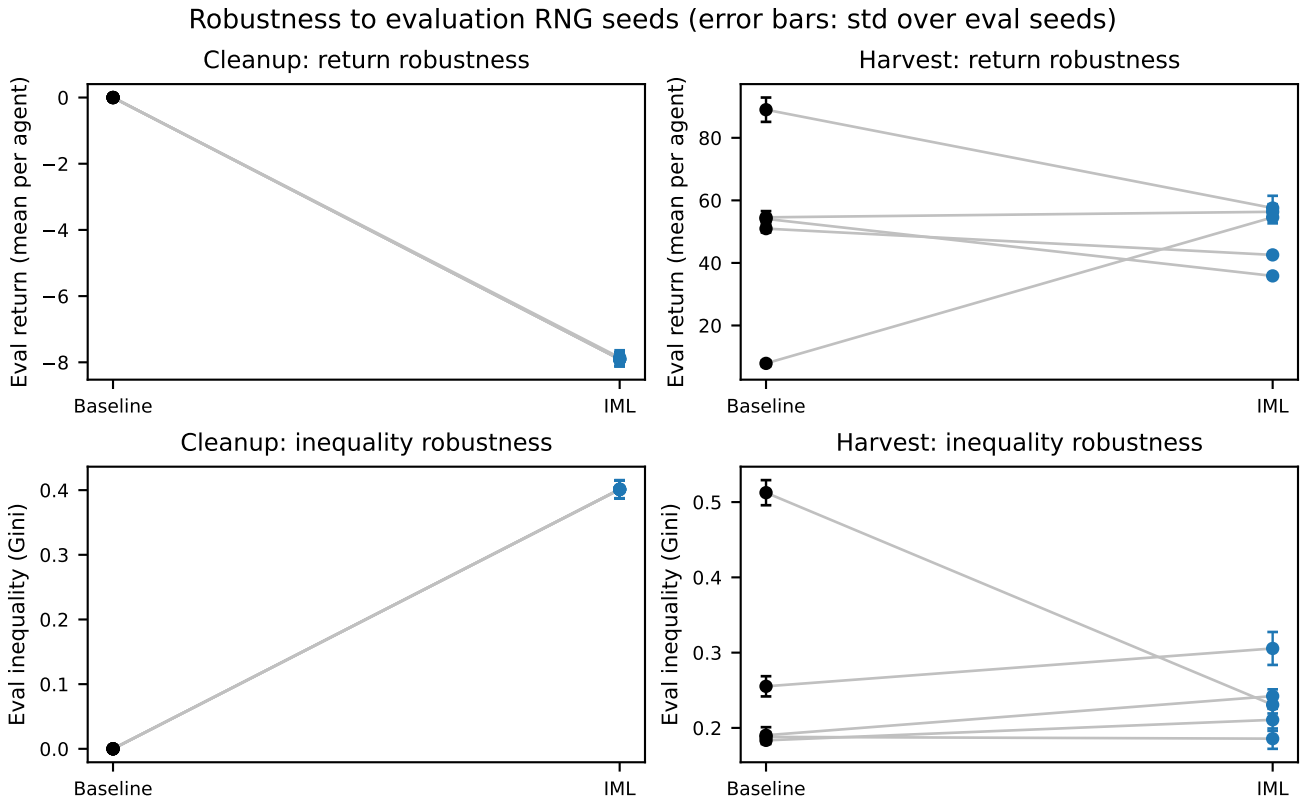
Robustness to evaluation RNG seeds (error bars: std over eval seeds)
Cleanup: return robustness
Eval return (mean per agent)
0
−2
−4
−6
−8
Baseline
IML
Harvest: return robustness
Eval return (mean per agent)
80
60
40
20
Baseline
IML
Cleanup: inequality robustness
Eval inequality (Gini)
0.4
0.3
0.2
0.1
0.0
Baseline
IML
Harvest: inequality robustness
Eval inequality (Gini)
0.5
0.4
0.3
0.2
Baseline
IML