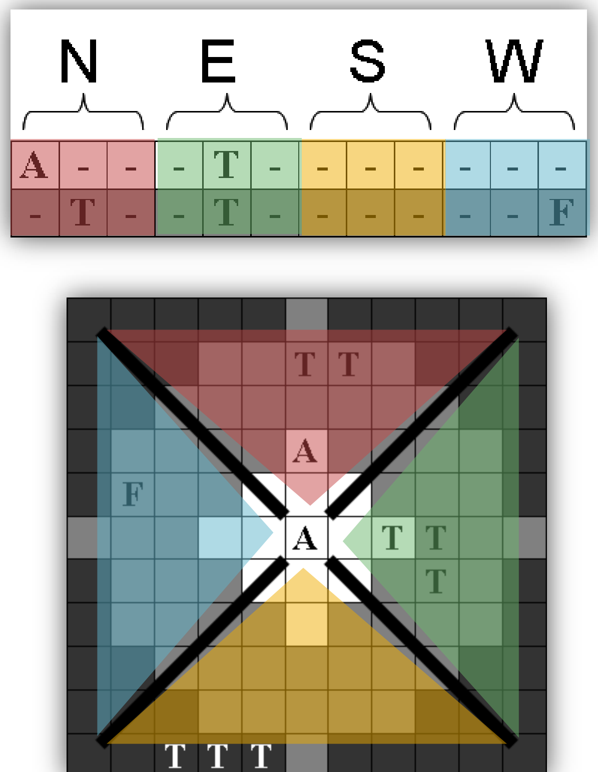


Sensors can sense either far or near (observation range / sight range)

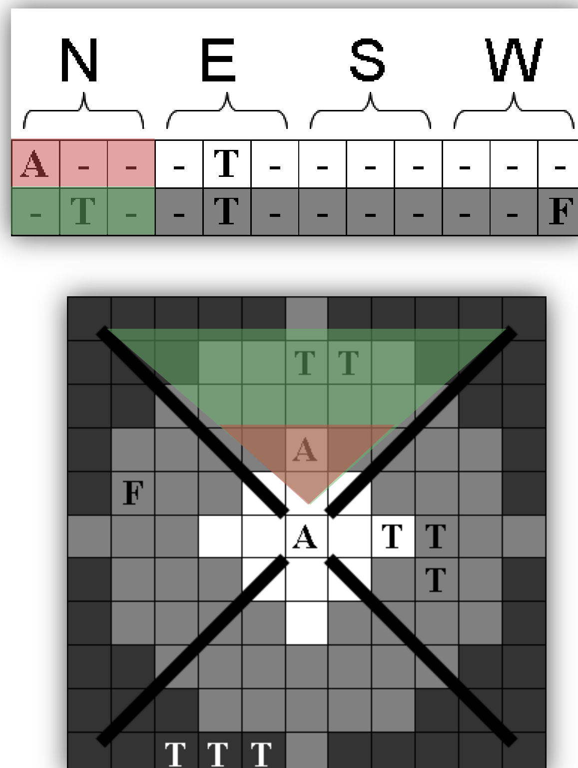


Sensors

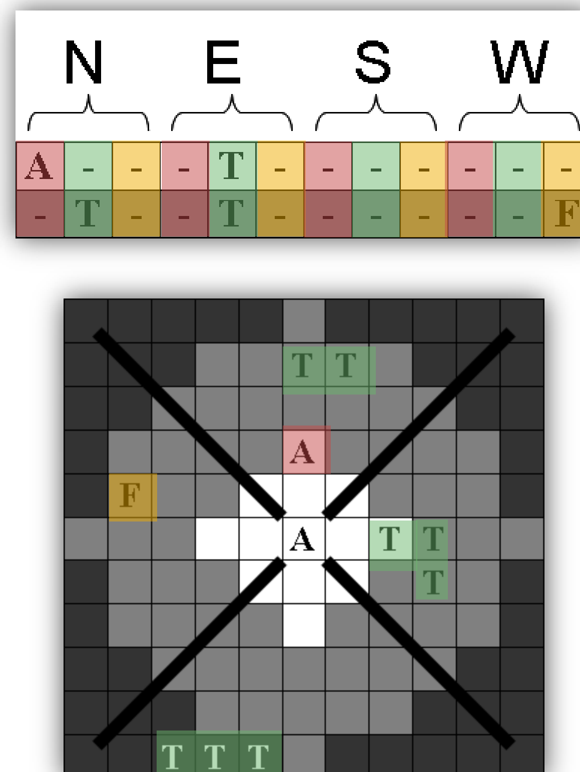
One sensor array for each direction



Sensors can sense either far or near (observation range / sight range)



Sensors can distinguish between predators, prey, and obstacles



Adaption of the Standard XCS Reward Function

- Standard implementation:
- Adapted implementation

Adaption of the Standard XCS Reward Function

- Standard implementation:
  - Reward:
    - Prey is in a neighboring cell
- Adapted implementation
  - Reward:
    - Prey is in observation range ("XCS obs")
    - Prey is in sight range ("XCS sight")

Adaption of the Standard XCS Reward Function

- Standard implementation:
  - Reward:
    - Prey is in a neighboring cell
  - Action:
    - Assign reward
    - Restart scenario
    - Switch between explore/exploit phase

- Adapted implementation
  - Reward:
    - Prey is in observation range ("XCS obs")
    - Food is in sight range ("XCS sight")
  - Action:
    - Assign reward
    - Continue scenario
    - Always use exploit phase

Classification of Predator/Prey Scenarios

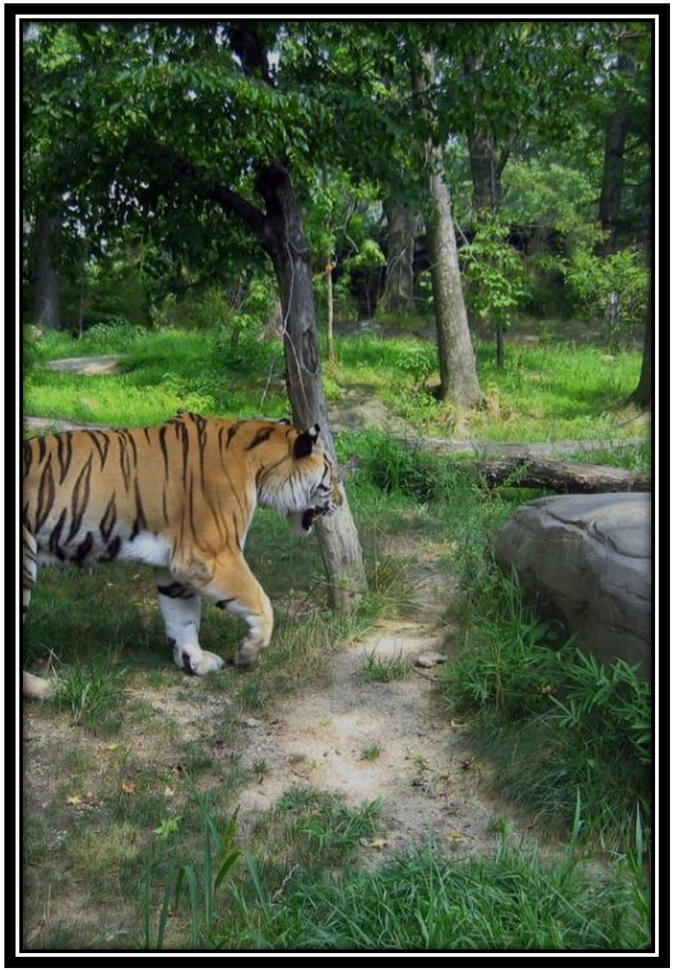
- Global knowledge cannot be reconstructed
  - Memory becomes invalid after each step

- The predator/prey scenario is a Non-observable Markov Decision Processes

Classification of Predator/Prey Scenarios

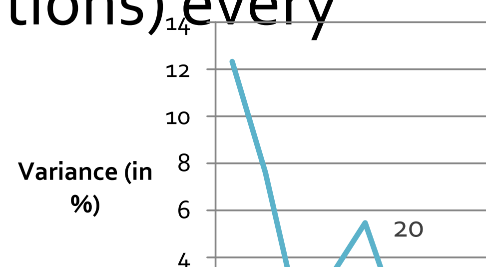
- Global knowledge cannot be reconstructed
  - Memory becomes invalid after each step
- The predator/prey scenario is a Non-observable Markov Decision Processes
- Despite being a NOMDP, can the XCS still learn?

