We thank the reviewers for their thoughtful comments and helpful suggestions, and respond to comments below. We are pleased that the reviews are enthusiastic. R1: *"Interesting theoretical connections of maximum expected hitting cost to potential based reward shaping ... This work may be useful in practice to craft informative shaping rewards"* and *"good significance. A tighter upper bound to diameter under the same set of assumptions is introduced while the connections to potential based reward shaping are interesting. Furthermore, this paper may be useful in practice to craft better shaping rewards."*; R2: *"The paper is well-written and provides an interesting insight on the fact that the reward function in itself should ideally be taken into account in complexity measures."*; and R3: *"the paper improves the regret bound on a very important algorithm by simple analysis"*

R1: "bugs" in designed rewards (minor comment 1) – In the particular context, we meant that from the perspective of the reward designer, the specified reward function might not be consistent with the desired behaviors (see L171). An example is the commonly cited OpenAI blog post on a faulty reward function in a game called CoastRunners (regrettably we cannot include any external links in the rebuttal). The quotation marks are to indicate the colloquial use of the term bug. We appreciate your suggestion and we shall add an explicit reference in any future versions.

R1: discussion of contributions (minor comments 2-4) – We agree with your assessment that the discussion section can be enhanced by emphasizing the *formal* results we established in this work. In contrast, in prior works, PBRS's learning efficiency was only supported by numerical evidence. We are also excited to mention some prospects to extend this work in both theory and practice in any future versions.

R2: applicability of MEHC beyond `UCRL2` – We are also curious about the same question for non-optimism-based algorithms. We plan to pursue it in future works.

R2: r_{max} in the definition of MEHC – This is an interesting suggestion and we will consider it seriously. One benefit we currently see in keeping r_{max} in the definition is to remind readers (and users) that we assume some knowledge of r_{max} in the bounded MDP setting (see footnote 2). Note also that a regret bound has a *unit* of "rewards" and even in the case of the unitless diameter, the resulting regret bound would include r_{max}, e.g., $\tilde{O}(r_{max}DS\sqrt{AT})$ from the original `UCRL2` analysis.[1]

R3: MEHC and diameter – We agree that in *some* MDPs the gap between these two quantities can be small (or zero). However, we think this new complexity measure is worth studying as it provides a valuable tool to study the impact of rewards on learning efficiency as we have shown.

R3: limited impact of PBRS on MEHC – We refrain from making a harsh judgment on the merit of PBRS because in this work, we focus on the average reward setting and `UCRL2` whose regret scales with MEHC. It is conceivable that for a different setting and a different RL algorithm that does not scale with MEHC (or scales poorly with a larger exponent), such as SARSA with epsilon-greedy exploration as used in [NHR99, footnote 4], PBRS might create a larger impact on learning efficiency as you suggested. We do agree that the discussion section may be enhanced by pointing out this future research direction and by adding qualifications lest the claims sound exaggerated.

R3: PBRS vs other techniques – As noted in the paper, PBRS is restrictive as the shaped rewards and the original rewards are Π-equivalent (see L156). It is reasonable to expect a pair of non-Π-equivalent rewards—under some other weaker notion of equivalence—to have a greater difference in their learning efficiencies (under some algorithm). Furthermore we want to remark that [SBC05] studies a different RL setting with *salient events* and a comprehensive comparison of different means to incorporate expert knowledge is beyond the scope of this work.

R3: a proof sketch – We tried to make the detailed proofs (included as an appendix in the supplementary material) easy to follow, but we agree that a proof sketch will further enhance the paper by providing more intuition to readers. In particular, as you have complimented, Theorem 2 is a very interesting result and our proof, instead of constructing the best/worst potentials, relies only on the definitions of MEHC and that PBRS does not change rewards on a loop.

[1] In [JOA10], r_{max} is assumed to be 1 reward-unit making the expression *look* unitless.