We are grateful to the reviewers for their careful reading and thoughtful comments. Reviewer #1 states that "The two contributions are from in-depth analyses and thus, are substantially more rich than the standard submission...I think this could be a highly important paper for the field, and inspire a number of extensions by different researchers." Reviewer #2 writes "I found this to be a good paper and enjoyed reading it" and "I found the discussion of the experiments to be very clear." Reviewer #3 notes that the paper "although the idea is simple, such a study would be of great importance for the continual learning community" while Reviewer #5 writes that we "show empirically on some strong experiments that [our algorithm] has both stability and plasticity."

We agree with the reviewers that our paper offers (i) a simple method for preventing catastrophic forgetting in continual learning that makes no assumptions about task boundaries (unlike other algorithms), (ii) innovative empirical analyses that set new standards for probing the properties of continual learning algorithms, (iii) demonstrations on hard problems, including standard benchmarks in this area, that we achieve both stability and plasticity, surpassing state-of-the-art. We respond to particular comments and questions below.

Reviewer #1. *(a) Task setup.* "Why did you pick those three?" They represent very distinct tasks within the DMLab suite. "How is your method affected by the number of tasks it needs to learn?" As shown in Figures 4 and 7, it is possible for the method to perform well up to (at least) six tasks, which was the maximum we tested with. "Is there a tradeoff in the number of examples needed per task as the number of tasks grows?" Figure 4 (plasticity) shows that the time required to learn a new task does not immediately degrade as the number of tasks grows, though clearly this will not hold for arbitrarily many tasks. Please also see our response to Rev. #3, point (a), below.

(b) Interference and forgetting. We will make this difference clearer in the text: we see interference as a consequence of the nature of the tasks involved making them more difficult to learn together than separately - e.g. learning how to drive on the right side of the road may be difficult while also learning how to drive on the left side of the road. Forgetting, by contrast, results from sequentially training on one and then another task - e.g. learning how to drive and then not doing it for a long time. Most pairs of tasks seem to interfere minimally, if at all, while forgetting occurs in deep RL for essentially any pair of tasks.

(c) RL terminology. We will make this clearer in the revision, expanding on the definitions of terms such as "rollout" and "Importance Weighted Actor-Learner" and giving intuition for the V-Trace algorithm.

Reviewer #2. We will make the exposition of RL techniques much clearer in the revision, as noted above. Thank you for calling this to our attention. In particular, we will provide greater description and intuition for the V-Trace algorithm, in addition to being more explicit about overall RL terminology.

Reviewer #3. *(a) Task setup.* The cyclic training paradigm is versatile, and we wanted to explore our method thoroughly in the simplest non-trivial forgetting scenario. Note that the agent does indeed spend a very large amount of time (75 million environment steps) on each task before switching to the next, causing, as we demonstrate, baseline RL algorithms to thoroughly forget each task. Please also see our response to Rev. #1, point (a). Other task setups would indeed be interesting to consider in future work, but we also think this paradigm merits adoption in subsequent papers.

(b) Other off-policy algorithms. Yes, we could use a different off-policy RL algorithm – though the demonstration of the principle and its utility would be largely the same. Rather than compare multiple off-policy algorithms, we focused on exploring performance in different experimental setups. Making Retrace work is an exciting avenue for more work in the future. We've now commented in the paper how our algorithm can be straightforwardly combined with stronger off-policy learning algorithms.

(c) Destructive interference. We were trying to fix the specific problem of forgetting, and therefore tried to disentangle these various effects by minimizing both constructive and destructive interference. While very minimal constructive interference resulted naturally from tasks being drawn from the same suite, we expect that destructive interference is a less common phenomenon that one may explicitly need to seek out, arising primarily when two tasks impose conflicting constraints on the learner. See Figure 1 and Subsection 4.1.

(d) Small memory buffers. Surprisingly, we actually didn't observe any significant decrease in performance from reducing the memory size (see Figure 6), even with a buffer as little as 0.5% of past experience. Exploring what happens at even smaller buffers would indeed be an interesting future direction.

(e) New-replay ratio. For Atari, 75-25 performed slightly better than 50-50, but in both DMLab and Atari, both 75-25 and 50-50 worked very well (see Figure 5).

Reviewer #5. We would respectfully disagree that the paper is missing an interpretation of empirical results or justification for the plasticity and stability properties we observe. We would be very happy to address specific criticisms or failures in clarity in the revision. We will certainly provide more background for terms from the RL literature such as "behavioural cloning", "V-Trace", and "historical policy distribution" (see Rev. #1, point (c)).