Testing Methodology

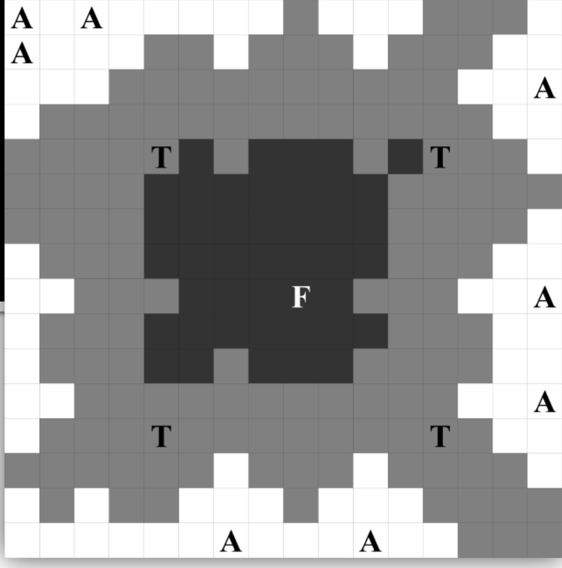


http://www.flickr.com/photos/james_crowley

- "XCS obs", "XCS sight"
- "Obstacle-evading prey"
- "Predator-evading prey"
- "Blinded Prey"
- Standard XCS parameter settings
- 2,000,000 steps
- Reset of XCS every 20,000 steps
- Reset of scenario (new random positions) every 2,000 steps

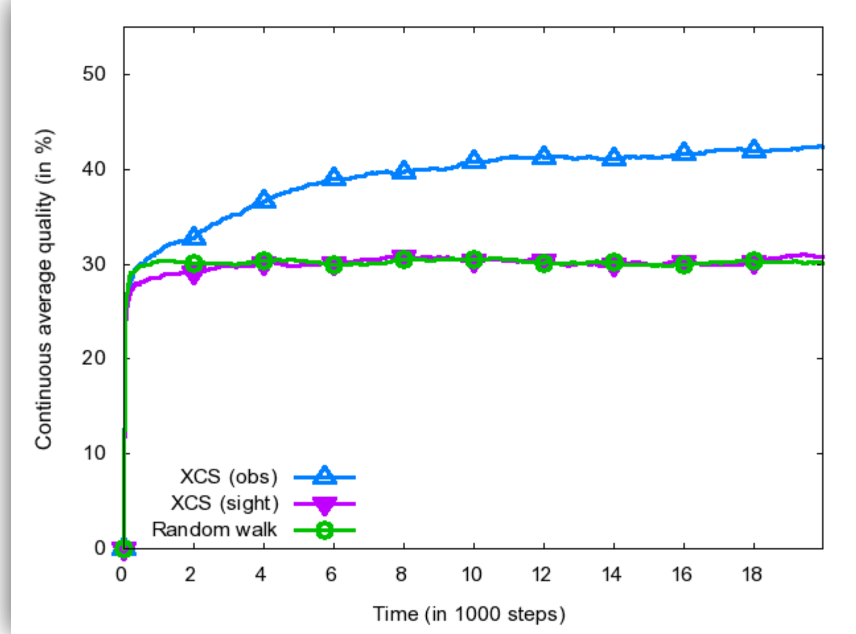


XCS Experimental Results "Pillar Scenario"

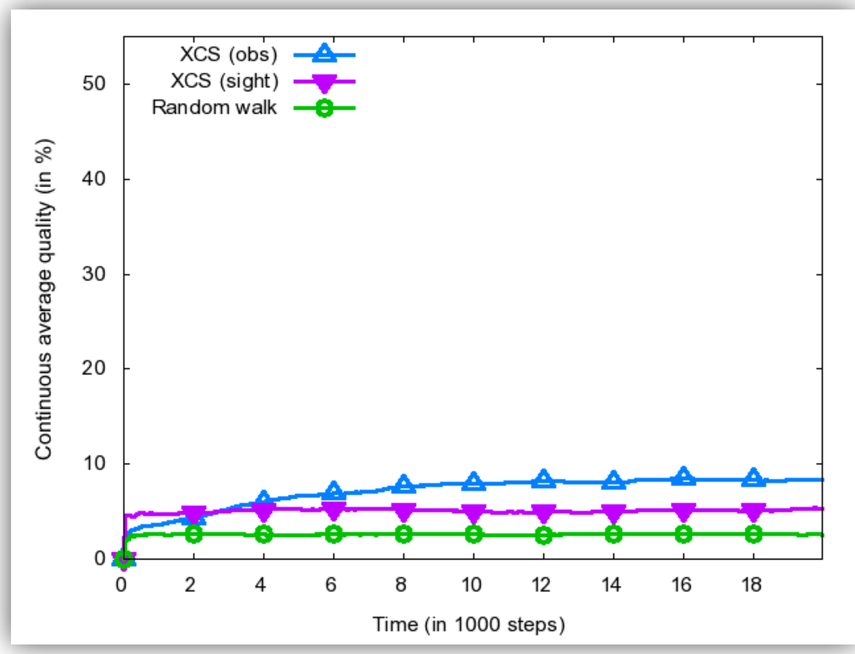


- XCS (obs) shows some learning

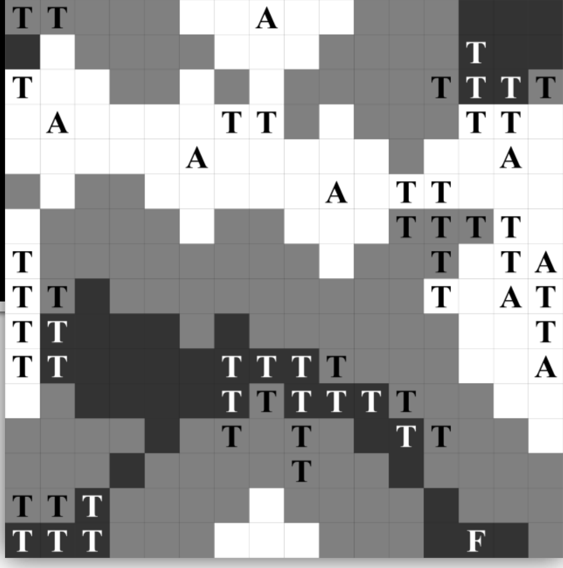
Obstacle-evading prey



Predator-evading prey

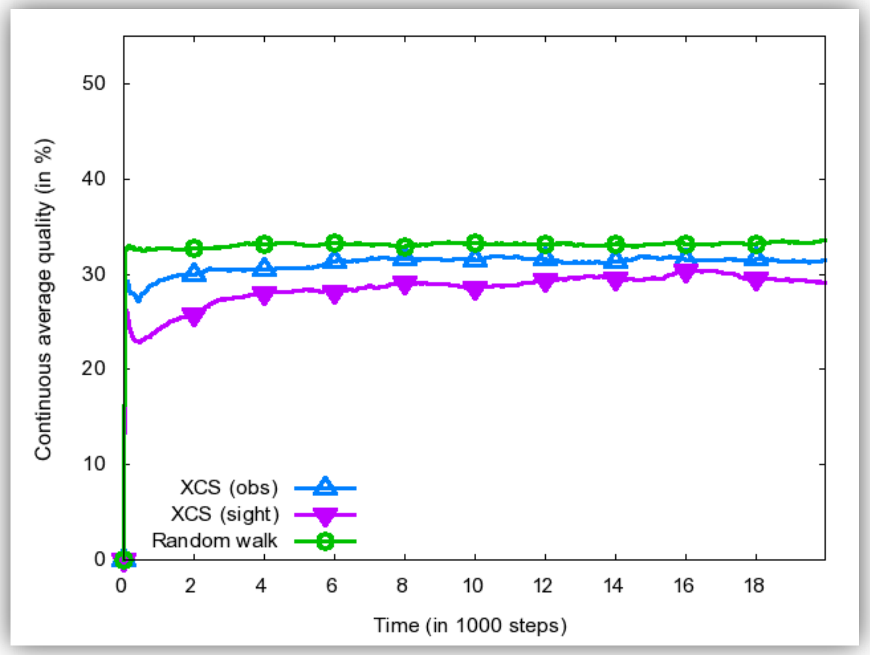


XCS Experimental Results "Random Scenario"

