Figure 1: Comparison between 1) DARC's training reward in the source domain, i.e. $\mathbb{E}_{\pi_{\text{DARC}},p_{\text{src}}}[\sum_t r(s_t,a_t)]$, 2) DARC's evaluation reward in the target domain, i.e. $\mathbb{E}_{\pi_{\text{DARC}},p_{\text{trg}}}[\sum_t r(s_t,a_t)]$, and 3) the reward of the target optimal policy in HalfCheetah. Here, we followed the reviewer's suggestion to add a new comparison with the optimal policy in the target domain. According to the DARC training objective, $\mathbb{E}_{\pi_{\text{DARC}},p_{\text{src}}}[\sum_t r(s_t,a_t)]$ is expected to be similar to the reward of target optimal policy. In practice, there is a gap due to learning errors of DARC. We can see that the DARC's evaluation reward is worse than the target optimal policy, further showing a suboptimal performance of DARC in more general settings.

Table 1: Evaluation of DARAIL in more general off-dynamics settings. Here we changed the coefficient of gravity (the coefficient of gravity changed from 1.0 to 0.5 for the target domain). DARAIL outperforms the DARC evaluation reward.

	DARC Evaluation	DARC Training	DARAIL
HalfCheetah	1544 ± 127	5828 ± 417	5818 ± 326
Reacher	-16.9 ± 0.9	-16. 5 ± 1.4	-16.4 ± 1.6

Table 2: DARAIL results in broken source and intact target environment. We also included comparisons with the target optimal policy and direct imitation of the source optimal policy. DARAIL can improve DARC in this setting and significantly outperforms mimicking source optimal policy.

	Target Optimal Policy	Mimic Source Optimal Policy	DARC Evaluation	DARC Training	DARAIL
HalfCheetah	9235 ± 307	4512±398	4133±828	6995±30	**7067**±176
Reacher	-13.4 ± 1.4	-18.2 ± 3.7	-26.3±3.3	-11.2±2.9	**-13.7**±0.9

Table 3: DARAIL results in intact source and broken target environment (DARC settings). We also included comparisons with the target optimal policy and directly imitation of the source optimal policy. DARAIL can improve DARC (on target) in this setting and significantly outperforms mimicking source optimal policy.

	Target Optimal Policy	Mimic Source Optimal Policy	DARC Evaluation	DARC Training	DARAIL
HalfCheetah	8417 ± 263	1014 ± 73	7156 ± 828	7864 ± 32	7793 ± 237
Reacher	-12.2 ± 0.5	-19.6 ± 5.3	-18.8 ± 5.1	-17.6 ± 2.3	-18.4 ± 1.1

Table 4: Comparison with DARC with the same amount of rollout from the target. The number in the columns represents the amount of rollout from the target. More target domain rollout will not improve the DARC's performance further.

	DARAIL 5e4	DARC on Target 2e4	DARC on Source 2e4	DARC on Target 5e4	DARC on Source 5e4
HalfCheetah	7067 ± 176	4133 ± 828	6995 ± 30	4037 ± 798	6988 ± 27
Ant	4752 ± 872	4280 ± 33	5197 ± 155	4342 ± 42	5207 ± 172
Walker2d	4366 ± 434	2669 ± 788	3896 ± 523	2538 ± 802	3782 ± 510

Table 5: Comparison with DARC with the same amount of rollout from target, on Reacher. The number in the columns represents the amount of rollout from the target. More target domain rollout will not improve the DARC's performance further.

	DARAIL 5e3	DARC on Target 3e3	DARC on Source 3e3	DARC on Target 5e3	DARC on Source 5e3
Reacher	-13.7 ± 0.9	-26.3 ± 3.3	-11.2 ± 2.9	-29.7 ± 4.1	-10.2 ± 1.2