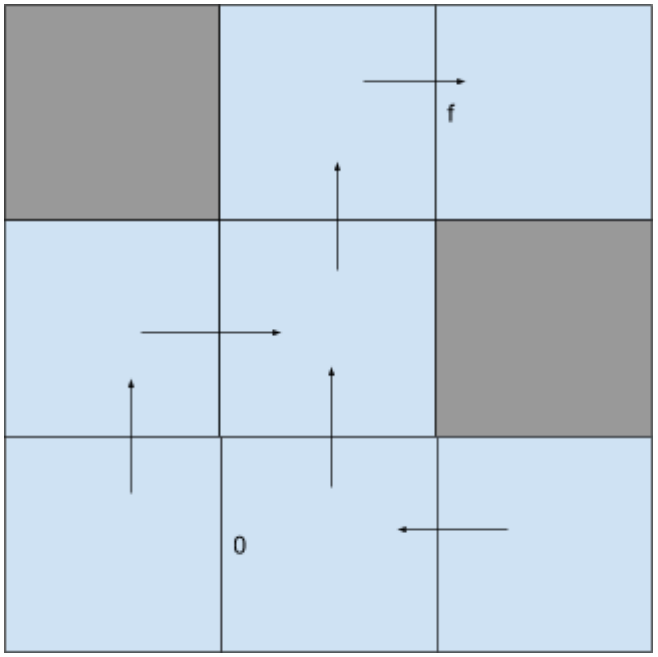
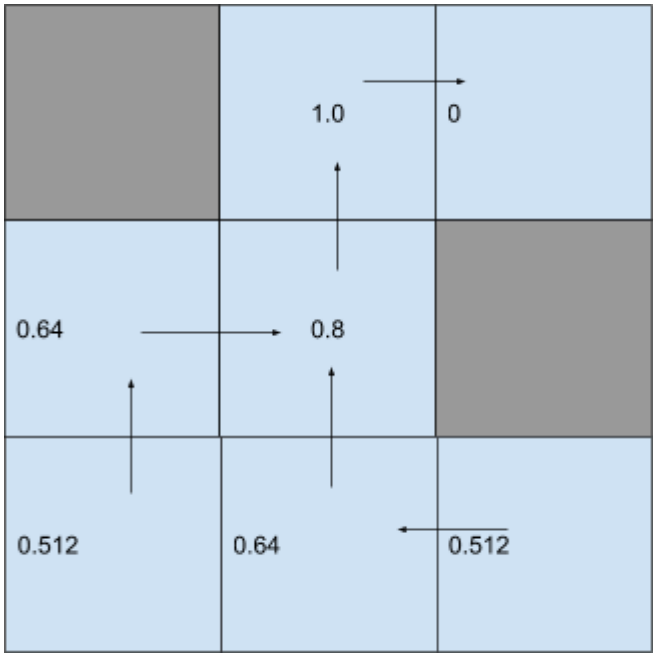
Assignment 2 report
344.057 Artificial Intelligence WS2019/2020

G32
Balázs Varga K11932199, thebalu1@gmail.com
Salvijus Mazeikis K11932182, Salvijus.mazeikis2@mail.dcu.ie

(A) An optimal policy can be seen in the drawing. There are 2 optimal policies, the other differing from this only in the bottom left cell's action. If we go to the right from there, that also results in an optimal policy.

