

TRPO dynamic self-learned Hanabi policy

mean=1.5, stddev=1.5

