APPENDICES

A COVERAGE RESULTS

Figure 9 and Figure 10 show the coverage achieved across 15 replications in each of the experiments. The results show that all QD algorithms manage to find solutions which span the entire descriptor space. Since SPEA2 and NSGA-II do not seek diversity over the descriptor space, they achieve comparably poor coverage.

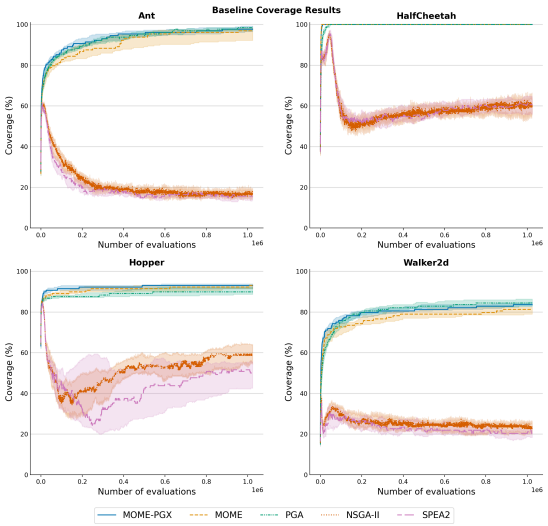


Figure 9: Coverage performance of MOME-PGX and all baseline algorithms for each of the tasks. The curves show the median performance across 15 replications and the shaded regions show the inter-quartile range.

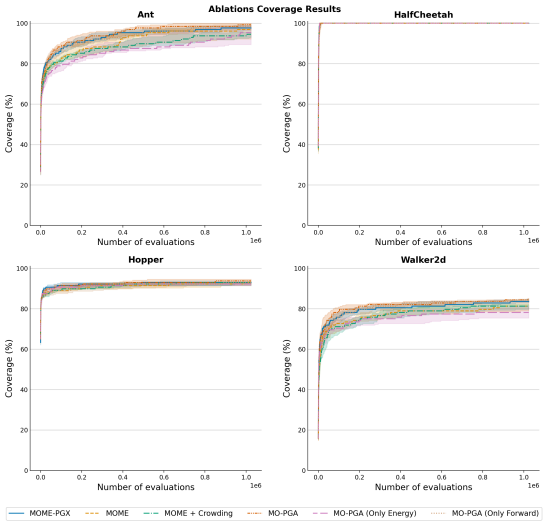


Figure 10: Coverage performance of MOME-PGX and all ablation algorithms for each of the tasks. The curves show the median performance across 15 replications and the shaded regions show the inter-quartile range.

B HYPERVOLUME REFERENCE POINTS

Table 1 presents the reference points used to calculate the hypervolume metrics in each of the tasks. The same reference points were used for all of the experiments.

Table 1: Reference points

Ant	[-350, -4500]
HalfCheetah	[-2000, -800]
Hopper	[-50, -2]
Walker2d	[-210, -15]

C POLICY GRADIENT HYPERPARAMETERS

Table 2 presents all of the policy gradient hyperparameters that are used for our algorithms. All hyperparameters were kept the same for each task and for all algorithms which used PG variations.

Table 2: Policy Gradient Network Hyperparameters

Replay buffer size	1,000,000
Critic training batch size	256
Critic layer hidden sizes	[256, 256]
Critic learning rate	3×10^{-4}
Actor learning rate	3×10^{-4}
Policy learning rate	1×10^{-3}
Number of critic training steps	300
Number of policy gradient training steps	100
Policy noise	0.2
Noise clip	0.2
Discount factor	0.99
Soft τ-update proportion	0.005
Policy delay	2