(earning) Agent will learn now using Q-Learning in the Public Civility Game.

Learning Process started. Will finish when Episode = 15000

Policy with the weights: [1.0, 0.31, 0.013]
Expected value is: [7.0, 0.0, -3.375]

The Learnt Policy has the following Value for alpha = 0.05 is:
[1.53054195 -2.35091535 -2.76690103]

The Learnt Policy has the following Value for alpha = 0.1 is:
[7.79691574 -5.73111446 -2.51638429]

The Learnt Policy has the following Value for alpha = 0.15000000000000002 is:
[5.82987613 -3.71064028 -1.98546862]

The Learnt Policy has the following Value for alpha = 0.2 is:
[6.62344396 -3.64613105 -5.76251303]

The Learnt Policy has the following Value for alpha = 0.25 is:
[6.30765716 -1.75468716 -2.69952199]

The Learnt Policy has the following Value for alpha = 0.3 is:
[5.93987805 -0.98120438 -6.19280676]

The Learnt Policy has the following Value for alpha = 0.35000000000000003 is:
[6.81104278 -0.57074722 -2.26420377]

The Learnt Policy has the following Value for alpha = 0.4 is:
[7.33007967 -1.66822698 -3.25778544]

The Learnt Policy has the following Value for alpha = 0.45 is:
[6.99836394e+00 -7.31296001e-08 -2.86347825e+00]

The Learnt Policy has the following Value for alpha = 0.5 is:
[6.98800165e+000 -2.83943608e-182 -3.30230459e+000]
This alpha works!

The Learnt Policy has the following Value for alpha = 0.55 is:
[6.99937097e+00 -3.42968167e-91 -4.96501285e+00]

The Learnt Policy has the following Value for alpha = 0.6000000000000001 is:
[6.99990624e+00 -2.05054755e-64 -4.40707875e+00]

The Learnt Policy has the following Value for alpha = 0.6500000000000001 is:
[6.99999577e+000 -3.30949433e-126 -1.28039720e+000]

Policy with the weights: [1.0, 0.372, 1.5]
Expected value is: [6.03125, 0.0, 0.0]

The Learnt Policy has the following Value for alpha = 0.05 is:
[0.11200033 -0.29715442 0.   ]

The Learnt Policy has the following Value for alpha = 0.1 is:
[3.37863411e+00 -1.75840575e+00 -5.22102813e-17]

The Learnt Policy has the following Value for alpha = 0.15000000000000002 is:
[4.69044531 -2.73749467 0.   ]

The Learnt Policy has the following Value for alpha = 0.2 is:
[4.10521079e+00 -1.82381702e+00 -1.98538033e-13]

The Learnt Policy has the following Value for alpha = 0.25 is:
[4.30336375e+00 -7.48614315e-01 -5.42149136e-11]

The Learnt Policy has the following Value for alpha = 0.3 is:
[5.76307362e+00 -8.20036116e-01 -9.29875740e-06]

The Learnt Policy has the following Value for alpha = 0.35000000000000003 is:
[5.89410134e+00 -3.63362387e-01 -3.68068865e-04]

The Learnt Policy has the following Value for alpha = 0.4 is:
[6.38082107e+00 -4.62371156e-02 -9.48352171e-10]

The Learnt Policy has the following Value for alpha = 0.45 is:
[5.71246045e+00 -6.41807423e-01 -5.84265382e-18]

The Learnt Policy has the following Value for alpha = 0.5 is:
[5.41669979e+00 -7.15797114e-08 -2.60277116e-16]

The Learnt Policy has the following Value for alpha = 0.55 is:
[6.35579733e+00 -8.41698248e-07 -7.83588298e-07]

The Learnt Policy has the following Value for alpha = 0.6000000000000001 is:
[6.31089453e+00 -4.35108575e-07 -4.47101042e-07]

The Learnt Policy has the following Value for alpha = 0.6500000000000001 is:
[5.78176252e+00 -2.60680881e-38 -4.00486492e-38]

Policy with the weights: [1.0, 0.09, 1e-09]
Expected value is: [10.0, -20.0, -4.0]

The Learnt Policy has the following Value for alpha = 0.05 is:
[10.   -20.05983427 -4.01695478]
This alpha works!

