We thank the reviewers for the feedback. We first highlight our main contributions and then respond to specific points.

R1, R2, R4: Contributions: (1) We emphasize that to the best of our knowledge, Theorem 4 in our paper is the **first result** to show that for a specific class of test functions, one obtains **improved asymptotic CLT rates** by discretizing **Underdamped Langevin Diffusion (ULD)** using Euler or Randomized Mid-Point (RMP) methods. Prior works demonstrated improved W_2 rates for sampling with Euler or RMP discretizations of ULD compared to Overdamped Langevin Diffusion (OLD). However, such results are not insightful as far as the CLT rate for numerical integration is concerned, which is the focus of our paper. (2) Surprisingly, while **RMP discretization** obtains improved W_2 rates for sampling over Euler discretization, it **has the same CLT rates as Euler discretization** in both OLD and ULD settings. This points to the second major contribution of our submission: One should take the W_2 rate improvements with a pinch of salt if the main goal is to compute integrals, arguably the main application of samplers in machine learning.

R1: Regarding Proposition 2.1 - As discussed in Lines 142-145, we would like to clarify that, to obtain the same W_2 rates based on OLD, other discretizations require higher-order smoothness assumption whereas Proposition 2.1 only requires gradient-smoothness. **Initialization:** For the RMP discretization, if the initialization is far from the minimizer, we have to decrease the step size much quicker as large step size in the region where gradient is large (far from minimizer) can result in very inaccurate discretization of the process. In this sense, the rate will deteriorate. Also in the regime we consider, obtaining a point that is ϵ close to such a minimizer only costs $\mathcal{O}(\log(1/\epsilon))$ iterations. Compared to the sampling rate, which is $\mathcal{O}(1/\epsilon)$, this is significantly cheaper. **Thms 1 and 3, and Ergodicity:** Ergodicity results in Thms 1 and 3 are important properties of RLMC and RULMC, and they form the basis of the CLT results. In other words, to obtain confidence intervals (CIs) which is through establishing a CLT, stated (asymptotic) ergodicity results are sufficient. We emphasize that it is possible to obtain geometric ergodicity results (with rates); however, we preferred not to add this, as they are not needed for establishing CLT results, which are our main focus. **Prop 2.2:** Indeed this is an important observation and the bias is higher than vanilla Euler discretization of OLD. This is also reflected in the CLT bias term ρ in Lines 208-209. **Prop 3.1:** We emphasize that the conclusion in Prop 3.1 does not follow from [SL19]. Indeed, one possibility to use the result in [SL19] to obtain bias in W_2 metric is to let the iterations go to infinity in their main theorem. However, their result (and proof) assume an upper bound on the number of iterations (depending on the choices of ϵ). This leads to a worse bound than our stated result. **Biased/Unbiased CLT & Limit of $\hat{\gamma}_n$ always exist?** These two aspects are related and form the crux of our main result. Note that in Theorem 2, lines 205, 206 and 207 respectively, we consider three cases: when the limit $\hat{\gamma}_\infty := \lim_{n\to\infty} \hat{\gamma}_n$ is (i) equal to 0, (ii) between 0 and ∞, and (iii) equal to $+\infty$ (and hence doesn't exist). Next, note that in the definition of $\pi_n^\gamma(\varphi)$, we have already centered with the true expectation (which is zero in our setup by definition). Only when the limit $\hat{\gamma}_\infty$ equals zero, we obtain an unbiased CLT, meaning the normal distribution is exactly centered at the true expectation. If the limit $\hat{\gamma}_\infty$ is in the interval $(0, \infty)$, the CLT is biased meaning it is no longer centered around the truth – as a consequence, we need to do appropriate bias-correction when obtaining practical confidence intervals. When the limit $\hat{\gamma}_\infty$ is $+\infty$, we converge to a degenerate random variable (i.e., a constant). We hope this clarifies our main results. The same interpretation applies also to Theorem 4. **Comparison & No surprising result is shown:** Please see Lines 2-10 above for our main contributions. Apart from the result in [LP02] (which we have compared against), we are not aware of any CLT results, in particular for ULD. We would greatly appreciate pointers to specific papers if the reviewer thinks otherwise – we would be happy to cite them and compare our results with those in the suggested papers.

R2: Bias variance Trade-off: With respect to sampling (i.e., W_2 rates), roughly speaking, the optimal rate obtained in [SL19] is exactly based on picking a constant step-size by trading off bias and variance. However, as we show, to obtain a CLT for numerical integration, especially the one centered on the true value of the integral, one needs to have a specific decreasing step-size choice. **Complicated terms in CLT:** We will add specific examples in our revision, to provide more insights. However, the main take-away from our general results is the asymptotic rate improvement of CLT with RMP/Euler discretization of ULD. **Simulations:** We definitely agree with the reviewer that adding numerical experiments would be enlightening. We will add simulations to the camera-ready version if the paper is accepted.

R4: Comparison: Note that [LS19] provides a table comparing results only for the W_2 rates for which there are several related works to compare against. Our main contribution in this work is the CLT rate improvement with the RMP discretization of ULD. We are not aware of prior works in this direction, to the best of our knowledge. Nevertheless, we will take R4's advice and add a table comparing CLT rates for RMP and Euler discretizations of both ULD and OLD established in our paper and [LP02]. If the reviewer is aware of any other state-of-the-art CLT results for some discretization of ULD or OLD, we would greatly appreciate it and would add it to the table in our revision. **L235-237:** We mean the tuple (U_1, U_2, U_3) is in R^{3d} i.e., $U_i \in \mathbb{R}^d$, for $i = 1, 2, 3$. Also, we missed an identity matrix in the covariance definition. Thanks for catching this. We will fix these typos in our revision. **L131, L202:** These suggestions are well-taken and we would incorporate them in our revision. **L305:** $\mathcal{O}(n^{5/8})$ is the best achievable rate by picking for α among polynomially decreasing step-size choices of the form $\gamma_k = k^{-\alpha}$. Hence, we call it *optimal* following the terminology in [LP02]. We will clarify it in our revision as *optimal among polynomially decreasing step-size choices*.