Simple things destroy RL
Examples from [Henderson et al. 2019]

- Using different random seeds:
  - 2 Sets of 5 seeds for exact same RL-Setup
- Results in statistically significant different distributions
- Avoidance:
  - [Colars et al. 2022] proposes to use either 10 or more seeds and mean or median + STD or IPR
  - or if plot every run as well as mean or median. In addition use STD or CI

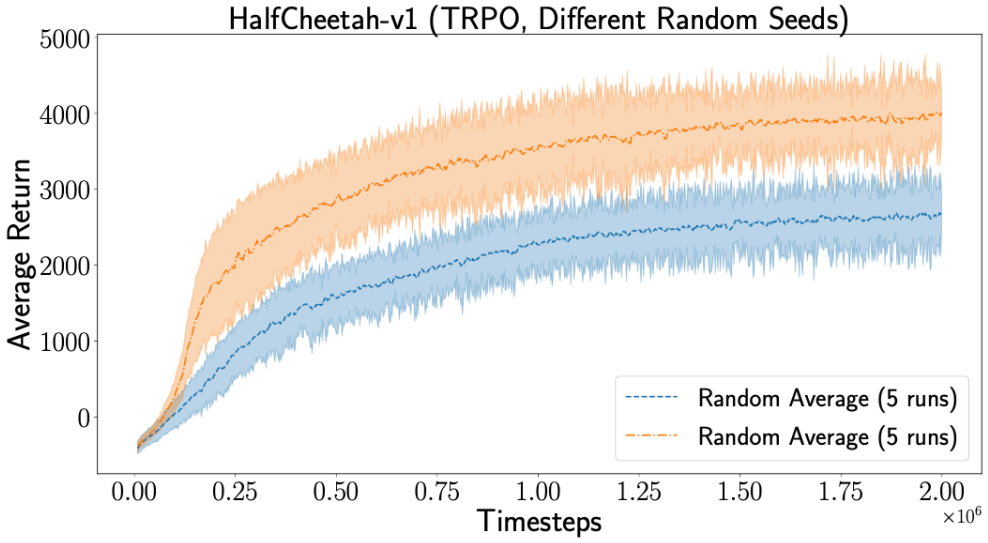

Figure 5: TRPO on HalfCheetah-v1 using the same hyperparameter configurations averaged over two sets of 5 different random seeds each. The average 2-sample t-test across entire training distribution resulted in $t = -9.0916$, $p = 0.0016$.

Simple things destroy RL
Examples from [Henderson et al. 2019]

- Using different code bases
  - TRPO and DDPG for 5 seeds each with code bases:
    - Original TRPO Code [Schulman et al. 2015]
    - Rllab [Duan et al. 2016]
    - OpenAI baselines [Schulman et al. 2017]
  - Different Code for DDPG
- Avoidance:
  - Use stable-baselines 😁

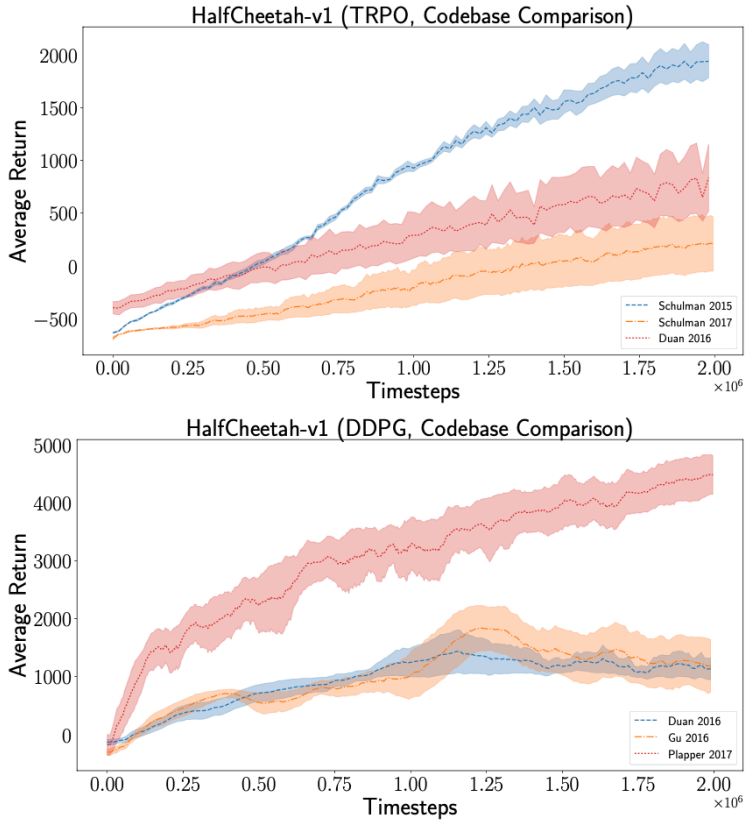

Figure 6: TRPO codebase comparison using our default set of hyperparameters (as used in other experiments).