Table 1: PPO Hyperparameter values used across Breakout, HalfCheetah, Humanoid and MountainCar.

Hyperparameter	**Breakout**	**HalfCheetah**	**Humanoid**	**MountainCar**	**Hopper**
lambda	0.95	0.95	1.0	1.0	1.0
gamma	0.99	0.99	0.955	0.99	0.99
kl coeff	0.5	1.0	1.0	0.2	1.0
clip rewards	True	False	False	False	False
clip param	0.1	0.2	0.3	0.3	0.3
vf clip param	10.0	10.0	10.0	10.0	10.0
vf loss coeff	1.0	0.5	1.0	1.0	1.0
entropy coeff	0.01	0	0	0	0
num sgd iter	10	32	20	30	20
sgd minibatch size	500	4096	32768	128	32678
sample batch size	100	200	200	200	200
train batch size	5000	65536	320000	4000	160000
model - free log std	False	False	True	False	False
use gae	True	True	False	True	True
batch mode	truncate	truncate	complete	truncate	complete
vf share layers	True	False	False	False	False
observation filter	NoFilter	MeanStdFilter	MeanStdFilter	NoFilter	MeanStdFilter
lr	[1e-2, 5e-3, 1e-3, 5e-4, 1e-4, 5e-5, 1e-5, 5e-6]	[1e-3, 5e-4, 1e-4, 5e-5]	[1e-3, 5e-4, 1e-4, 5e-5]	[1e-2, 5e-3, 1e-3, 5e-4, 1e-4, 5e-5]	[1e-3, 5e-4, 1e-4, 5e-5]

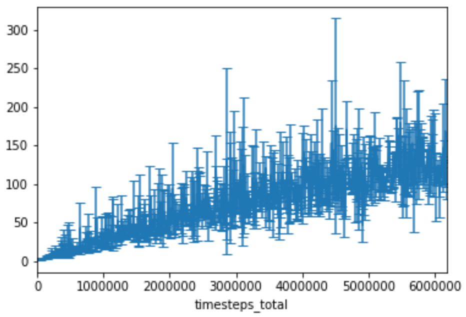

Figure 1: FedPBT on Breakout

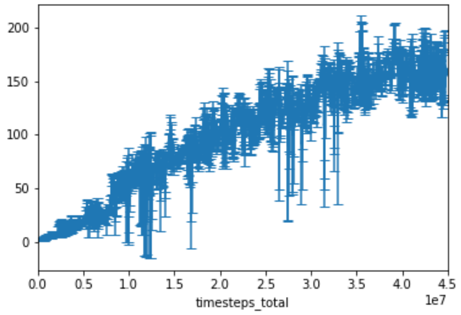

Figure 2: PBT on Breakout

Learning curves plotting the mean (across hyperparameter configurations) with error bars of smoothed episode reward for all results follow. When multiple hyperparameter configurations significantly differ (i.e. learning rates for gridsearch) we omit the worst-performer after some time for visual clarity. The best trial for PBT is shown with cumulative timesteps, whereas all trials for gridsearch are shown with individual timesteps. The perturbation interval for MountainCar and HalfCheetah is 4 training iterations, which works well with a β of $1-2$. The perturbation interval for Humanoid, Hopper and Breakout is 4 training iterations, which matches a β of $3-4$. Full implementation of FedPBT is available on GitHub. Experiments were performed on a g3.16xlarge instance (64 CPU, 4 GPU).