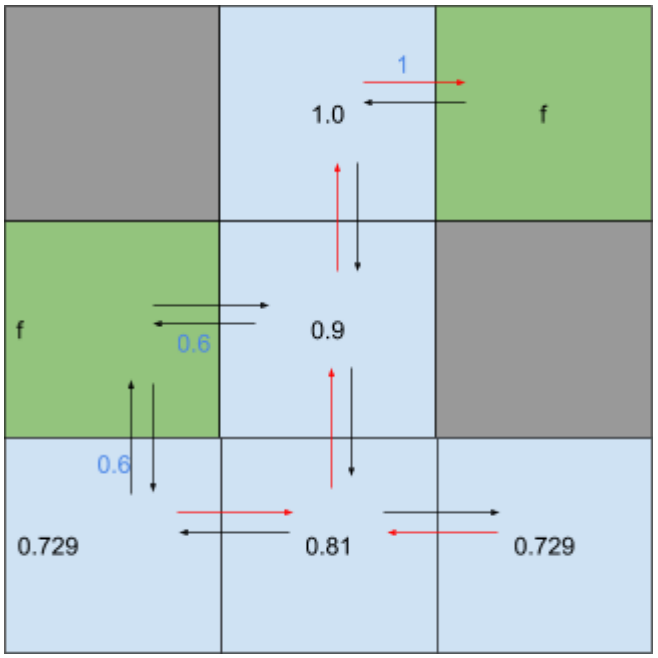


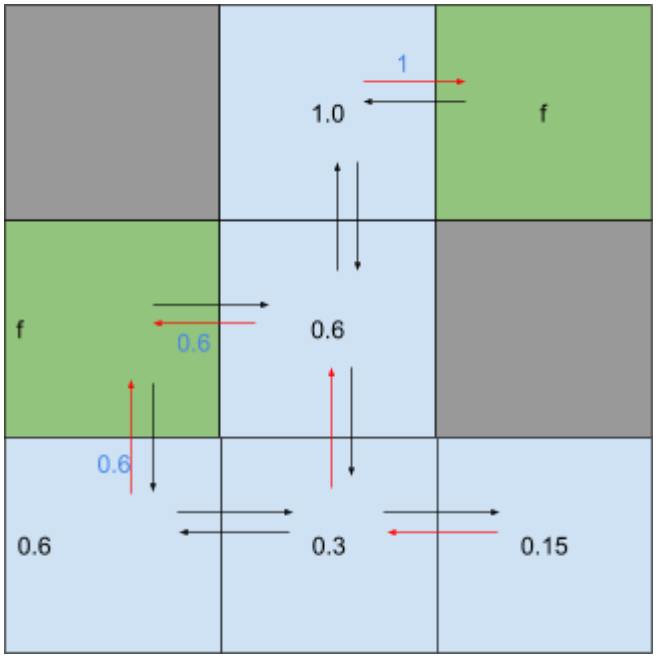
(B) In the case of discount factor = 0, only the top-middle square would keep its value. All others would be 0. In the case of 1, all states from which the goal is reachable would have a V* value of 1. Both of these discount factors would produce a useless result, because no decisions can be made based on the calculated values.



(C)

In the first picture, we have γ = 0.9, in the second γ = 0.5. This results in a different optimal policy, for example for the cell in the center. In the first example, the optimal decision is to go up, in the second, it is to go left.





(D)

EPISODE 1

	LEFT	UP	DOWN	RIGHT
bottom-left		0		0
bottom-mid (0)	0	0		0
bottom-right	0			
mid-left			0	0
mid-mid	0	0	0	

top-mid			0	0
top-right (f)	0			

UP

	LEFT	UP	DOWN	RIGHT
bottom-left		0		0
bottom-mid (0)	0	0		0
bottom-right	0			
mid-left			0	0
mid-mid	0	0	0	
top-mid			0	0
top-right (f)	0			

UP

	LEFT	UP	DOWN	RIGHT
bottom-left		0		0
bottom-mid (0)	0	0		0
bottom-right	0			
mid-left			0	0
mid-mid	0	0	0	
top-mid			0	0
top-right (f)	0			

RIGHT

	LEFT	UP	DOWN	RIGHT
bottom-left		0		0
bottom-mid (0)	0	0		0
bottom-right	0			
mid-left			0	0
mid-mid	0	0	0	
top-mid			0	1.0
top-right (f)	0			

EPISODE 2

UP

	LEFT	UP	DOWN	RIGHT
bottom-left		0		0
bottom-mid (0)	0	0		0
bottom-right	0			
mid-left			0	0
mid-mid	0	0	0	
top-mid			0	1.0
top-right (f)	0			

UP

	LEFT	UP	DOWN	RIGHT
bottom-left		0		0
bottom-mid (0)	0	0		0
bottom-right	0			
mid-left			0	0
mid-mid	0	0.8	0	
top-mid			0	1.0
top-right (f)	0			

DOWN

	LEFT	UP	DOWN	RIGHT
bottom-left		0		0
bottom-mid (0)	0	0		0
bottom-right	0			
mid-left			0	0
mid-mid	0	0.8	0	
top-mid			0.64	1.0
top-right (f)	0			

LEFT

	LEFT	UP	DOWN	RIGHT
bottom-left		0		0

bottom-mid (0)	0	0		0
bottom-right	0			
mid-left			0	0
mid-mid	0	0.8	0	
top-mid			0.64	1.0
top-right (f)	0			

