

TRPO guided-learned medium Hanabi policy

mean=2.4, stddev=1.96

