To all reviewers, thank you very much for your thoughtful comments and suggestions.

R#1 & R#2: Limitations of the linear regression. Although linear regression has limited application, it removes the optimization error (problem has an analytical solution), and enable us to study statistical and modeling trade-off. Our extensive empirical study on meta-RL complements this and fully describe the trade-off in a more general setting.

R#1: *"...importance of similarity among the selected tasks..."* In Theorem 1&2, similarity in the tasks can be described by $\mathrm{Var}_\gamma(\mathbf{Q}_\gamma)$ and $\mathrm{Var}_\gamma(\mathbf{Q}_\gamma \boldsymbol{\theta}_\gamma)$, implying that greater similarity leads to smaller statistical error.

R#1: *"...domain randomization, when enough samples are used, is a better alternative to meta-learning..."* In many practical deep-RL scenarios (highly nonlinear, low sample-case), we find optimization error dominates and DRS outperforms MAML. Although Theorem 4 shows that DRS needs some small amount of training data to be effective, empirical results (Fig 1&2) suggest that the amount of data required is as small as a few rollouts.

R#2: *"...Theorems 1 and 2 are asymptotic..."*: Only the first sentence of each theorem is asymptotic, the rest (starting with "specifically") holds for finite samples. Indeed, asymptotic statements are obtained via the limits of those finite bounds as $M, N \to \infty$. Hence, the theorems are NOT asymptotic. We will remove the asymptotic parts for clarity.

R#2: *'Assumption 2 ... the per-task optimal models are centered around the corresponding optimal solutions."*: We do NOT assume that the task optimal models are centered around the DRS/MAML optimal solutions, only that their distance is bounded. This assumption can easily be dropped with the cost of including the distance as a term.

R#3: *'...trust region alone cannot justify why ... TRPO fares better..."* Thanks for this insightful comment. We agree that the use of a trust region does not fully explain the behavior; we will investigate this and add further discussion.

R#3: Other points. We will add the links to the supplement. We will share the source code when the paper is public.

R#4: *"...as the learning rate goes to zero, MAML becomes DRS..."* We respectfully disagree that MAML with $\alpha = 0$ is the same as DRS. This statement is only true for the asymptotic objectives (Eq 9) and is not true in practice, when there are finite samples. MAML uses some data for the inner optimization regardless if the parameters are updated ($\alpha > 0$) or not ($\alpha = 0$). Our main contribution is analyzing MAML and DRS in the practical, finite-sample case.

R#4: *"...[1] reports consistent improvements over joint training..."* Our meta-RL experiments are more extensive. [1] considers locomotion with varying reward, while we consider locomotion and manipulation with varying reward or dynamics. On the HalfCheetahRandVel environment, the only one that overlaps, our result (Fig 6c) is consistent with [1] as MAML slightly outperforms DRS. Evaluations on a wider range of environments results in a different conclusion.

R#4: *"...learning rates..."* For TRPO-MAML's inner learning rate & trust region size, we use values from [16] for the locomotion environments (as we use their algorithm implementations) and [25] for the others; DRS+TRPO's trust region size was set to be the same as TRPO-MAML. The comparison is fair and follows practice of previous work.

R#4: *"...Joint training plus fine-tuning ... never does as well as it seems to in this paper..."* Our contribution is to introduce and discuss the trade-off between DRS and MAML. Existing literature compare them on a single point within this trade-off, whereas we explore the entire spectrum. Hence, those works happen to experiment on one side of the spectrum. Our paper does not conflict with existing works, it complements them by considering a larger picture.

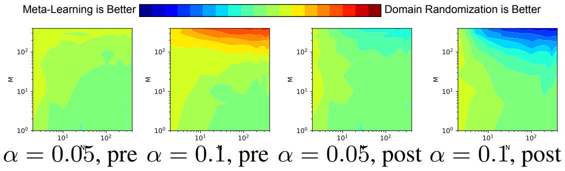
R#4: *"... Figures 2, 5, and 6... Why do these figures not align with each other?...* The one-sided Welch t-test obtains an estimate of the distribution of the difference between the average rewards of DRS and MAML. The variance of that distribution combines the variances of the two average rewards and is fairly large. Thus, even if the average reward for MAML is visually above that of DRS (Figure 5), the probability that DRS is better (Figure 2) may not be small. In other words, the difference in Figure 5 is not statistically significant considering the variances.

R#4: *"In Section 2, DRS does better with more training steps...Section 3, DRS does better with *fewer* training tasks..."* In Section 3, the optimization error is 0 for both methods (problem has analytical solution), and only the statistical and modeling error matter. MAML has smaller modeling error and so it is better with more data. In Section 2, the optimization error is significantly larger for MAML because of its bilevel structure; it dominates the trade-off for a wide range of training budgets. The difference between these behaviors is one of the main points of this paper. We show that the trade-off between DRS and MAML is not straightforward and requires careful analysis.

R#4: *"...it's shown that DRS requires a large training budget to be successful...* We showed that DRS needs some data (can be very small). Results in Figures 5&6 suggest that this budget is very small (only a few rollouts).

R#4: *"...[2] has an analysis..."* For meta-linear regression, [2] derive the MAML and DRS estimates when the task-specific losses are known. In contrast, we consider the practical scenario where the task-specific losses are not known, analyzing the estimates, their finite-sample behaviors, and their meta-test performances. We will add a discussion.

R#4: *"... if α is reduced to 0.1 (or below) ..."* The adjacent figure shows meta-linear regression results for $\alpha = 0.05, 0.1$. The same conclusions hold, but the difference between MAML and DRS is not as pronounced; MAML requires a larger data set to provide a clear improvement over DRS after test-time optimization.



$\alpha = 0.05$, pre $\alpha = 0.1$, pre $\alpha = 0.05$, post $\alpha = 0.1$, post