RIGHT

	LEFT	UP	DOWN	RIGHT
bottom-left		0		0
bottom-mid (0)	0	0		0
bottom-right	0			
mid-left			0	0.64
mid-mid	0	0.8	0	
top-mid			0.64	1.0
top-right (f)	0			

DOWN

	LEFT	UP	DOWN	RIGHT
bottom-left		0		0
bottom-mid (0)	0	0		0
bottom-right	0			
mid-left			0	0.64
mid-mid	0	0.8	0	
top-mid			0.64	1.0
top-right (f)	0			

UP

	LEFT	UP	DOWN	RIGHT
bottom-left		0		0
bottom-mid (0)	0	0.64		0
bottom-right	0			
mid-left			0	0.64

mid-mid	0	0.8	0	
top-mid			0.64	1.0
top-right (f)	0			

UP

	LEFT	UP	DOWN	RIGHT
bottom-left		0		0
bottom-mid (0)	0	0.64		0
bottom-right	0			
mid-left			0	0.64
mid-mid	0	0.8	0	
top-mid			0.64	1.0
top-right (f)	0			

RIGHT

	LEFT	UP	DOWN	RIGHT
bottom-left		0		0
bottom-mid (0)	0	0.64		0
bottom-right	0			
mid-left			0	0.64
mid-mid	0	0.8	0	
top-mid			0.64	1.0
top-right (f)	0			

EPISODE 3

LEFT

	LEFT	UP	DOWN	RIGHT
bottom-left		0		0
bottom-mid (0)	0	0.64		0
bottom-right	0			
mid-left			0	0.64
mid-mid	0	0.8	0	
top-mid			0.64	1.0

top-right (f)	0			

RIGHT

	LEFT	UP	DOWN	RIGHT
bottom-left		0		0.512
bottom-mid (0)	0	0.64		0
bottom-right	0			
mid-left			0	0.64
mid-mid	0	0.8	0	
top-mid			0.64	1.0
top-right (f)	0			

LEFT

	LEFT	UP	DOWN	RIGHT
bottom-left		0		0.512
bottom-mid (0)	0.4096	0.64		0
bottom-right	0			
mid-left			0	0.64
mid-mid	0	0.8	0	
top-mid			0.64	1.0
top-right (f)	0			

UP

	LEFT	UP	DOWN	RIGHT
bottom-left		0.512		0.512
bottom-mid (0)	0.4096	0.64		0
bottom-right	0			
mid-left			0	0.64
mid-mid	0	0.8	0	
top-mid			0.64	1.0
top-right (f)	0			

DOWN

	LEFT	UP	DOWN	RIGHT

bottom-left		0.512		0.512
bottom-mid (0)	0.4096	0.64		0
bottom-right	0			
mid-left			0.4096	0.64
mid-mid	0	0.8	0	
top-mid			0.64	1.0
top-right (f)	0			

RIGHT

	LEFT	UP	DOWN	RIGHT
bottom-left		0.512		0.512
bottom-mid (0)	0.4096	0.64		0
bottom-right	0			
mid-left			0.4096	0.64
mid-mid	0	0.8	0	
top-mid			0.64	1.0
top-right (f)	0			

UP

	LEFT	UP	DOWN	RIGHT
bottom-left		0.512		0.512
bottom-mid (0)	0.4096	0.64		0
bottom-right	0			
mid-left			0.4096	0.64
mid-mid	0	0.8	0	
top-mid			0.64	1.0
top-right (f)	0			

UP

	LEFT	UP	DOWN	RIGHT
bottom-left		0.512		0.512
bottom-mid (0)	0.4096	0.64		0
bottom-right	0			

mid-left			0.4096	0.64
mid-mid	0	0.8	0	
top-mid			0.64	1.0
top-right (f)	0			

RIGHT

	LEFT	UP	DOWN	RIGHT
bottom-left		0.512		0.512
bottom-mid (0)	0.4096	0.64		0
bottom-right	0			
mid-left			0.4096	0.64
mid-mid	0	0.8	0	
top-mid			0.64	1.0
top-right (f)	0			

(E)

benefit:

- It allows for exploration of unknown actions, does not get stuck with a suboptimal solution

drawback:

- Slow learning of optimal policy. Even if it is found, the algorithm will keep exploring randomly.

solution:

- Make a compromise: with a certain probability, follow the already learned best path; or choose an action randomly. This way, we might still explore unknown, possibly more optimal paths, but many times we will take the known good path, allowing for a quicker solution.