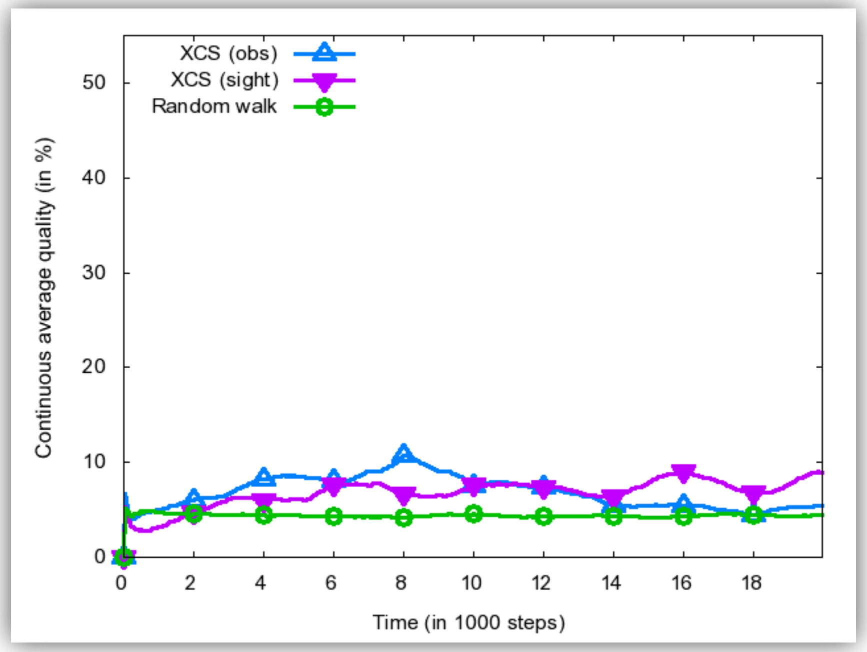


- XCS shows very little learning

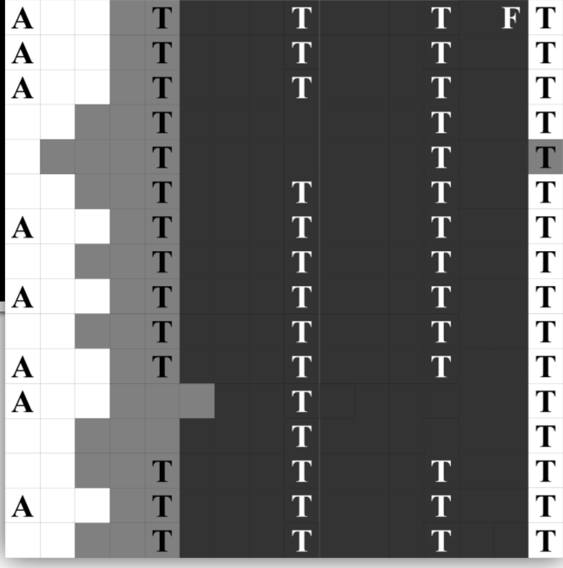
Obstacle-evading prey



Predator-evading prey



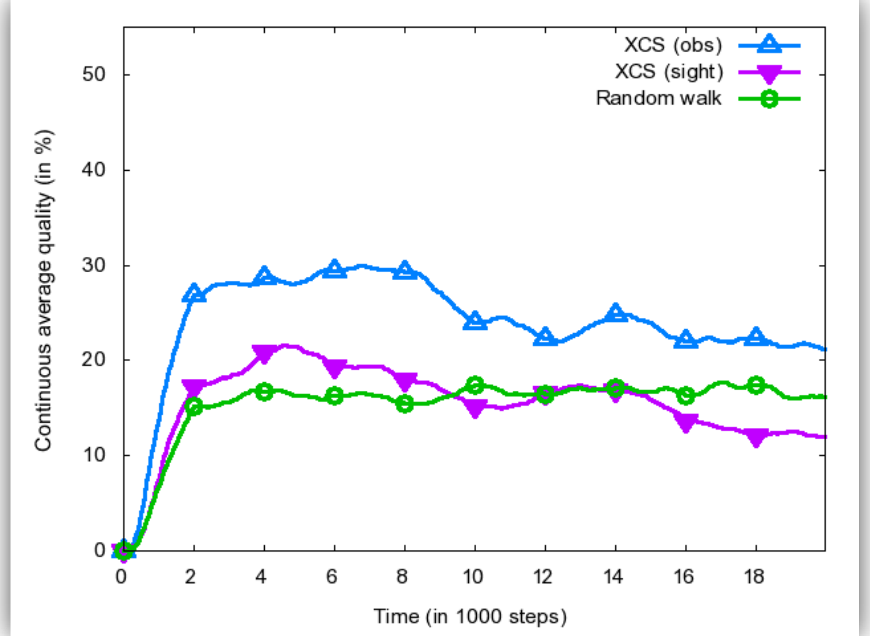
XCS Experimental Results "Difficult Scenario"



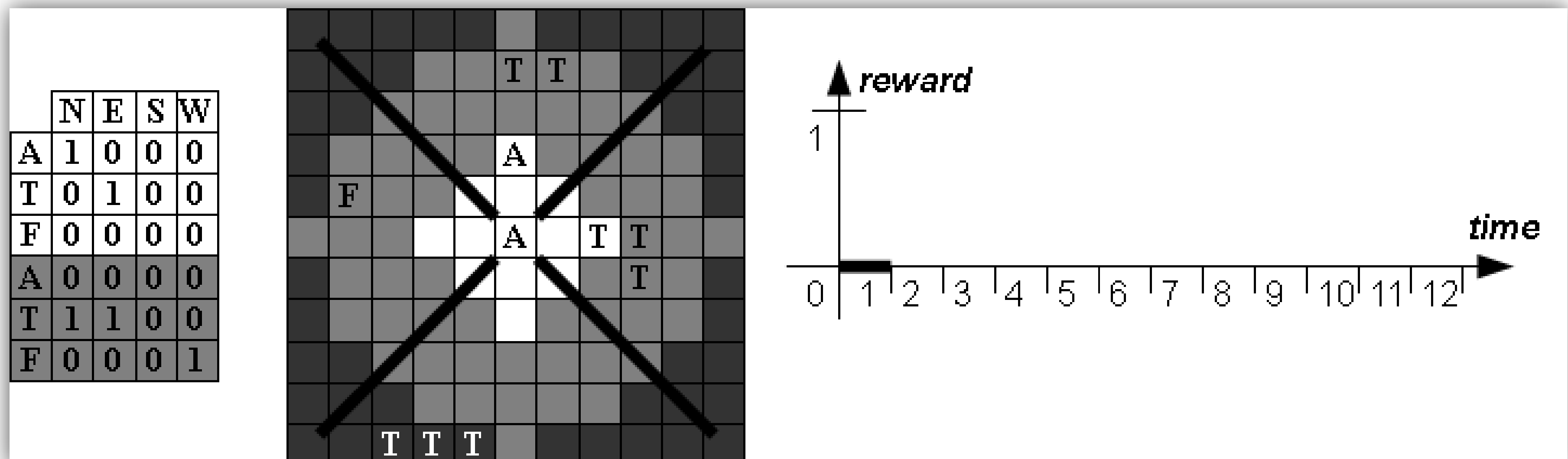
- XCS shows significant learning
  - But also unlearning after 8,000 steps

- "Difficult Scenario" is a maze-like scenario, this result was expected

Blinded prey

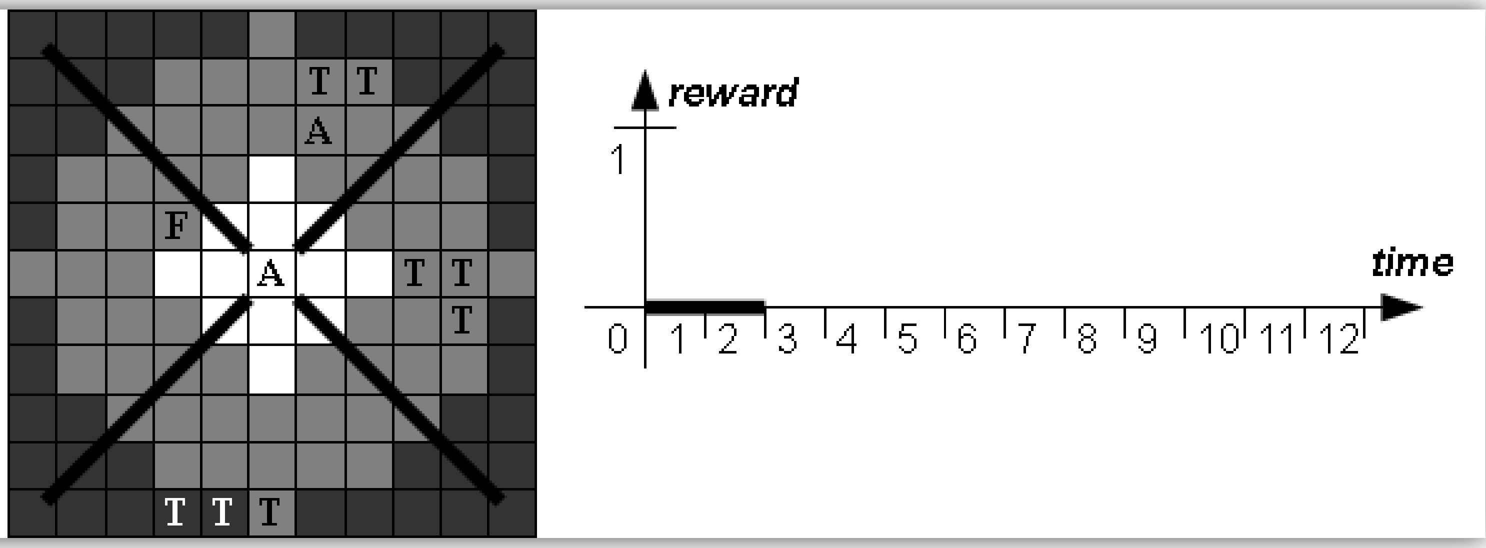


Reward Events “eventXCS”



