

TRPO guided-learned Hanabi policy

mean=2.6, stddev=1.114

