

TRPO self-learned mini Hanabi policy

mean=1.869, stddev=1.466