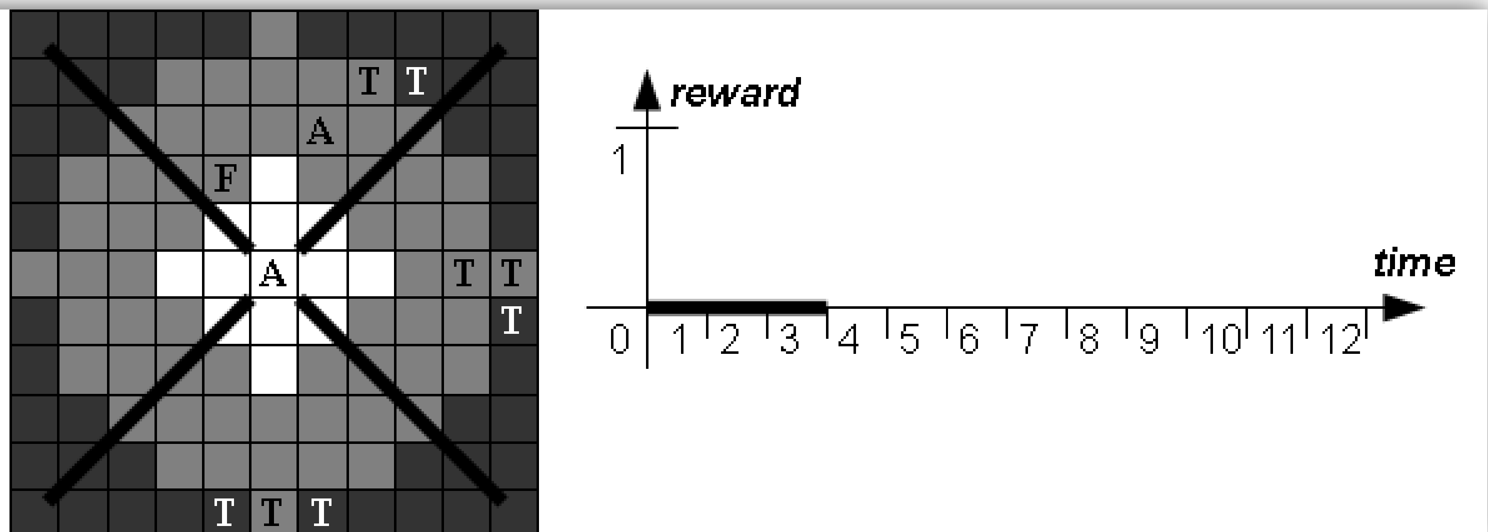
Reward Events "eventXCS"

	N	E	S	W
A	0	0	0	0
T	0	0	0	0
F	0	0	0	0
A	1	0	0	0
T	1	1	1	0
F	0	0	0	1



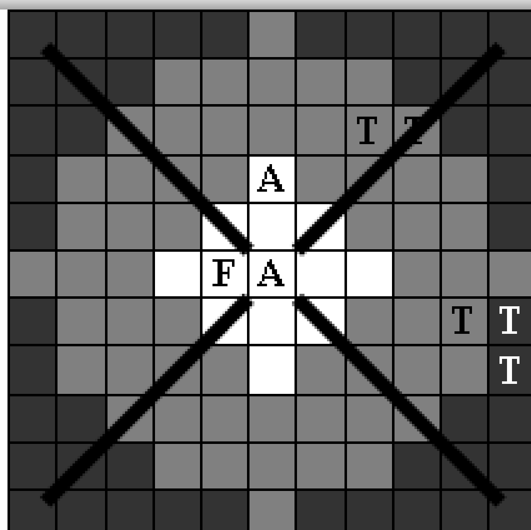
Reward Events "eventXCS"

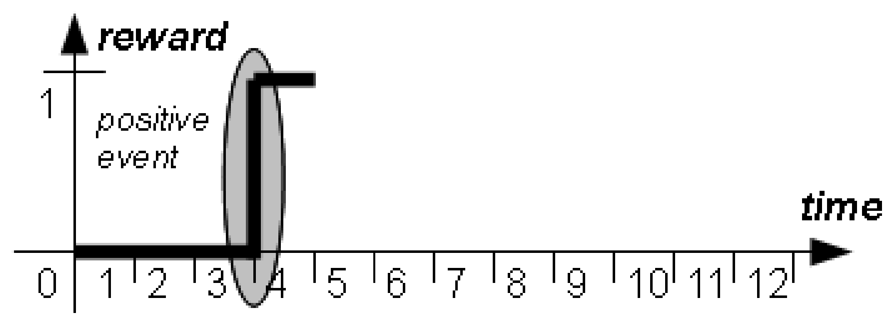
	N	E	S	W
A	0	0	0	0
T	0	0	0	0
F	0	0	0	0
A	1	0	0	0
T	1	1	1	0
F	1	0	0	0



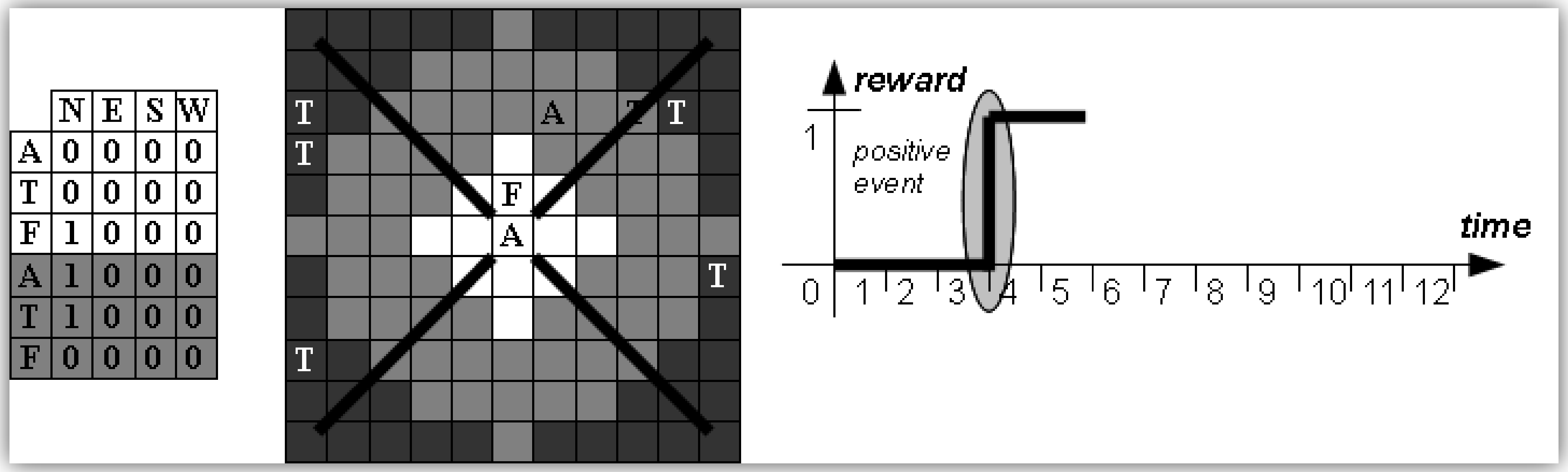
Reward Events “eventXCS”

