Single Agent	Explorer and Solver	Multi Agent
	SET agents ← solver (1), explorer (2)	**SET** agents ← solver-start (1), solver-finish (2)
SET epsilon ← epsilon initial value	**SET** epsilon ← epsilon initial value	**SET** epsilon ← epsilon initial value
REPEAT for each episode	**REPEAT** for each episode	**REPEAT** for each episode
SET solved ← False	**SET** solved ← False	**SET** solved ← False
	FOR each agent in agents	**FOR** each agent in agents
RESET agent current state	**RESET** agent current state	**RESET** agent current state
REPEAT for each step **UNTIL** solved or maximum steps are reached.	**REPEAT** for each step **UNTIL** solved or maximum steps are reached.	**REPEAT** for each step **UNTIL** solved or maximum steps are reached.
	FOR each agent in agents	**FOR** each agent in agents
	IF agent is the solver **THEN**	
DETERMINE action A from state S using exploration policy	**DETERMINE** action A_1 from state S_1 using exploration policy	**DETERMINE** action A from state S using exploration policy
	ELSE	
	DETERMINE action A_2 randomly	
	ENDIF	
COMPUTE agent movement according to action A	**COMPUTE** agent movement according to action A	**COMPUTE** agent movement according to action A
IF agent subsequent state S' is the goal state **THEN**	**IF** agent subsequent state S' is the goal state **THEN**	**IF** S'_1 is the same as S_2 or S_2' the same as S_1 **THEN**
SET reward R ← goal reward and solved ← True	**SET** reward R ← goal reward	**SET** reward R ← goal reward and solved ← True
	IF agent is the solver **THEN**	
	SET solved ← True	
	ENDIF	
ELIF agent subsequent state S' is the same as current state S **THEN**	**ELIF** agent subsequent state S' is the same as current state S **THEN**	**ELIF** agent subsequent state S' is the same as current state S **THEN**
SET reward R← off grid penalty	**SET** reward ← off grid penalty	**SET** reward ← off grid penalty
ELSE	**ELSE**	**ELSE**
SET reward R ← move penalty	**SET** reward ← move penalty	**SET** reward ← move penalty
ENDIF	**ENDIF**	**ENDIF**
IF reward is goal reward **THEN**	**IF** reward is goal reward **THEN**	**IF** reward is goal reward **THEN**
SET Q(S,A) ← goal reward	**SET** Q_1(S,A) ← goal reward	**SET** Q1(S1,A1), Q2(S2,A2) ← goal reward
ELSE	**ELSE**	**ELSE**
SET Q(S,A) ← Q(S,A) + α[R + γ $\max_a$ Q(S',a) - Q(S,A)]	**SET** Q_1(S,A) ← Q_1(S,A) + α[R + γ $\max_a$ Q_1(S',a) - Q_1(S,A)]	**SET** Q_1(S,A) ← Q_1(S,A) + α[R + γ $\max_a$ Q_1(S',a) - Q_1(S,A)]
ENDIF	**ENDIF**	**ENDIF**
	ENDFOR	**ENDFOR**
SET S ← S'	**SET** S ← S'	**SET** S ← S'
SET epsilon ← epsilon * epsilon decay	**SET** epsilon ← epsilon * epsilon decay	**SET** epsilon ← epsilon * epsilon decay