

TRPO dynamic self-learned medium Hanabi policy

mean=1.6, stddev=1.428

