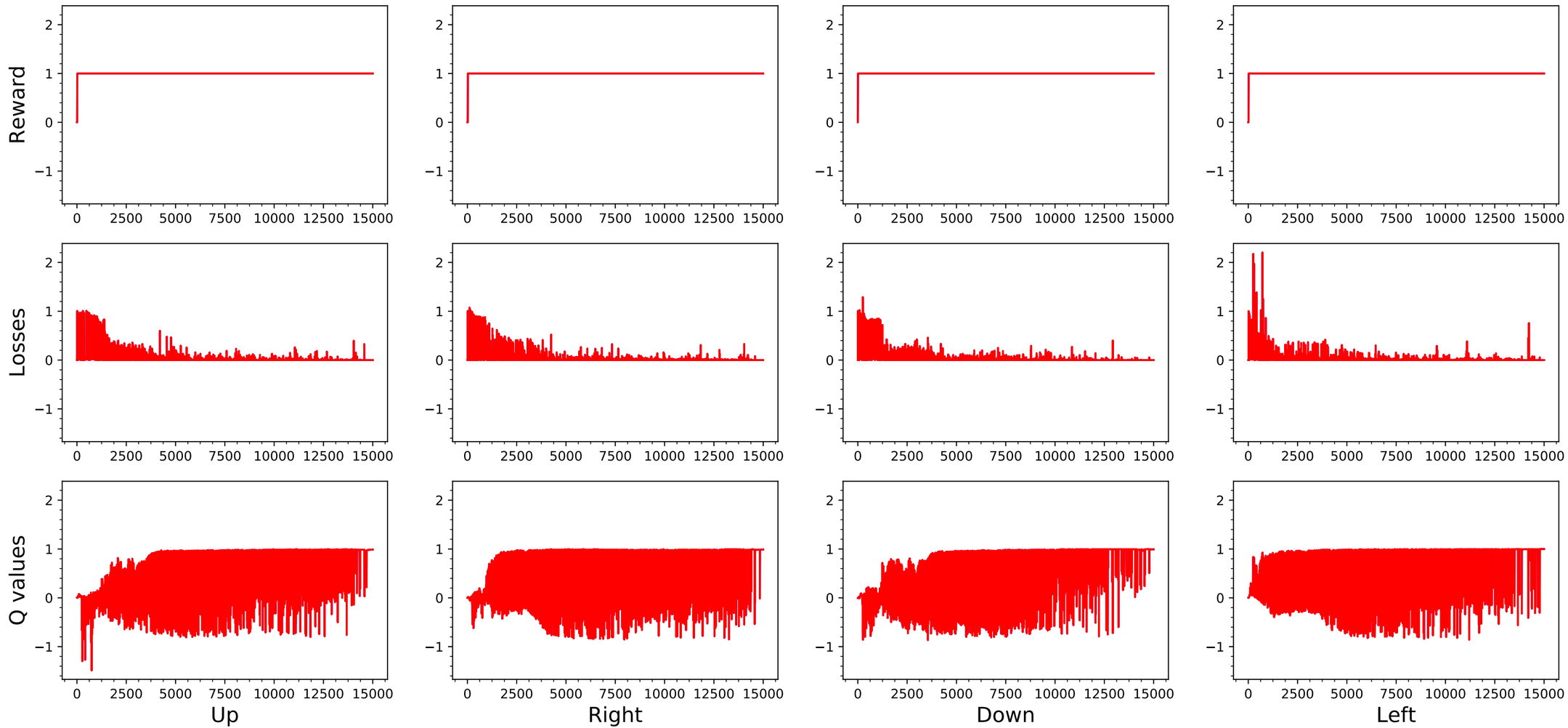
Reward, Qval, Loss of actions over epochs. Blue: train, Red: test
Trial info: lr:0.001, mna:2, nepochs:15000, nsamples:1. Epsilon annealed linearly.
Reward
Losses
Q values
Up
Right
Down
Left