	N	E	S	W
A	1	0	0	0
T	0	0	0	0
F	0	0	0	1
A	0	0	0	0
T	1	1	0	0
F	0	0	0	0

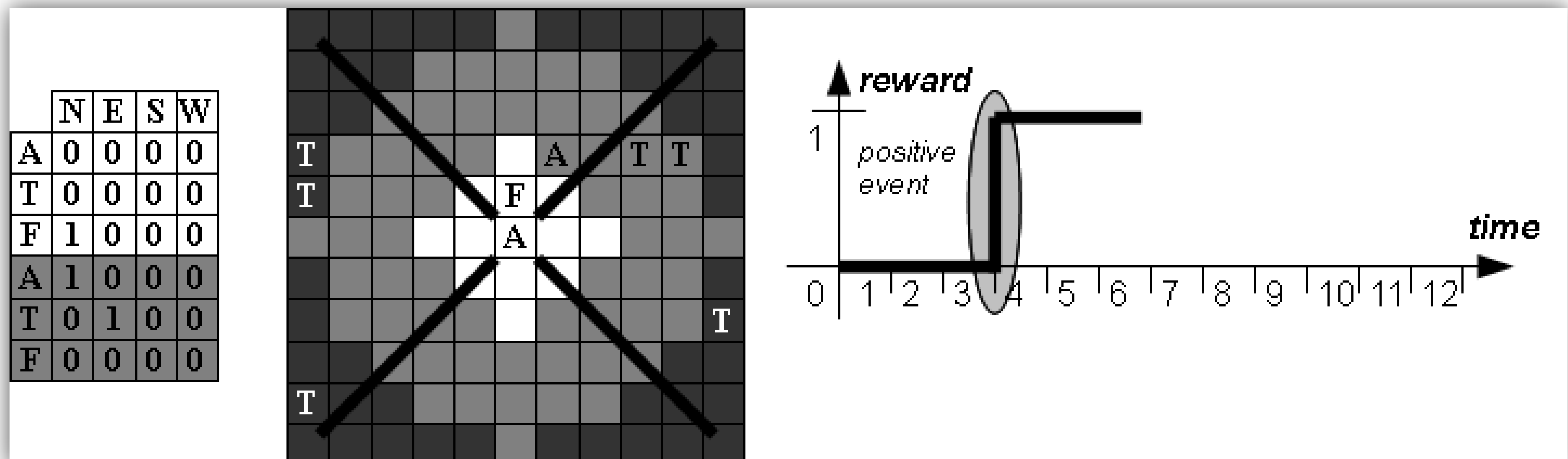




Reward Events "eventXCS"

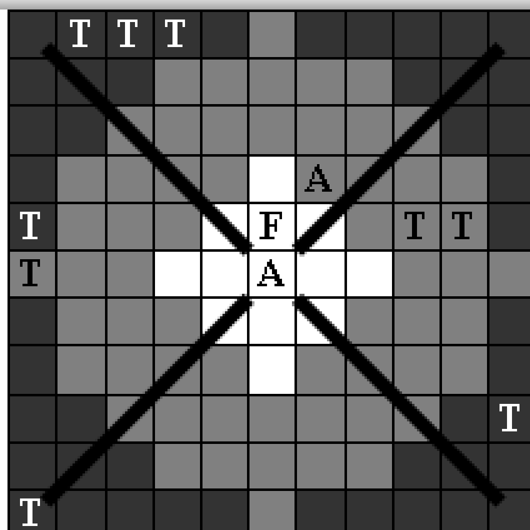


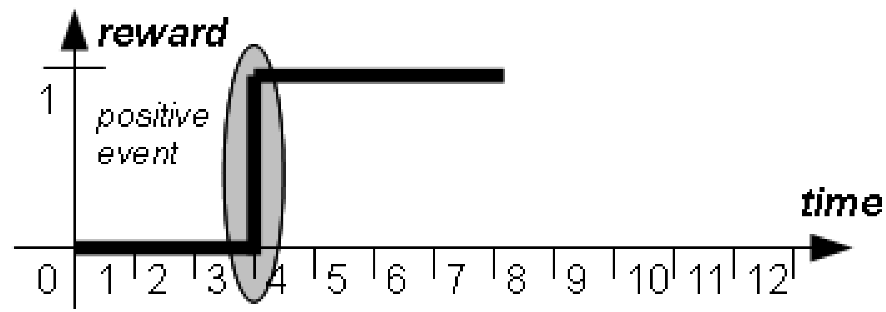
Reward Events "eventXCS"



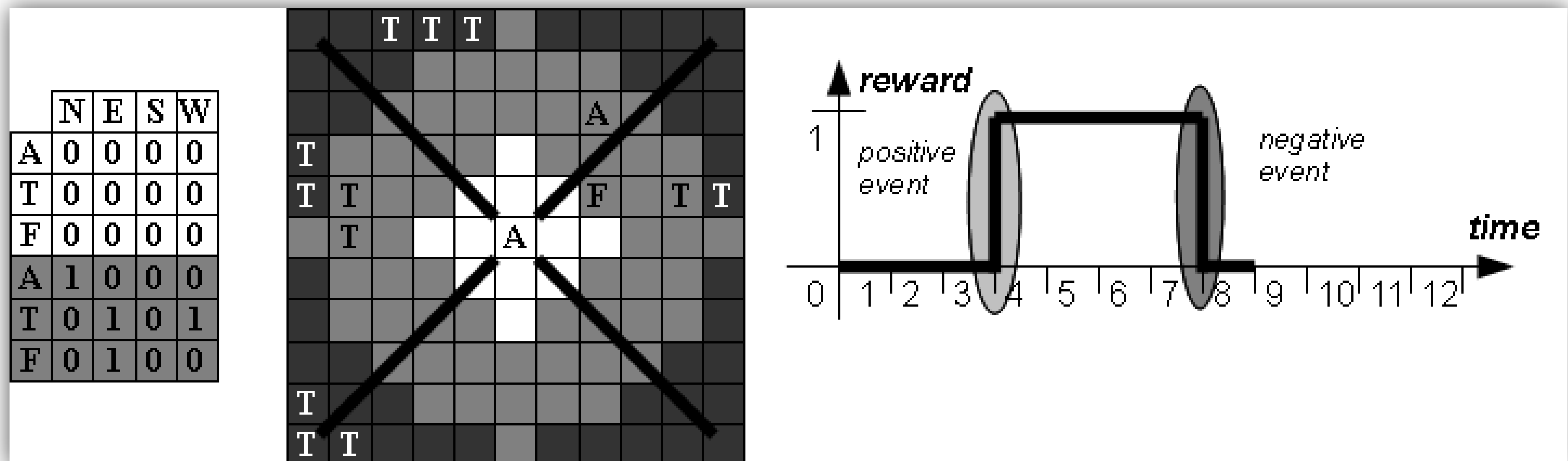
Reward Events "eventXCS"

	N	E	S	W
A	0	0	0	0
T	0	0	0	0
F	1	0	0	0
A	1	0	0	0
T	0	1	0	1
F	0	0	0	0



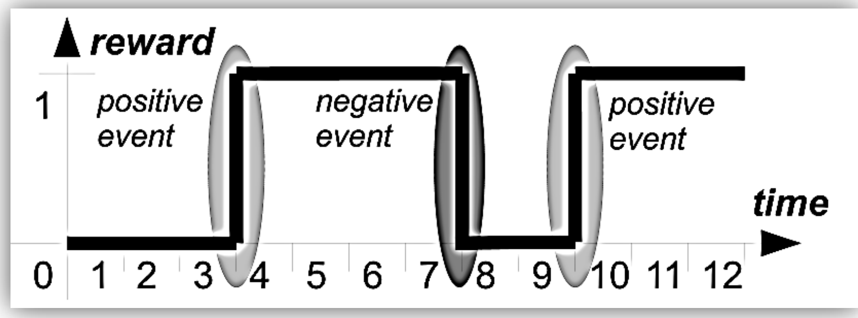


Reward Events "eventXCS"



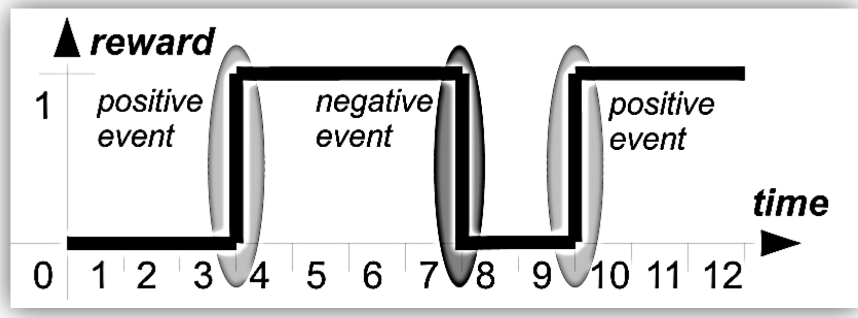
Reward Distribution "eventXCS"

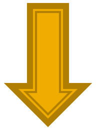
- Analyze succession of positive and negative events

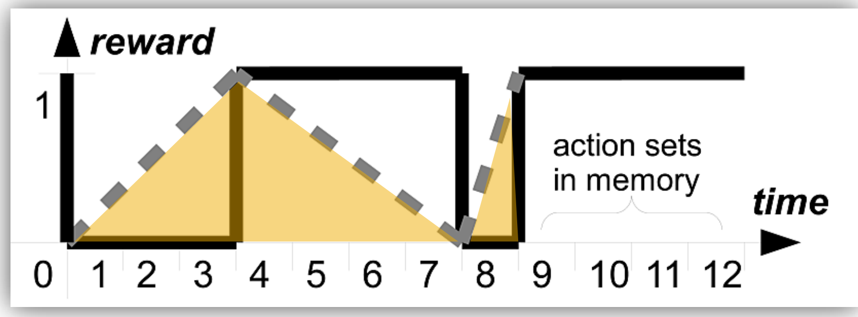


Reward Distribution "eventXCS"

