|  |
| --- |
|  |
| User Guide |
|  |
|  |
|  |
|  |

|  |
| --- |
|  |

Contents

[List of Figures 1](#_Toc387164590)

[Running the Program 2](#_Toc387164591)

[Main Screen 2](#_Toc387164592)

[Trace on console 3](#_Toc387164593)

[Show the tile qValue 4](#_Toc387164594)

[Discount 5](#_Toc387164595)

[Greedy Exploit 5](#_Toc387164596)

[Fixed alpha 5](#_Toc387164597)

[Move Button 5](#_Toc387164598)

[Reset Button 6](#_Toc387164599)

[Initial Value 6](#_Toc387164600)

[Number of Move 6](#_Toc387164601)

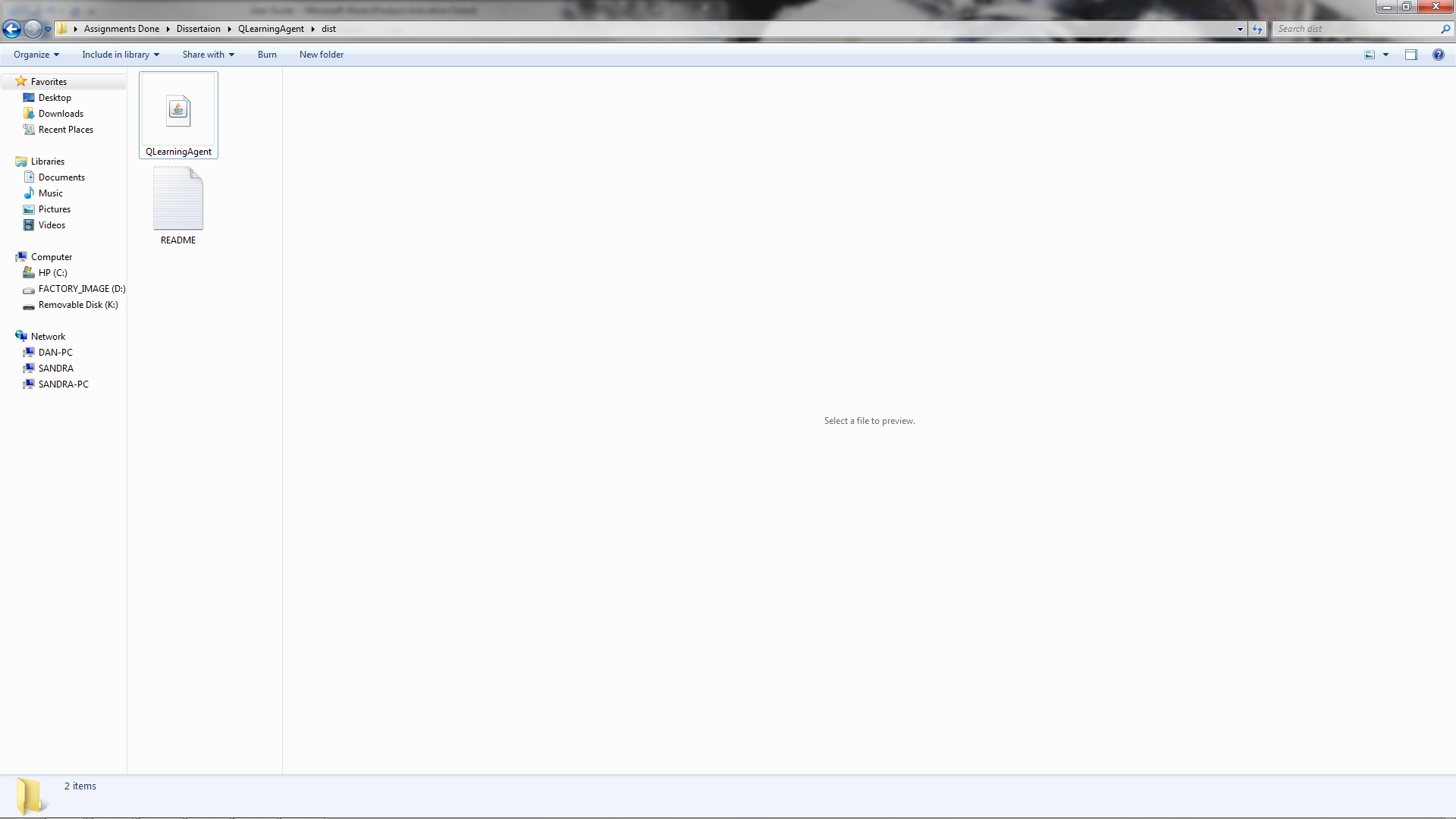
[Cumulative Reward 6](#_Toc387164602)

