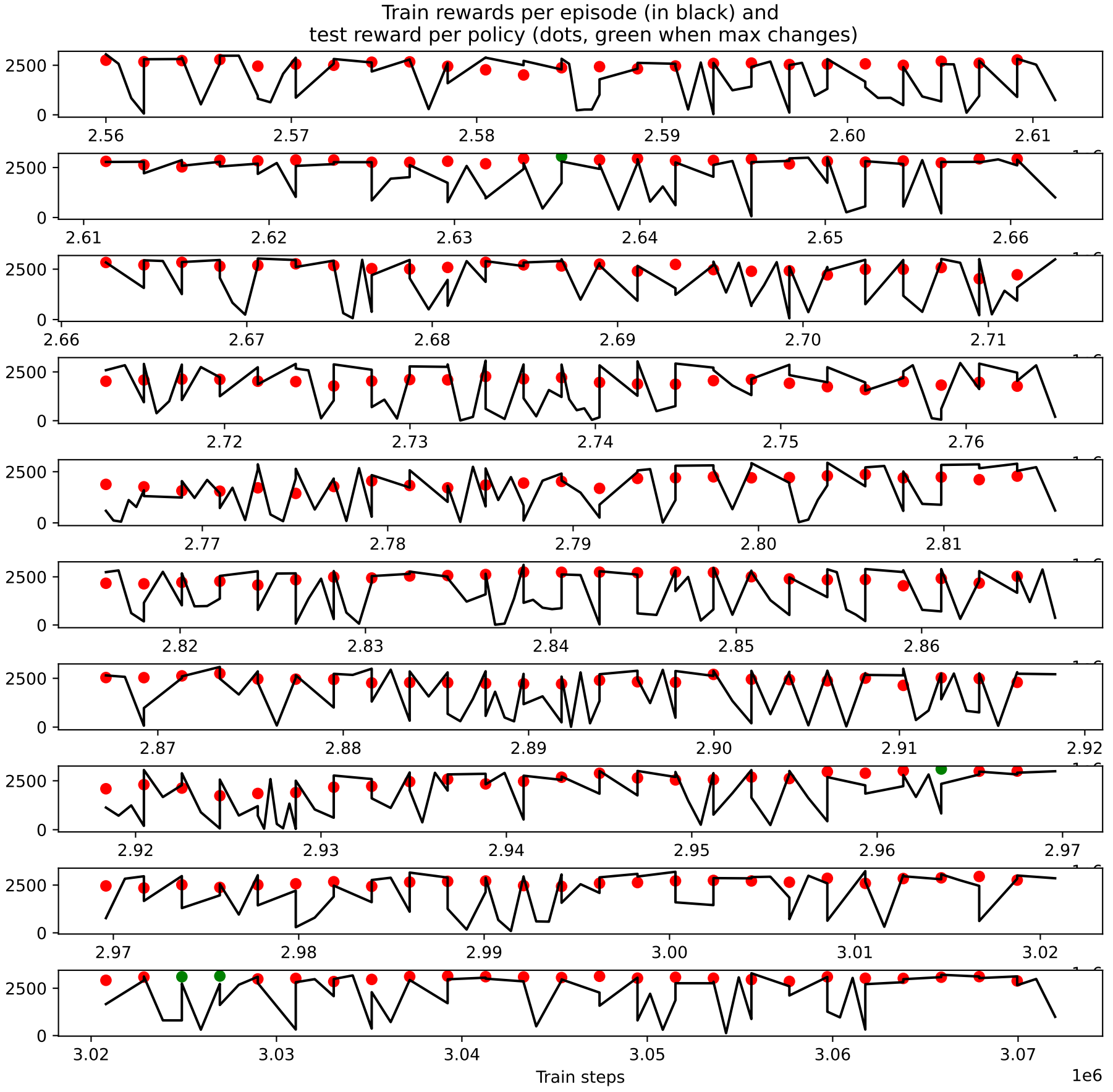
Train rewards per episode (in black) and
test reward per policy (dots, green when max changes)
2500
0
2.56
2.57
2.58
2.59
2.60
2.61
2.62
2.63
2.64
2.65
2.66
2.67
2.68
2.69
2.70
2.71
2.72
2.73
2.74
2.75
2.76
2.77
2.78
2.79
2.80
2.81
2.82
2.83
2.84
2.85
2.86
2.87
2.88
2.89
2.90
2.91
2.92
2.93
2.94
2.95
2.96
2.97
2.98
2.99
3.00
3.01
3.02
3.03
3.04
3.05
3.06
3.07
Train steps
1e6