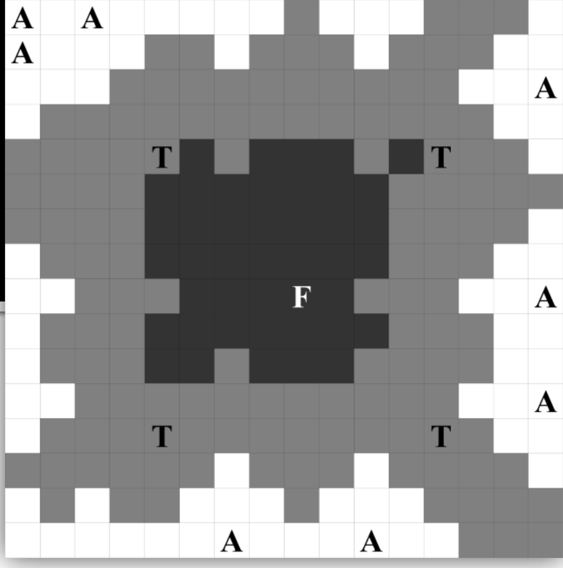
- Analyze succession of positive and negative events
- Distribute the reward as soon as possible (i.e. at each event)
- Idea:
  - Action sets close to an event probably contributed TODO





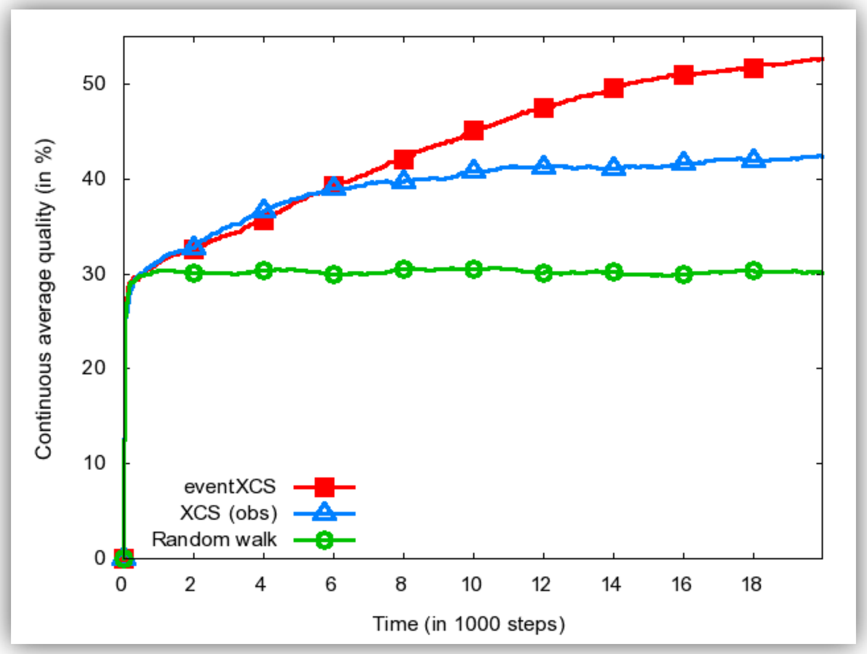


Experimental Results "Pillar Scenario"

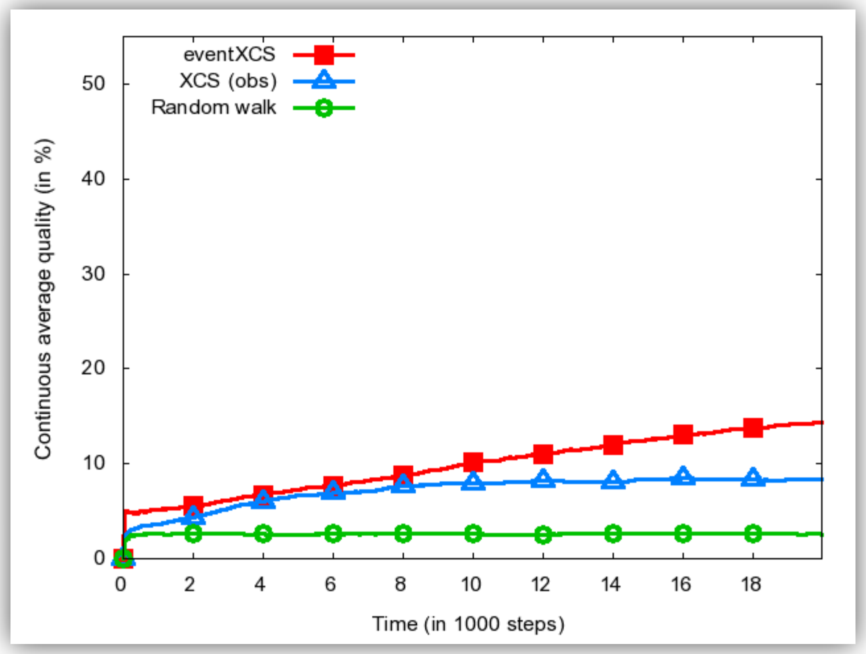


- eventXCS clearly outperforms XCS

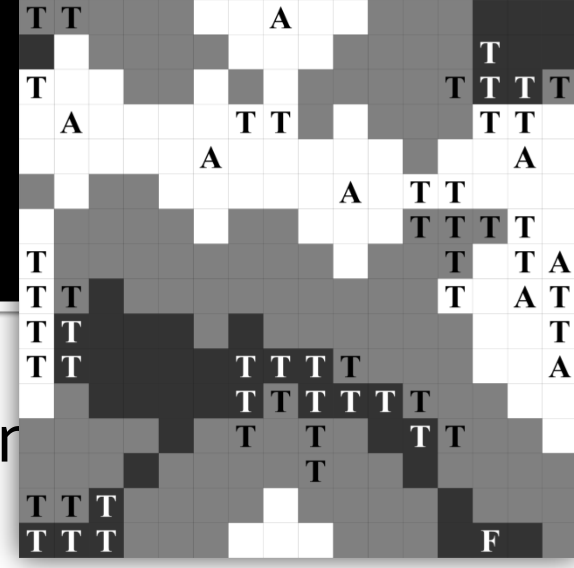
Obstacle-evading prey



Predator-evading prey

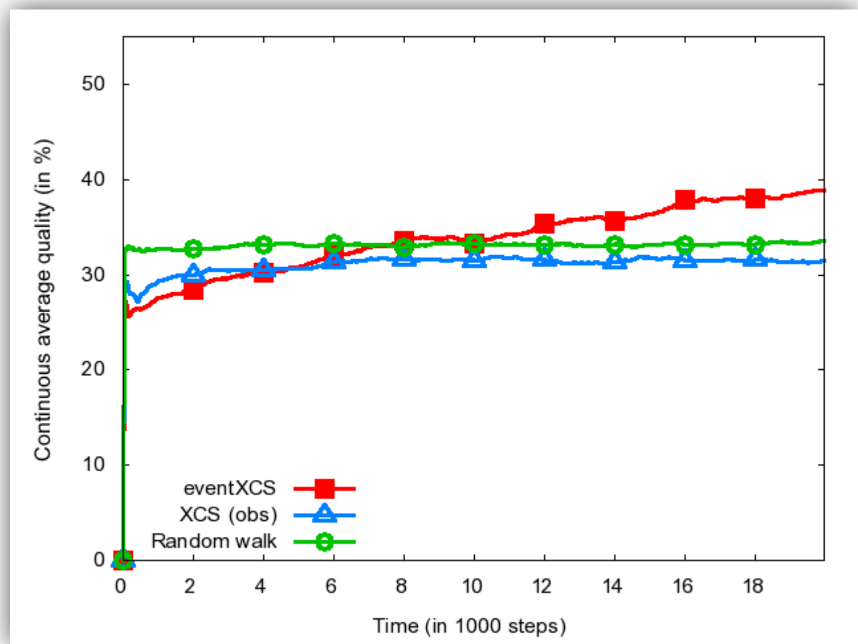


Experimental Results "Random Scenario"