## List of Figures

Figure 1…………………………………………………………………………….………..2  
Figure 2………………………………………………………………...…………………...2  
Figure 3...……………………………………………………..…………………………….3  
Figure 4……………………………………………………………………………………...3  
Figure 5………………………………………………………………………………..…….4  
Figure 6………………………………..…………………………………………………….5

## Running the Program

To run the program, run the QLearningAgent.exe.



Figure

## Main Screen

When the program has been loaded the main menu screen will appear (Figure 2). This screen shows the controls to the left with the console output underneath and the maze environment to the right.

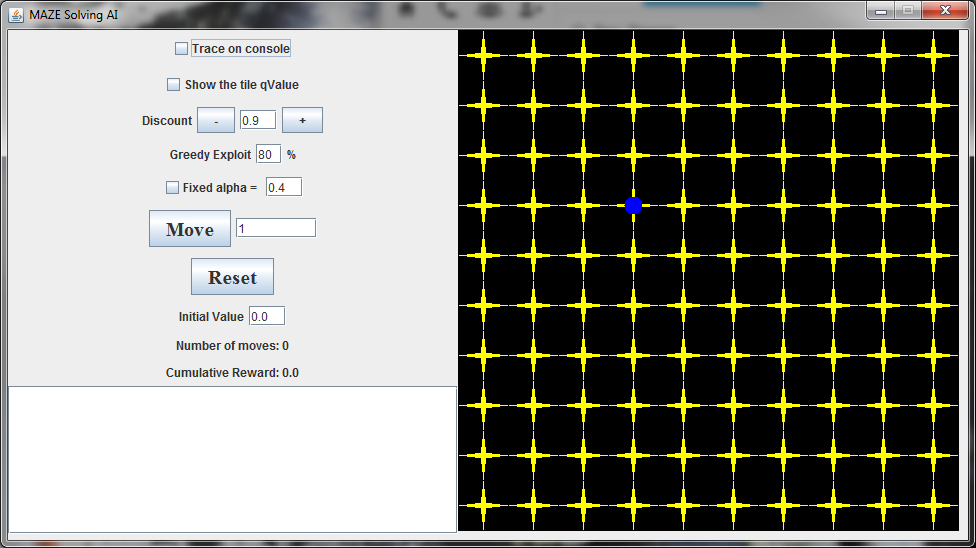
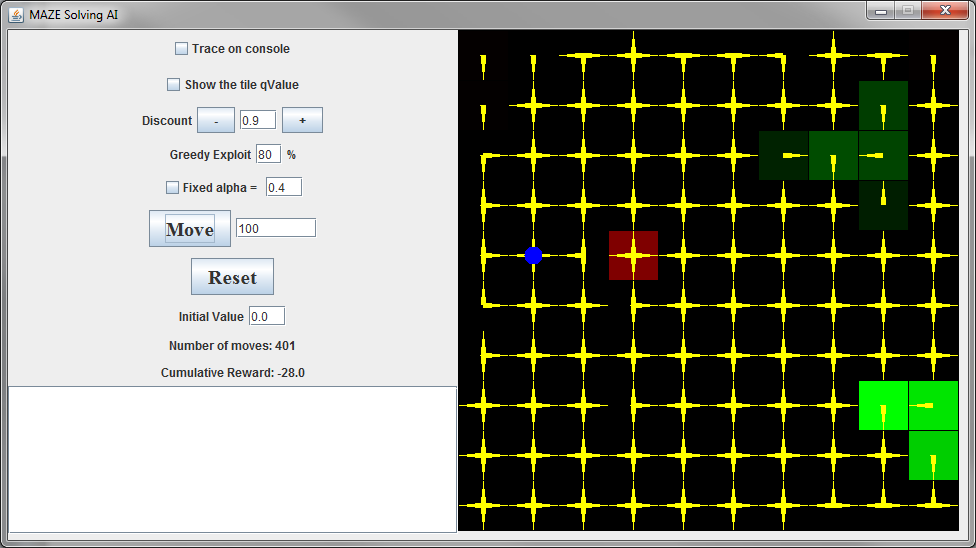


