

Mean Policy Reward Averaged Over All Agents

-160  
-180  
-200  
-220  
-240

0

200

400

600

800

Training Iterations ( $\times 10^3$  Environment Steps)

Policy IDs

A3	A1
B3	B1
C3	C1
D3	D1
A2	A0
B2	B0
C2	C0
D2	D0