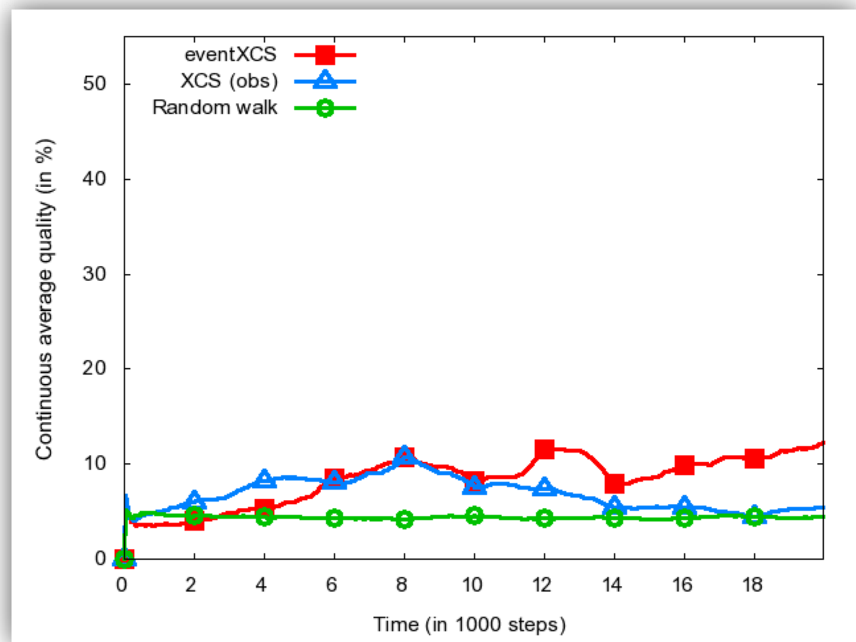


- eventXCS shows slow but steady learn obstacle-evading prey

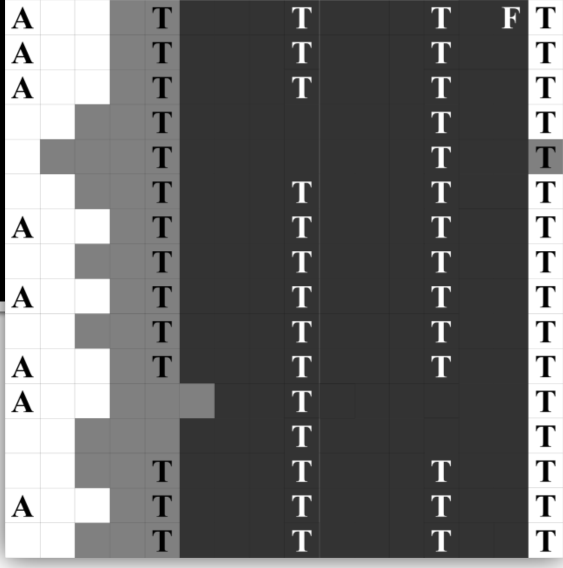
Obstacle-evading prey



Predator-evading prey

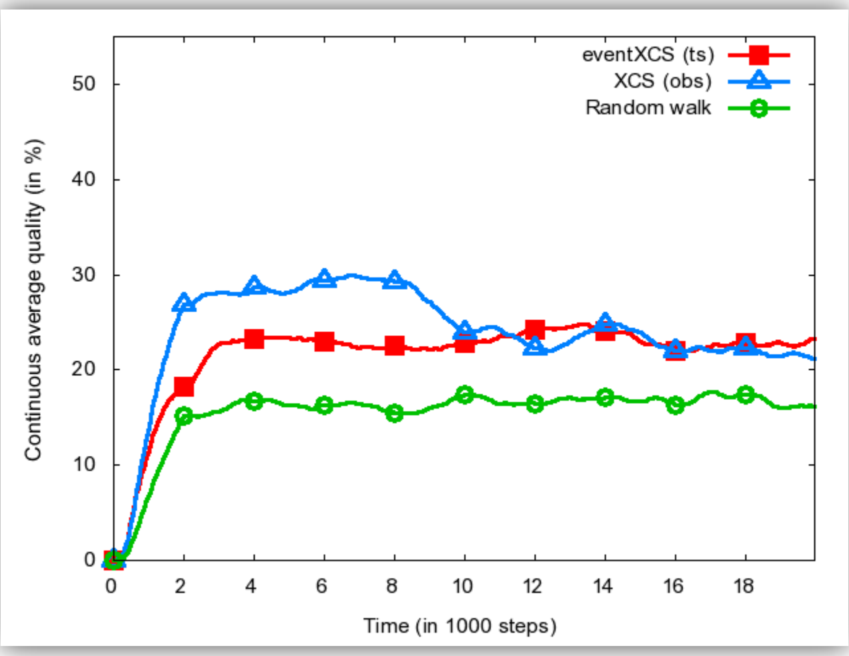


Experimental Results "Difficult Scenario"



- eventXCS fails in this scenario
- Using "tournament selection" shows acceptable results with no sign of unlearning

Blinded prey



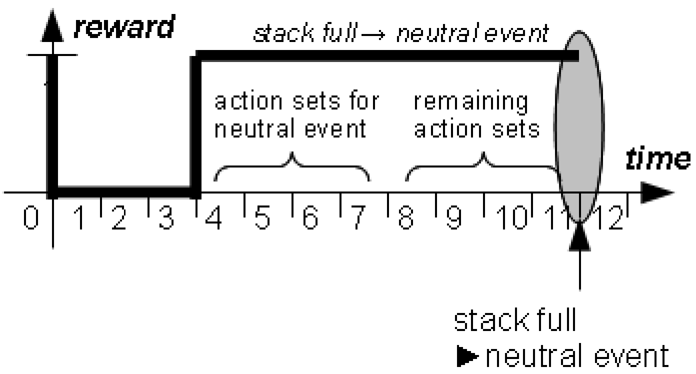
Conclusion

- XCS with minimal adaptions can learn
  - Unable to use sight range
- Event XCS superior

Backup slides

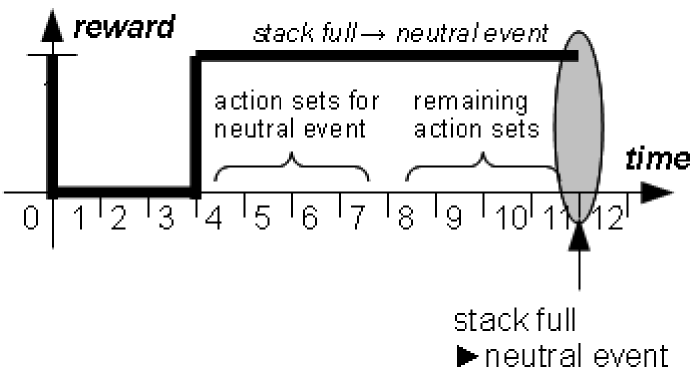
Neutral Events

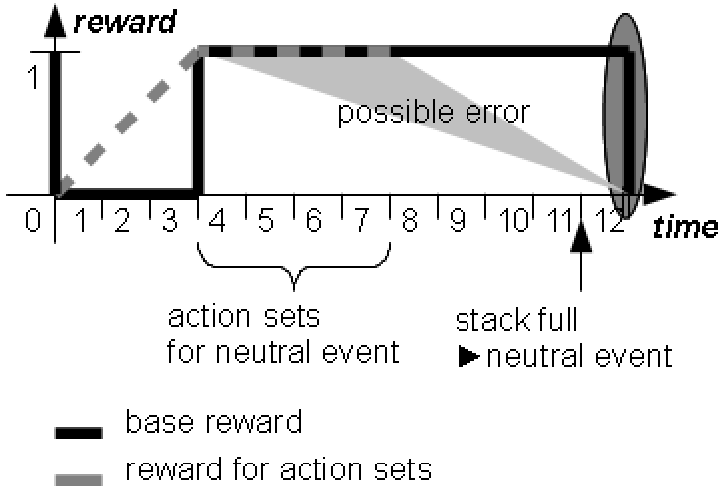
- Neutral Event
  - No positive or negative event for a number of steps
  - Half of the action sets is discarded and receives reward
  - Idea:
    - Good actions are rewarded earlier
    - Preventing of dead ends



Neutral Events

- Neutral Event
  - No positive or negative event for a number of steps
  - Half of the action sets is discarded and receives reward
  - Idea:
    - Good actions are rewarded earlier
    - Preventing of dead ends
  - Problem:
    - Error possibility high if directly followed by an event.

