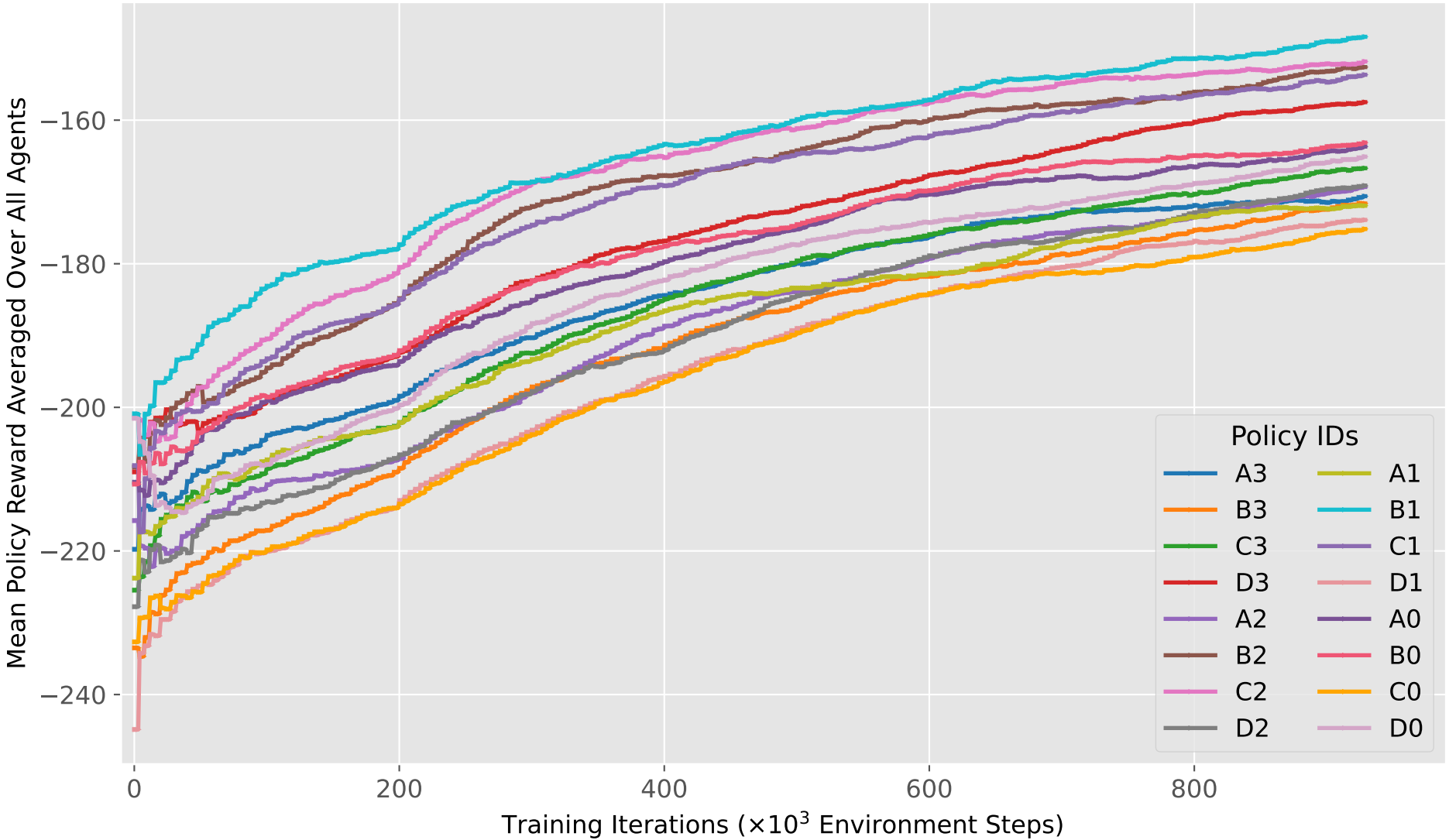
Mean Policy Reward Averaged Over All Agents
−160
−180
−200
−220
−240
0
200
400
600
800
Training Iterations (×10^3 Environment Steps)
Policy IDs
A3
B3
C3
D3
A2
B2
C2
D2
A1
B1
C1
D1
A0
B0
C0
D0