We thank the reviewers for clear and thoughtful feedback, and respond to specific points raised by reviewers below.

R2: “how the approach compares to [22].” **R3:** “absence of prior work that it out-performs”.

To address the primary concerns of **R2** and **R3**, we present results of new comparisons to Gupta et al. [22] on the Fixed ViZDoom experimental setting in Table 1. This comparison ([22]) is representative of “train[ing] an agent and task distribution using one of the 10s of DIAYN-like approaches” (**R3**) before freezing the task distribution and running meta-learning in a “pipelined” manner. However, we note that [22] considers environments with simpler, ground-truth state, as opposed to pixel observations.

The compared approaches are: **(i)** [22], which uses DIAYN [13] for task acquisition, adapted for pixel observations; **(ii)** an ablation of our method – "pipelined CARML" – more similar to [22], for an apples-to-apples comparison; **(iii)** [22], but initializing the DIAYN discriminator of with the image encoder of **(ii)**, to address failure modes of applying [22] in visual domains; and **(iv)** CARML, our full method.

Table 1: Comparing to [22].

	Avg. Succ.
i. [22]	0.291
ii. Ours, pipeline	0.535
iii. [22], smart-init	0.405
iv. Ours, full	**0.625**

Our approach outperforms [22] on transfer to test tasks. The benefit of our task acquisition method over that of DIAYN (which [22] uses) is indicated by the improvement from **(i)** and **(iii)** to **(ii)**. The benefit of using a curriculum for meta-learning over the pipelined approach of [22] is indicated by the improvement from **(ii)** to **(iv)**. **Please find discussion of these results at the end of the page.** We will include these and further experiments on the remaining settings in our revision.

R3: “Show that ... the newly proposed task is super useful".

We note that the environments considered are nearly identical to the navigation setting of [55] (though ours is more challenging insofar as no task description is given) and the manipulation setting used in [34], among others. Our work is among the first to study unsupervised meta-RL in visual domains, addressing challenges of pixel observation trajectories and partial observability, among others, which exacerbate the challenges of unsupervised RL and meta-RL.

Populating $\mathcal{D}$ (R2). We choose the simplest strategy that keeps complexity constant: sample a fixed number of trajectories uniformly at random from the entire history, i.e. reservoir sampling. We used a reservoir of 1000 trajectories (not tuned). We agree with **R2** that more sophisticated sampling strategies are worth pursuing in future work.

Comparison Details. Differing from [22], we use RL^2 instead of MAML for more direct comparability; to our knowledge, policy gradient MAML has yet to be successfully implemented in RL domains with pixel observations. Comparison **(ii)** uses a contextual policy to co-adapt with the task distribution before freezing the task distribution and meta-learning with RL^2. Results are reported for transfer to the Fixed ViZDoom test tasks, analogous to results in Figure 5a of submission. We use the same hyper-parameters for skill acquisition (i.e. number of skills) as existing experiments. In Table 1, we report the average of two runs per approach, but will use more in our revision.

Comparison Discussion (R2, R3). We find the task acquisition of DIAYN variants **(i, iii)** to suffer from an effect akin to mode-collapse; the policy’s data distribution collapses to a smaller subset of the trajectory space (one or two modes), and tasks correspond to minor variations of these modes. Skill acquisition methods such as DIAYN rely purely on discriminability of states/trajectories under skills, which can be more easily satisfied in high-dimensional observation spaces and can thus lead to such mode-collapse (related to the instability of GAN methods noted by **R1**). Moreover, they do not a provide a direct mechanism for furthering exploration once skills are discriminable.

On the other hand, the proposed task acquisition approach (Alg. 2, Sections 3.2, 3.4) **fits a generative model over jointly learned discriminative features**, and is thus not only **less susceptible to mode-collapse** (w.r.t the policy data distribution), but also allows for density-based exploration (Section 3.3). Indeed, we find that **(iii)** seems to mitigate mode-collapse – benefiting from a pretrained encoder from **(ii)** – but does not entirely prevent it. Overall, in terms of meta-transfer to hand-crafted test tasks, the DIAYN variants **(i, iii)** perform worse than pipelined CARML **(ii)**, due to the poorer diversity in the task distribution. We will incorporate this comparison, as well as additional visualizations (i.e. skill maps) of all discussed methods, in the revised Appendix.

Moreover, **(ii)** performs worse than "full CARML" **(iv)**. As in the paper, we hypothesize that this is due to the challenge of meta-learning more complex task distributions – compared to full CARML, the distribution of trajectories eventually discovered by the contextual policy of **(ii)** may be just as diverse and structured, but meta-learning the corresponding task distribution directly from scratch is harder. This shows **the benefit of co-adapting tasks with the meta-learner (iv)** as opposed to using a separate agent **(ii)**, and the value of investigating the effects of curricula on meta-learning.

[13] Eysenbach, Gupta, Ibarz, Levine. Diversity is all you need: learning skills without a reward function. ICLR 2019.
[22] Gupta, Eysenbach, Finn, Levine. Unsupervised meta-learning for reinforcement learning. arXiv:1806.04640, 2018.
[34] Nair, Pong, Dalal, Bahl, Lin, and Levine. Visual reinforcement learning with imagined goals. NeurIPS 2018.
[55] Chaplot, Sathyendra, Pasumarthi, Rajagopal, Salakhutdinov. Gated-attention architectures for task-oriented language grounding. AAAI 2018.