We would like to thank all reviewers for their feedback, hints towards typos as well as improvements in notation. We will make sure to incorporate them in the final version of the paper. Reviewer comments are in **bold**.

R1: Unclear how to choose the penalty proportion ξ, nor how it was selected. N_α not explained in the text. ξ has a monotone behavior, controlling the average progression speed to the target distribution. In all experiments, the performance initially increased with an increase of ξ until reaching a maximum. Hence a simple line search always sufficed to find good values of ξ. We will move the discussion from the appendix to the main paper and ensure that N_α does not only appear in the algorithm box.

R2: SPDL requires explicit knowledge of the desired distribution over contexts. While this may seem a burden, it actually makes assumptions of existing CRL algorithms explicit. Many CRL algorithms (including GoalGAN and ALP-GMM) choose tasks for training based on a proxy for learning progress. In the absence of unlearnable tasks, such an approach explores the whole context space, comparable to a uniform target distribution. This explains the comparatively strong performance of SPDL compared to GoalGAN and ALP-GMM in Experiments 1 and 2, where only a specific task is to be learned. ALP-GMM and GoalGAN cannot exploit this additional knowledge. **SPDL requires a parametrization of the context space space (unlike e.g. GoalGAN).** Access to a parameterized context space is an assumption shared by all CRL algorithms known to us (at least for continuous task spaces). In the experiments of the GoalGAN paper, desired positions to be reached by the agent are used as parameterizations of the context space, as can be verified in the code accompanying the GoalGAN paper. **Theorem 1 additionally requires that $f(r) = \exp(r/\eta)$.** We indeed forgot the assumption $f(r) = \exp(r/\eta)$ in Theorem 1 placing it only in Theorem 2 and will correct the mistake. Luckily, the assumption is standard in the RL-as-Inference literature so that its presence does not impair the applicability of our algorithm. **Theorem 2: Choice of f is only possible with particular constraints on the MDP (i.e. non-positive rewards is sufficient).** We think there is a misunderstanding. As correctly mentioned, for a probabilistic interpretation of the rewards, these need to be assigned non-negative real numbers. This is, however, already enforced by the monotonic transformation f whose domain are the positive real numbers (introduced in line 127). f represents the (unnormalized) probability $p(\mathcal{O}|\tau, \mathbf{c})$. **What is meant by a p-test in the caption?** We will correct this typo to a t-test. **Is SPDL applicable when a low-dimensional context parameterization is not available?** As mentioned, a parameterization of the context space is unavoidable in the continuous setting. Regarding dimensionality, we believe that the method as it currently is will likely scale to 5- to 7-dimensional context spaces. For higher-dimensional spaces, we believe that more advanced representations of the context distribution than an anisotropic Gaussian are required.

R3: theoretical results provided in Section 4 disconnected from algorithmic design. We agree that the algorithm can be motivated from simpler arguments. However, our theoretical interpretation connects it to the rich literature of inference. We strongly believe that the theoretical grounding, although harder to process, allows for a more rigorous understanding of why the algorithm works, and hence is just as important as good algorithmic performance. Further, the theoretical motivation differs a lot from the commonly employed proxies based on Intrinsic Motivation which makes it additionally interesting. **How would SPDL perform on a set of discrete tasks with a Dirac delta distribution over one of them (e.g. creating a curriculum over finitely many initial states)?** SPDL can also be applied in discrete settings, since the (currently Gaussian) context distribution of SPDL can easily be replaced with a discrete one. The KL-Divergence employed in the RL-as-Inference framework, however, does not allow for a Dirac-Delta to be used. SPDL as of now could only employ a smoothed version of such a Dirac-Delta or a uniform distribution if desired. We ran a small experiment on an 8×8 Grid-World in which two keys need to be collected to open doors for reaching a goal position. Only a reward of 10 is given when reaching the goal. We selected 8 different starting states and ran SPDL as well as EXP3.S with "Absolute Learning Progress" (ALP), a standard formulation of intrinsic motivation in RL e.g. used by ALP-GMM, over 20 seeds to generate curricula over the starting states. Using PPO, both SPDL and Exp3.S generate curricula which prioritize states in a reverse order, i.e. moving from states close to the goal to the state furthest away from it. Both algorithms improve upon uniform sampling over contexts, although EXP3.S performs slightly better than SPDL. We are convinced that there are many possibilities for improving the algorithmic implementation of SPDL in discrete scenarios, as the current implementation focuses on continuous context spaces (e.g. for robotics) where a big challenge are intractable integrals. In a discrete setting, expectations can often be easily evaluated, allowing e.g. for more advanced sampling techniques from the context distribution. **How would you solve some Atari game via SPDL?** For Atari games, a context space encoding environment versions of different difficulty could be defined. For Space-Invaders an interesting parameterization could be movement speed and shooting frequency of enemies.

R4: Is [the curriculum] consistent in a task/independent of the underlying algorithm? How sensitive is it? We re-investigated the curricula generated by SPDL in the experiments. The evolution of the distributions looks consistent across algorithms. Depending on the task, they however exhibit a significant amount of variance, as e.g. in the point mass experiment: While the friction parameter is continuously annealed (although with varying pace), there are curricula that prioritize moving the gate to the target position before shrinking it to target size or, vice versa, first shrink the gate and then move it. Further, there exist all interpolations between these two extremes. Investigations of sensitivity were unfortunately out of scope given the short time of the rebuttal period, but are certainly interesting points for future work.