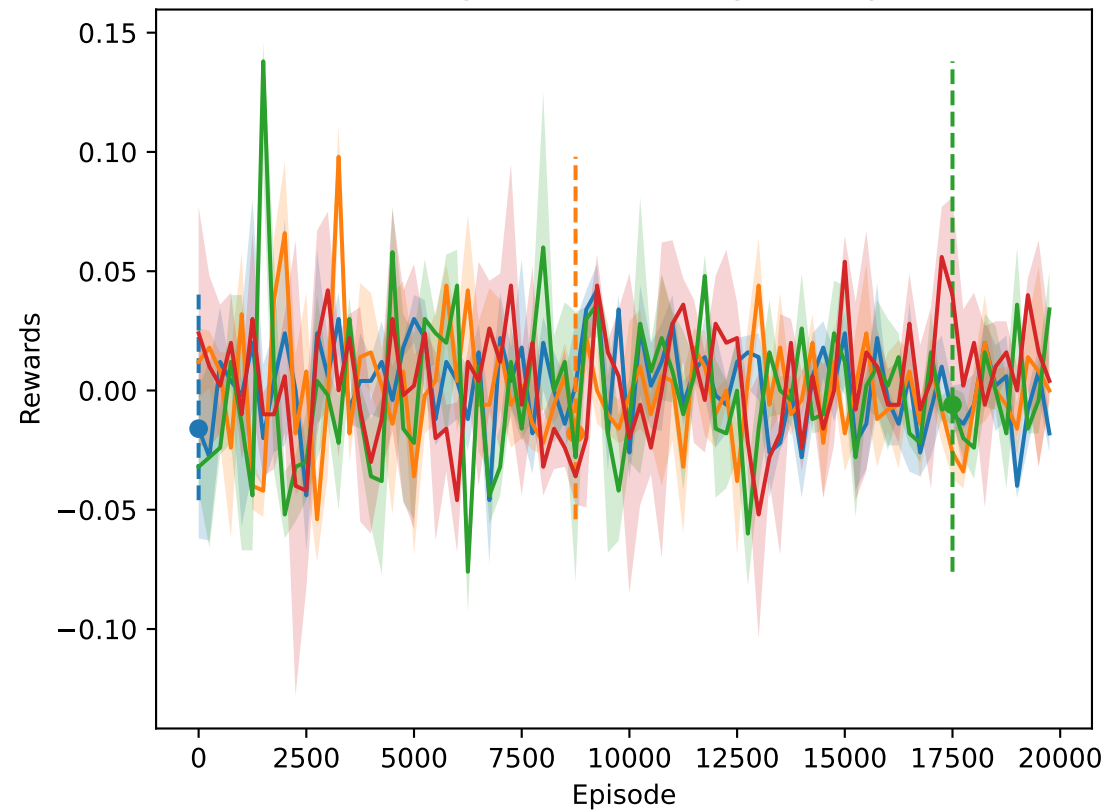
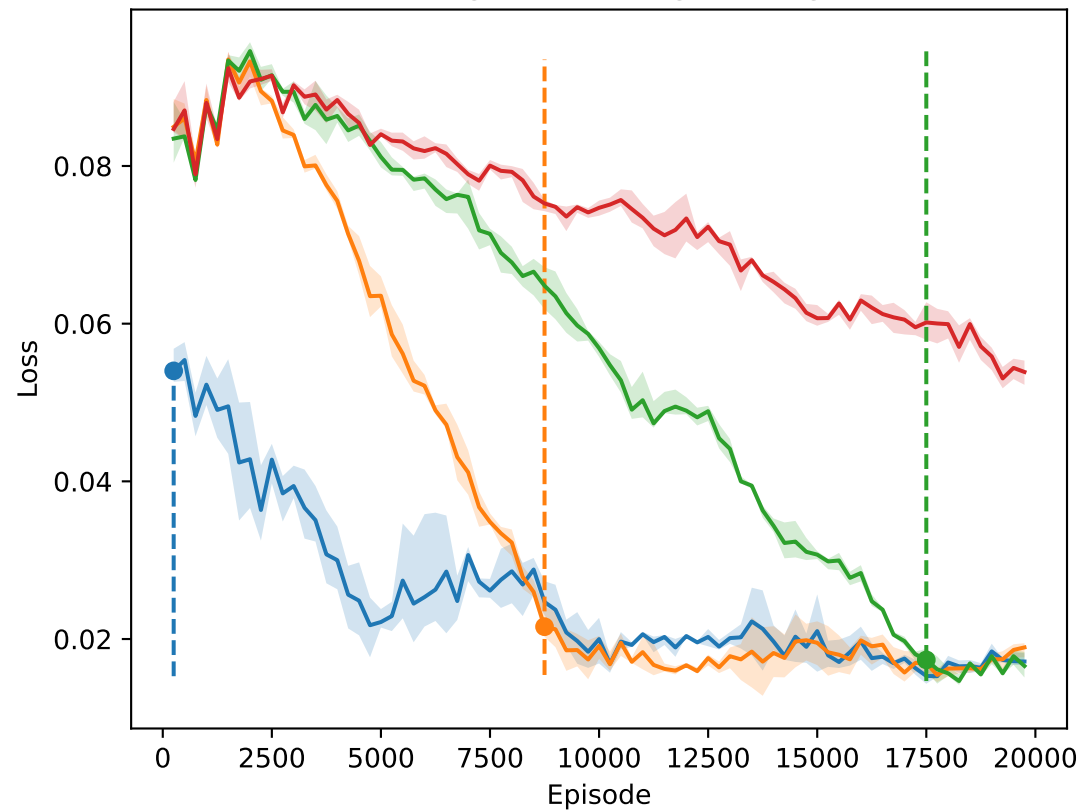


Average rewards during training



Average loss during training



$n^* = 1.0$ $n^* = 10000.0$ $n^* = 20000.0$ $n^* = 40000.0$