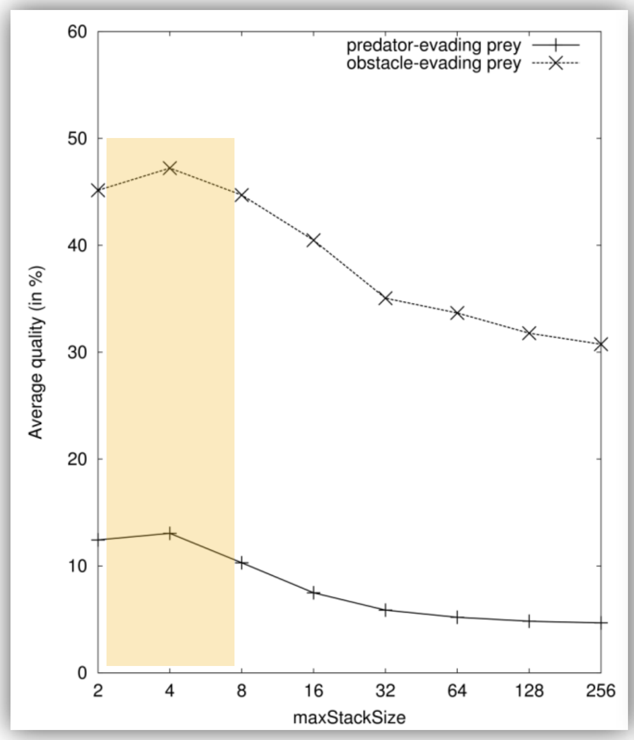




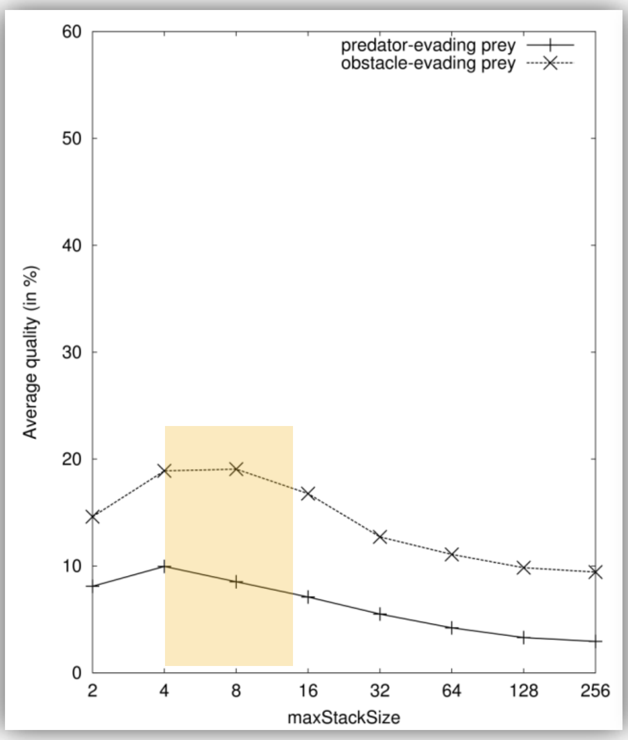
Neutral Events

- Tests have shown that a stack size of 8 is generally good for all three scenarios

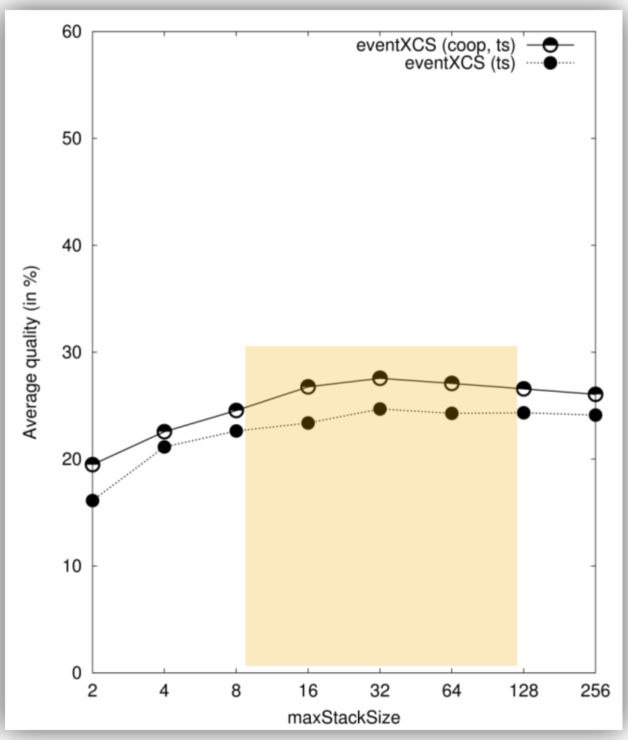
"Pillar Scenario"



"Random Scenario"

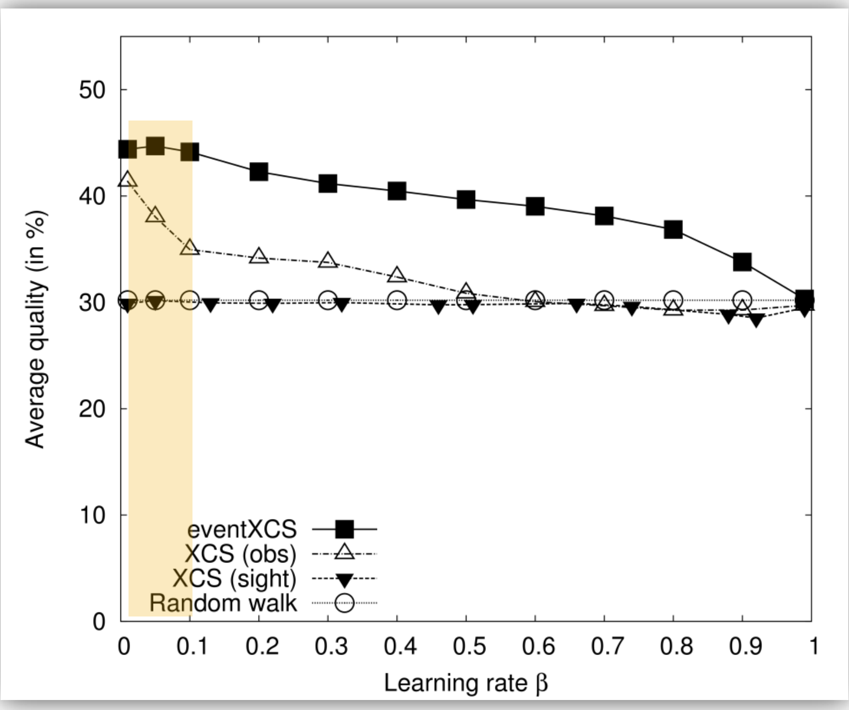


"Difficult Scenario"



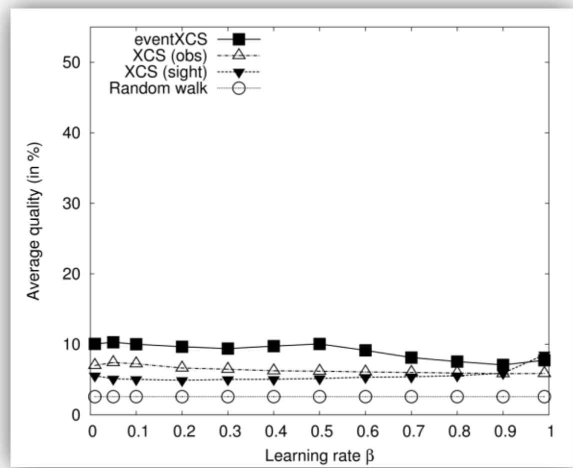
Learning Rate β

- Pillar Scenario
  - Obstacle-evading prey
  - Low learning rate (0.05) good, eventXCS very stable

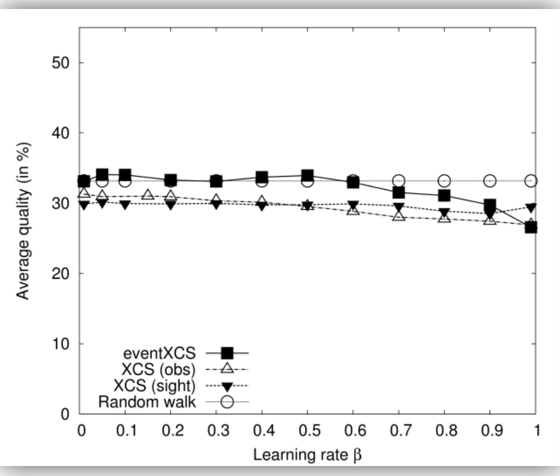


Learning Rate β

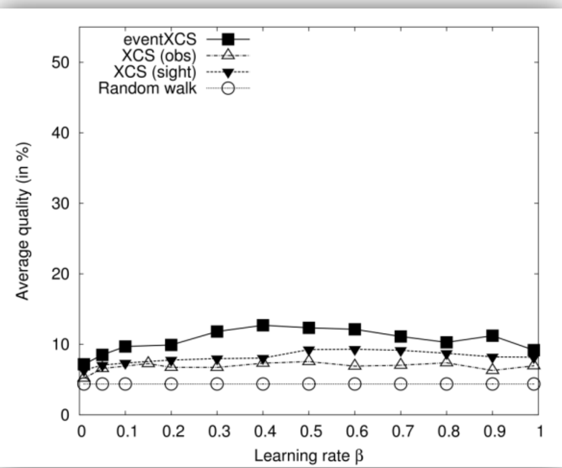
Pillar Scenario
Predator-evading
prey



Random Scenario,
Obstacle evading prey



Random Scenario,
Predator evading



Learning Rate β

- Difficult Scenario
  - Blind prey
  - High learning rates show an advantage because of long distance to the prey