The Learnt Policy has the following Value for alpha = 0.1 is:
[10.   -20.43310326 -4.31662349]

The Learnt Policy has the following Value for alpha = 0.15000000000000002 is:
[9.72012416 -17.27387425 -3.89978591]

The Learnt Policy has the following Value for alpha = 0.2 is:
[10.   -20.00398073 -4.05234347]
This alpha works!

The Learnt Policy has the following Value for alpha = 0.25 is:
[9.86501602 -20.00003928 -3.04329647]

The Learnt Policy has the following Value for alpha = 0.3 is:
[10.   -20.   -2.97022817]

The Learnt Policy has the following Value for alpha = 0.35000000000000003 is:
[8.71534162 -15.51475152 -4.08601386]

```
-------------------
The Learnt Policy has the following Value for alpha =  0.4  is:
[ 9.        -10.95627453 -5.24861079]
-------------------
The Learnt Policy has the following Value for alpha =  0.45  is:
[ 9.       -10.        -0.12567905]
-------------------
The Learnt Policy has the following Value for alpha =  0.5  is:
[ 9.       -10.        -0.49866495]
-------------------
The Learnt Policy has the following Value for alpha =  0.55  is:
[ 9.       -10.        -0.80501205]
-------------------
The Learnt Policy has the following Value for alpha =  0.6000000000000001  is:
[ 9.       -10.        -2.51997252]
-------------------
The Learnt Policy has the following Value for alpha =  0.6500000000000001  is:
[ 9.       -10.        -1.02882771]
-----


-----
Policy with the weights:  [1.0, 0.045, 0.233]
Expected value is:  [10.0, -30.0, -0.75]
-------------------
The Learnt Policy has the following Value for alpha =  0.05  is:
[ 10.     -30.      -1.733195]
-------------------
The Learnt Policy has the following Value for alpha =  0.1  is:
[ 9.50516452 -24.69766479  -0.30526365]
-------------------
The Learnt Policy has the following Value for alpha =  0.15000000000000002  is:
[ 9.93848069 -30.        -0.07029999]
-------------------
The Learnt Policy has the following Value for alpha =  0.2  is:
[ 9.00857358 -18.32899721  -0.70143732]
-------------------
The Learnt Policy has the following Value for alpha =  0.25  is:
[ 9.83144093 -28.31440935  -0.866243  ]
-------------------
The Learnt Policy has the following Value for alpha =  0.3  is:
[ 9.72464801 -24.97445356  -0.62305979]
-------------------
The Learnt Policy has the following Value for alpha =  0.35000000000000003  is:
[ 8.91486684 -10.        -0.01599989]
-------------------
The Learnt Policy has the following Value for alpha =  0.4  is:
[ 10.     -30.      -0.84030401]
This alpha works!
-------------------
The Learnt Policy has the following Value for alpha =  0.45  is:
[ 9.1971358  -15.80005648  -1.11107463]
-------------------
The Learnt Policy has the following Value for alpha =  0.5  is:
[ 8.99996792 -10.001297   -0.02230426]
-------------------
The Learnt Policy has the following Value for alpha =  0.55  is:
[ 8.99999961 -10.00000035  -0.90916897]
-------------------
The Learnt Policy has the following Value for alpha =  0.6000000000000001  is:
[ 9.64963546 -22.99270922  -1.94890638]
```

The Learnt Policy has the following Value for alpha = 0.6500000000000001 is:
[9.   -10.00000003 -0.25790476]

Policy with the weights: [1.0, 0.372, 1.5]
Expected value is: [6.03125, 0.0, 0.0]

The Learnt Policy has the following Value for alpha = 0.05 is:
[1.21690928e-01 -2.37051393e-03 -3.36108596e-12]

