

TRPO dynamic self-learned mini Hanabi policy

mean=1.1, stddev=1.136