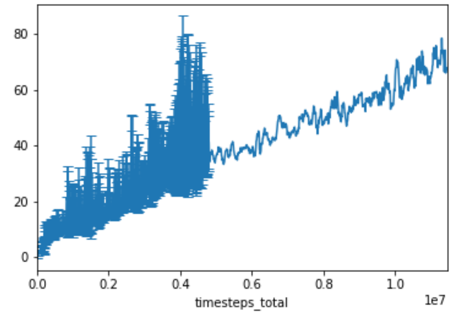

Figure 3: Gridsearch on Breakout

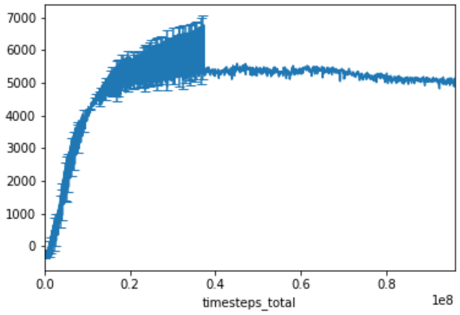

Figure 4: Gridsearch on HalfCheetah

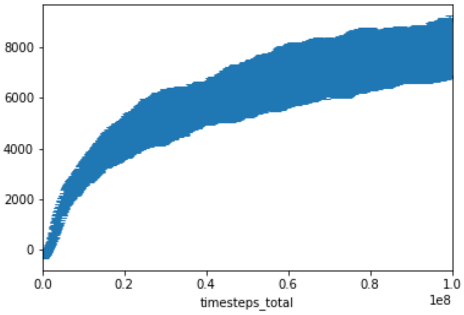

Figure 5: FedPBT on HalfCheetah

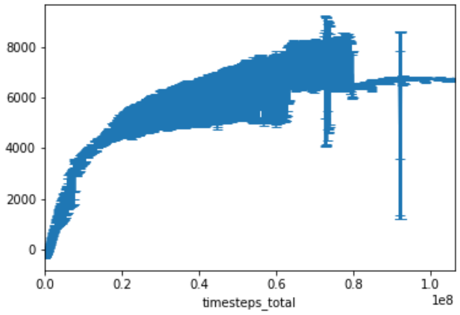

Figure 6: PBT on HalfCheetah

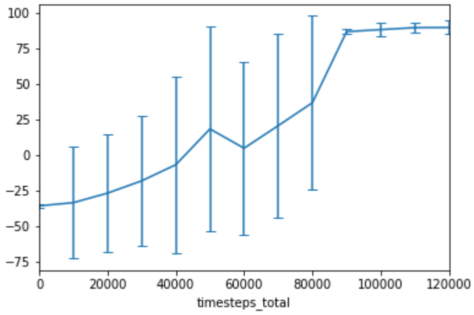

Figure 7: FedPBT on MountainCar

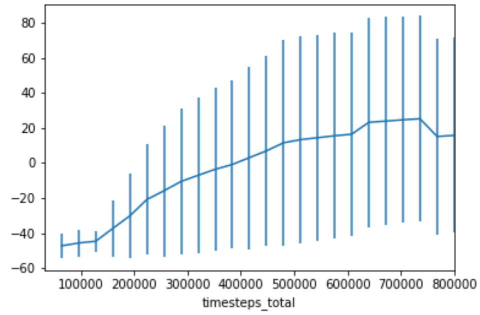

Figure 8: PBT on MountainCar

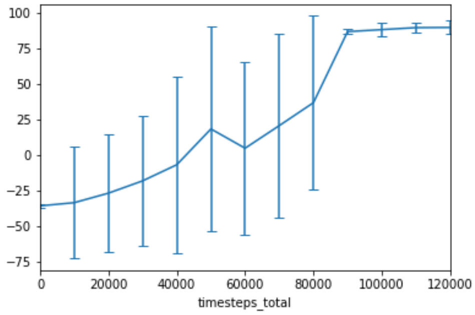

Figure 9: FedPBT on MountainCar

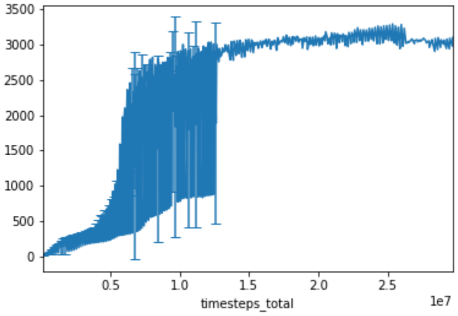

Figure 10: Gridsearch on Hopper

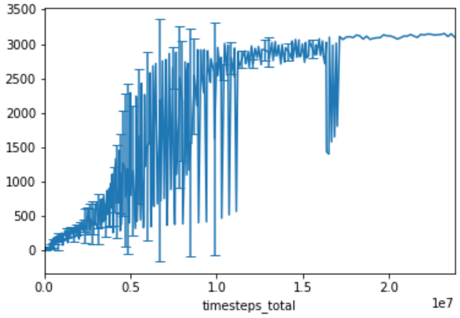

Figure 11: FedPBT on Hopper

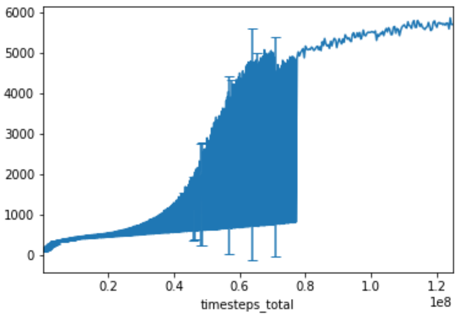

Figure 12: Gridsearch on Humanoid

