TRPO static self-learned medium Hanabi policy mean=2.4, stddev=1.4283.0 2.5 Frequency 0.2 2.0 1.0 0.5 0.0 6 8 10 12 14 Score