Figure 2

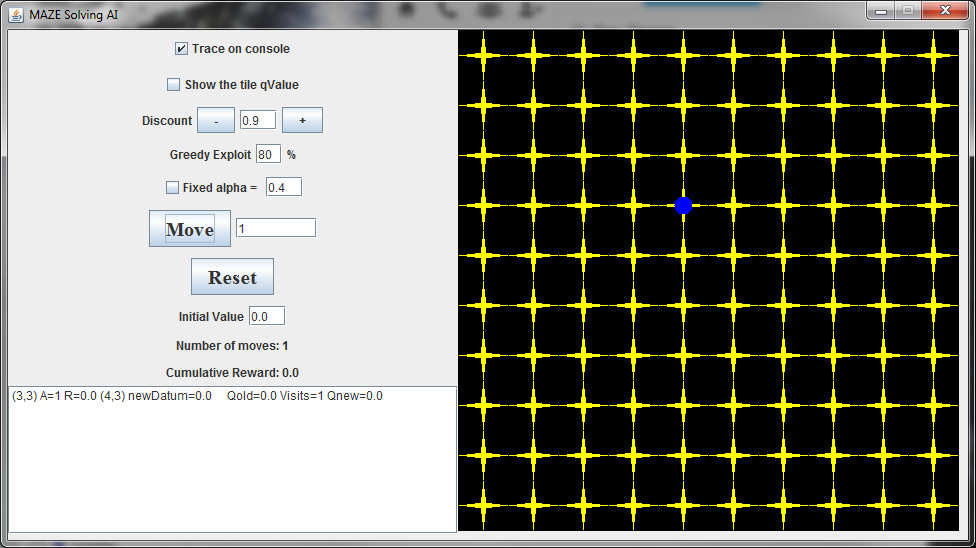
The maze it constructed of a black background that will change colour to show if the tile is good value (green) or bad (red) (Figure 3). As the agent makes more moves, the tiles will gradually become a colour forming a path to the higher value tiles. Another feature of the maze is the four optimal direction arrows that will point in the direction that has the highest q-value.



Figure

## Trace on console

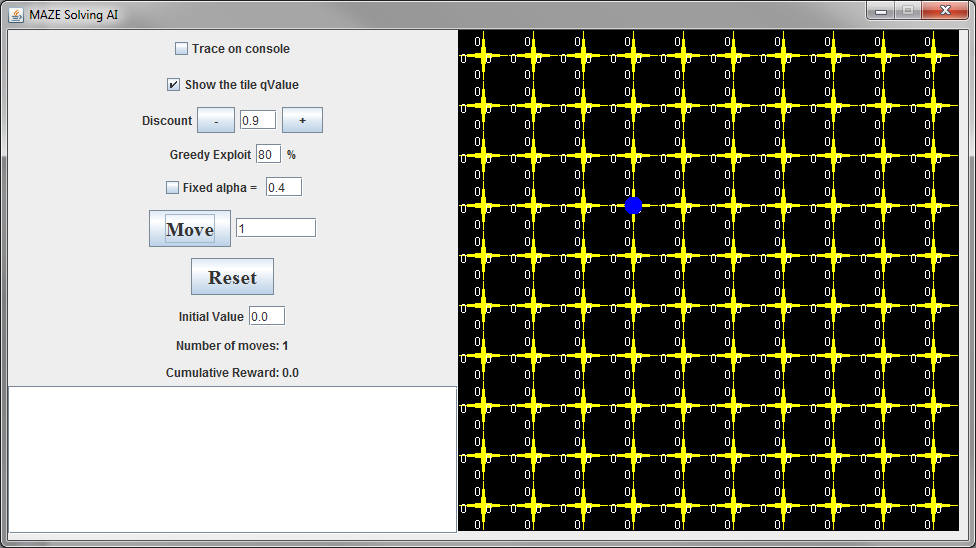
This is a checkbox that will output the learned values of the agent as it is move around the maze. Once selected and the move button is presses it will output those values (Figure 3). The output is the old co-ordinates, the action, the reward, the new co-ordinates and the datum. It will also print out the old Q value, the amount of visits to the tile and the new Q value.



