We thank the reviewers for their extremely helpful feedback. We also thank all reviewers for pointing out the typos, which we shall all correct in the final revision.

Organization, Clarity, Notation: We appreciate your feedback regarding the organization. Due to space constraints, we provided a brief proof sketch, and chose not to include further details of the StrongEuler algorithm. However, we recognize that this can obfuscate the intuition about both the algorithm and its analysis. We therefore propose the following reorganization: **Optimistic Algorithms:** After describing the general MDP setting but before the statement of our main results, we will introduce the notion of optimistic algorithms. We will explain how StrongEuler and related algorithms in the literature achieve optimism by combining empirical estimates of transition probabilities/rewards with confidence bound bonuses. The specific form of the bonuses for StrongEuler will remain in the appendix to save space, but this should give more intuition for how StrongEuler operates. This will also address help to clarify for which class of algorithms Theorem 2.3 applies (re: Reviewer 3's comments). **Proof Sektch:** In order to improve intuition and cut down notation overhead, we will instead state a simplified but informal version of Prop 3.2 that specifies the dominant term, albeit with a coarser dependence on the variances. This will eliminate clutter and allow us to simplify the rest of proof. In addition, this will free up additional room to further explain the intuition behind the clipping in Prop 3.1. In particular, we will provide intuition for the proof, and clarify why $\mathtt{gap}_{\mathtt{min}}/H$ appears in the analysis. This will in turn clarify the dependence of $\mathtt{gap}_{\mathtt{min}}$ in the final bounds. **Organization of Supplement:** The proofs were structured so that the proofs of the supporting lemmas were deferred until Appendix F. We now recognize that it may be easier to incorporate the proofs of these lemmas at the end of the section, and so we will refactor Appendix F so that the proofs of the lemmas come closer to their statements. We shall also include an "organization" section for the appendix to explain how the different sections fit together, and split the supplement into groups (i.e. Part I. proof of upper bounds Part II. Proof of clipping and other technical results, Part III proof of lower bounds). **Notational Overhead:** Regarding the notational overhead, this seems to be a difficulty encountered by many papers in the MDP regret space, and we intended our notation to be consistent with prior art. While we do include a table of notation, we appreciate the reviewers suggestions that we remind the reader of notation. We will also include remarks that explain notational rational (e.g. overline means "optimistic estimate"). We attempted to lighten notation by providing bounds at three levels of granularity: (in the body, in App A, and in App C), and because of this, we should be able to simplify notation in the proof sketch in the body of the paper by incorporating an informal statement of Prop 3.2 as described above.

Reviewer 1: Regarding your concern about the horizon dependence of the gaps, we define a horizon dependent gap in Definition 1 (i.e., for times $h \in [H]$). For simplicity, we state the results in the main text in terms of the minimum over the horizon dependent gaps, though more granular bounds are given in Appendix C. We confirm your suspicion that $\mathtt{gap}_{\mathtt{min}}$, as defined, may be zero if $\mathtt{gap}_h(x,a) = 0$ for all x, a. However, our results will go through with a more refined definition of $\mathtt{gap}_{\mathtt{min}} := \min_{x,a,h}\{\mathtt{gap}_h(x,a) : \mathtt{gap}_h(x,a) > 0\}$, and we will be sure to restate our results to use this definition instead. This way, $\mathtt{gap}_{\mathtt{min}}$ is strictly positive unless all actions are equally good for all states and all stages (in which case the regret is trivially zero). We will also include a remark to clarify this point.

Reviewer 2 Thank you for your helpful feedback. I hope your major concerns are addressed in the organization suggestions listed above. Regarding the general comments: **1.** Line 97 regarding "average" reward. We apologize for the imprecision. **2.** Line 168: We state in Sec 1.2 (the problem setting) that the rewards is a variable is bounded in [0,1]. **3.** Line 200: We will include the quantifier for $h \in [H]$. Thanks for pointing out the omission

Reviewer 3: Thank you for the detail feedback **1.** We hope we addressed your concerns about organization in the discussion above. **2.** Regarding point 2, yes you are correct an algorithm is optimistic with high probability. It would be more accurate to state in the definition that "an algorithm satisfies (strong) optimism if...", and then say that we informally refer to an optimistic algorithm as one that satisfies optimism with high probability. **3.** In appendix H, the formal statement of the lower bound does explicitly give the class of algorithms which suffer Thm 2.3, and we will do a better job to signpost this. To our knowledge, every low regret algorithm in the literature for this setting falls under this class, and we can add this clarification as well. Moreover, as mentioned in the discussion about organization above, explaining how optimistic algorithms are constructed from confidence bonuses will allow us to provide, in the main text, an intuitive explanation of the formal class of algorithms which suffer the lower bound. **Additional comments: 4.** We can explain the intuition for the ϵ parameter in the bound, and add a quantifier $(\forall \epsilon)$ in the definition of Zsub. Since the bound takes a minimum over ϵ, it is not a parameter that requires specification. **5.** Thanks for the clarification of Jaksch. We will clarify to state that their gap $\mathtt{gap}_*$ depends on measures of ergodicity that show up implicitly in the other asymptotic analyses. **6.** To clarify what we mean by "almost" gap-independent, we will add a term $\mathrm{poly}(\log(\mathtt{gap}_{\mathtt{min}}))$ to capture an at most poly logarithmic dependence on the min gap. **7.** $\widetilde{S}$ and capital P are typos - thanks for the catch! **8.** Thank you, we will clarify that uniqueness is essential for the proof of [15] **9.** For the inequality on 897 we will clarify that a, b are positive. Then $\sqrt{a} + \sqrt{b} \leq \sqrt{a+b} + \sqrt{a+b} = 2\sqrt{a+b}$, whence the inequality follows.