

TRPO self-learned mini Hanabi policy

mean=6.955, stddev=1.313