Figure

## Show the tile qValue

This checkbox will show the four q-values that the agent uses to make direction movements. The q-values will appear as white colours numbers alongside the yellow directions arrows. As the maze is explored the values will increase or decrease depending on their value.



Figure

## Discount

The discount factor (gamma) determines the importance of future rewards. A factor of 0 (short sighted) makes the agent consider current rewards and a factor of 1 (long sighted) makes it strive for long-term high rewards. If discount meets or exceeds 1 the action values may diverge.

## Greedy Exploit

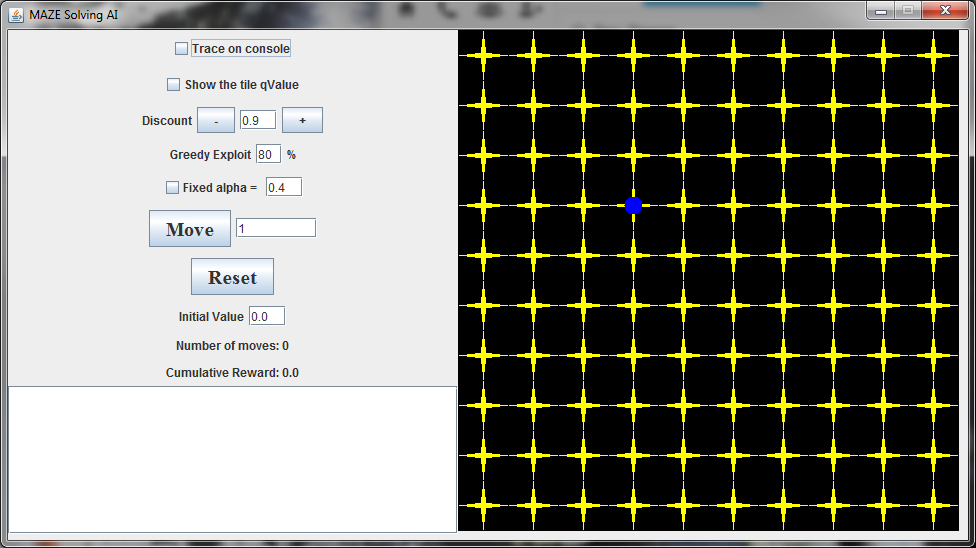
The greedy exploit is for the probability that the agent’s move is greedy. This means that the agent will make moves that will give it the highest cumulative reward. Figure 5 shows the Greedy Exploit as 80% meaning it will make movement decision to the highest reward 80% of the time and randomly move in a direction 20% of the time for exploration purposes.

## Fixed alpha

The learning rate (alpha) determines to what extent the newly acquired information will override the old information. A factor of 0 will make the agent not learn anything. A factor of 1 will make the agent consider only the most recent information.

## Move Button

This is the button that will run the algorithms that will allow the agent to move. When clicked it will move the agent the specified amount of moves in the textbox next to the button. All options in the options panel will not be put into action until the move button is pressed.



Figure

## Reset Button

The reset button resets the Q-values, sets all of the Q-values to initVal, and all of the visit counts to 0.

## Initial Value

InitVal is the initial value to set all the values of the tile directions to. This value can be change to show a different behaviour in the agent decision making.

## Number of Move

Number of moves is a label that will show how many moves the agent has made in total.

## Cumulative Reward

The cumulative reward is the overall reward that the agent has gotten. The agent aims to achieve a higher cumulative reward and makes decisions to increase its value.