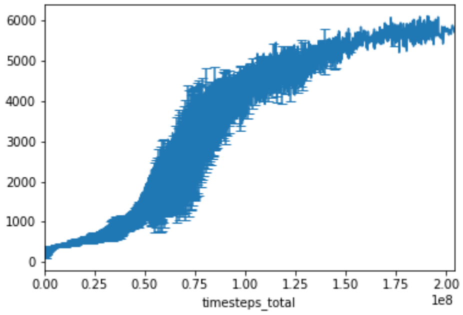

Figure 13: FedPBT on Humanoid

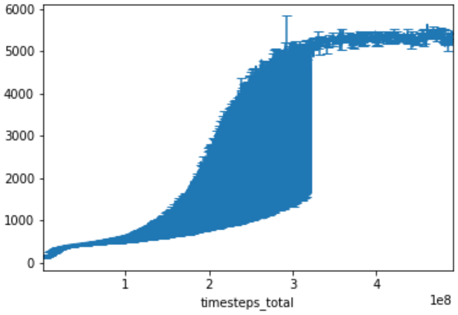

Figure 14: PBT on Humanoid

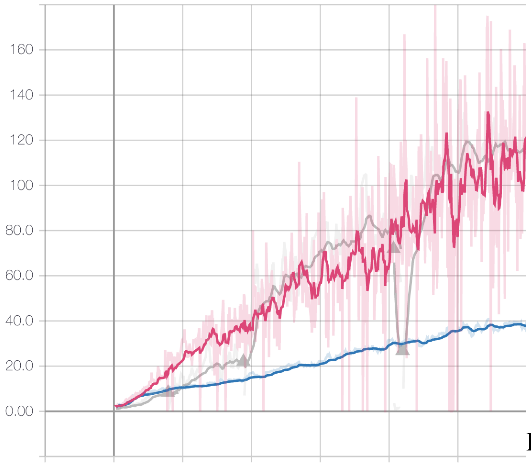

Figure 15: Breakout. PBT shown with 1/N samples. Pink: FedPBT, N=5, Temp=4. Grey: PBT, N=5. Blue: Fixed Lr 1e-4.

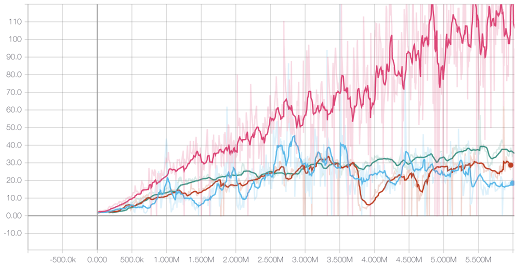

Figure 16: Pink: FedPBT, Breakout. Green: No exploration. Blue: No gradient sharing. Red: No exploration or gradient sharing.

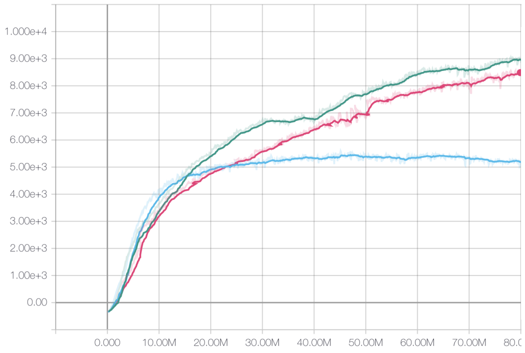

Figure 17: HalfCheetah. PBT shown with 1/N samples. Green: FedPBT, N=5, Temp=1.5. Pink: PBT, N=5. Blue: Fixed Lr 5e-5.

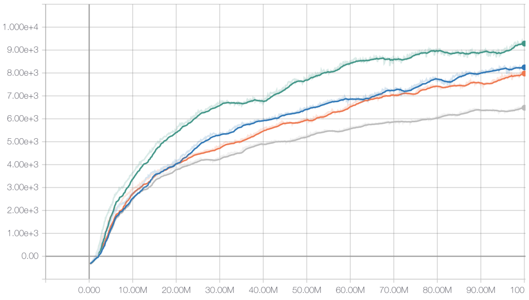

Figure 18: FedPBT on HalfCheetah. Green: Temp=1.5. Blue: Blue: Temp=1.0. Orange: Temp=2.0. Grey: Temp=0.5.