

TRPO guided-learned mini Hanabi policy

mean=1.6, stddev=0.6633