The Learnt Policy has the following Value for alpha = 0.1 is:
[4.16738889e+00 -1.49817959e+00 -9.72844338e-20]

The Learnt Policy has the following Value for alpha = 0.15000000000000002 is:
[3.71743332 -1.40252857 0.   ]

The Learnt Policy has the following Value for alpha = 0.2 is:
[6.04866751 -3.68271647 0.   ]

The Learnt Policy has the following Value for alpha = 0.25 is:
[4.34252289e+00 -3.40332903e-01 -2.64993356e-27]

The Learnt Policy has the following Value for alpha = 0.3 is:
[6.44336755e+00 -2.06457993e+00 -5.23893869e-25]

The Learnt Policy has the following Value for alpha = 0.35000000000000003 is:
[6.01357644e+00 -9.58724969e-01 -3.14880020e-07]

The Learnt Policy has the following Value for alpha = 0.4 is:
[5.44217510e+00 -2.51352611e-01 -2.92164441e-09]

The Learnt Policy has the following Value for alpha = 0.45 is:
[5.84235433e+00 -6.31926073e-01 -1.58721913e-11]

The Learnt Policy has the following Value for alpha = 0.5 is:
[6.74651947e+00 -4.04292279e-16 -9.19791197e-08]

The Learnt Policy has the following Value for alpha = 0.55 is:
[5.29239423e+00 -7.92267735e-19 -9.29471975e-15]

The Learnt Policy has the following Value for alpha = 0.6000000000000001 is:
[6.06018479e+00 -6.59706977e-08 -6.40200434e-12]
This alpha works!

The Learnt Policy has the following Value for alpha = 0.6500000000000001 is:
[6.44548175e+00 -2.36849222e-20 -1.30040015e-20]

Policy with the weights: [1.0, 0.005, 0.633]
Expected value is: [9.625, -30.0, 0.0]

The Learnt Policy has the following Value for alpha = 0.05 is:
[9.73568715 -29.9965614   -0.36950225]

The Learnt Policy has the following Value for alpha = 0.1 is:
[9.65994073 -30.   -0.44695708]

The Learnt Policy has the following Value for alpha = 0.15000000000000002 is:
[10.   -30.   -1.09787589]

The Learnt Policy has the following Value for alpha = 0.2 is:
[9.78459387e+00 -3.00000000e+01 -5.15097054e-79]

The Learnt Policy has the following Value for alpha = 0.25 is:
[9.55868497e+00 -3.00000000e+01 -4.47826508e-47]
This alpha works!

The Learnt Policy has the following Value for alpha = 0.3 is:
[9.52639122e+00 -2.99999985e+01 -1.08712917e-72]
This alpha works!

The Learnt Policy has the following Value for alpha = 0.35000000000000003 is:
[9.39782492e+00 -2.92497401e+01 -1.31484538e-06]

The Learnt Policy has the following Value for alpha = 0.4 is:
[9.73412726e+00 -3.00000000e+01 -1.30380490e-21]

The Learnt Policy has the following Value for alpha = 0.45 is:
[9.71418150e+00 -3.00000000e+01 -1.02767208e-50]
This alpha works!

The Learnt Policy has the following Value for alpha = 0.5 is:
[8.99468725e+00 -1.01948725e+01 -3.84573098e-13]

The Learnt Policy has the following Value for alpha = 0.55 is:
[8.97614757e+00 -1.30061497e+01 -2.05660639e-18]

The Learnt Policy has the following Value for alpha = 0.6000000000000001 is:
[8.57554729e+00 -1.47451461e+01 -2.84687380e-03]

The Learnt Policy has the following Value for alpha = 0.6500000000000001 is:
[9.89596067e+00 -3.00000000e+01 -5.64795339e-17]

Finnished!!!
[[0.5], [], [0.05, 0.2], [0.4], [0.6000000000000001], [0.25, 0.3, 0.45]]

Process finished with exit code 0