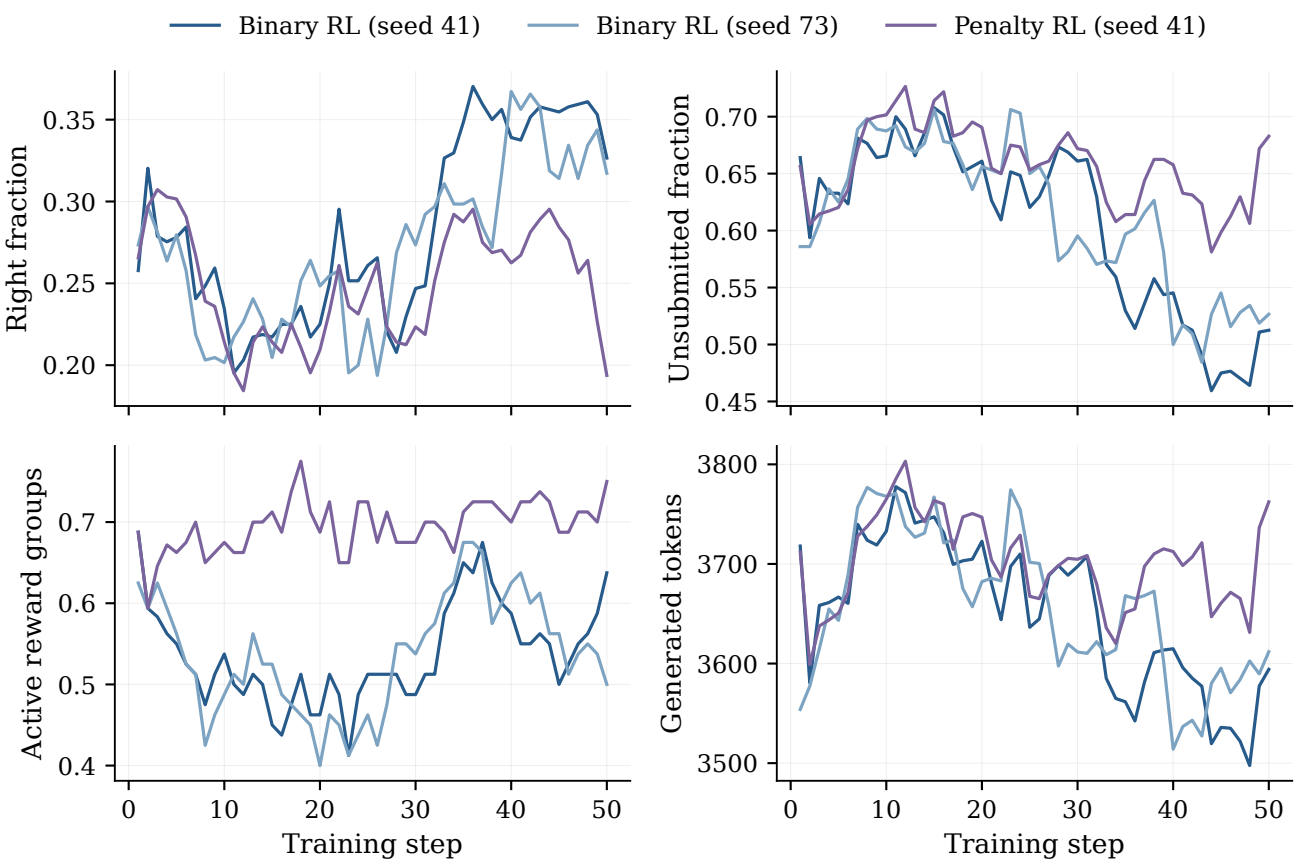

Binary RL (seed 41)
Binary RL (seed 73)
Penalty RL (seed 41)
Right fraction
0.35
0.30
0.25
0.20
Unsubmitted fraction
0.70
0.65
0.60
0.55
0.50
0.45
Active reward groups
0.7
0.6
0.5
0.4
Generated tokens
3800
3700
3600
3500
0
10
20
30
40
50
Training step
Training step