We thank all the reviewers for the constructive suggestions. To address concerns on readability, we included two extra figures illustrating the factored MDPs and the factored span. We could not include them in the response due to the page limitation. All the clarity questions are addressed in the main paper as well. This author response consists of two parts: 1) a table comparing computational oracles and regret bounds for different algorithms to clarify our main contributions; 2) point-to-point responses to questions from each reviewer.

Comparing computational oracles and regret bounds. To clarify our contributions, we provide the following table comparing oracles and regret bounds for the algorithms mentioned in the paper. Our proposed DORL eases the oracle in UCRL-Factored (adapted to non-episodic setting). FSRL is proposed for a tighter regret bound.

Algorithms	F-RMAX	UCRL-Factored	PSRL	DORL	FSRL
Works	Strehl (2007)	Osband et al. (2014)	Osband et al. (2014)	This work	This work
Regret	(mixing rate)$T^{3/4}$	$DT^{1/2}$	$DT^{1/2}$ (Bayesian)	$DT^{1/2}$	$Q(h(M))T^{1/2}$
Oracle	Planning oracle	Optimizing average reward within a confidence set	Planning oracle	Planning oracle	Optimizing average reward with bounded factored span

Response to reviewer 1. UCRL3 (Bourel et al., 2020) improves the dependence on actions and states by restricting the set of successor states of the state-action pair. We believe this construction is quite independent from the FMDP construction and the similar adaptation can easily be applied. Recently, Tian et al. (2020) closed the gap on horizon H in episodic case. However, the story is more complicated in non-episodic case as the horizon length H always serves as a tight bound on the connectivity.

The techniques are mainly adapted from Agrawal and Jia (2017); Osband and Van Roy (2014), while we point out two technical accomplishments: 1) we provide a new way to bound Bayesian regret of PSRL in non-episodic setting using the fact that simple deterministic horizons can also achieve a near-optimal regret bound (Lemma 4). In UCRL2 and Ouyang et al. (2017), length of episodes are (partly) determined by the doubling trick and are random variables. 2) The proof of Theorem 4 in Appendix G provides a tighter deviation bound using factored span, which we haven't seen in any previous literature.

On the clarity question, h^+ simply denotes the optimal bias vector of FMDP in the lower bound construction.

Response to reviewer 2. Our simulation settings are from Guestrin et al. (2002), while we agree that more experiments on general environments shall be tested in the followings works.

Since the accurate algorithm does not exist even for standard planing oracle, we discuss approximate algorithm for FSRL here. One can simply add an extra constraint on factored span for the approximate algorithm in Guestrin et al. (2003) and the optimization problem can still be written as a linear programming and be solved efficiently.

On the other question about DORL, extended factored MDP is proposed to ease the oracle in UCRL-factored. The correctness is preserved as the optimism can be preserved by only keeping a finite set of MDPs instead of a infinite set of all the possible MDPs. Theoretically, DORL has an regret bound of $\sqrt{T}$, while factored RMAX only achieves $T^{3/4}$, which is why we expect DORL to outperform factored RMAX.

Response to reviewer 3. Theorem 2 does not involve prior distribution as it considers the frequentist regret of DORL. In Theorem 1, we explicitly pointed out that the true prior distribution ϕ is over the set of MDPs with diameters $\leq D$.

As mentioned above, we agree that more experiments on general environments shall be tests in the followings works.

Our lower bound seems weaker because we state the theorem in terms of span rather than diameter. However, we suggest that when restricted to the tabular case, they are the same construction and have the same lower bound, because in the construction of Jaksch et al. (2010), the span is propositional to the diameter.

In Line 128, "adding extra discrete actions" means for each sampled MDP, we add an copy of actions corresponds to each of the FMDPs in the discrete set. In Line 131-132, instead of sampling, we directly construct a finite set of FMDPs.

Response to reviewer 4. We used the simplest demonstration in the theorems. In fact, $m\sqrt{W}$ can be easily replaced by $\sum_{i=1}^{m} \sqrt{|S_i|}$. We suggest that both diameter and span of bias vector are the most commonly used connectivity measure in non-episodic MDPs and their intuitions have been well-understood in Puterman (2014); Bartlett and Tewari (2009).

On the questions about simulations, we proposed two new algorithms DORL and FSRL. While FSRL is not implemented, DORL is tested in our simulation and achieved good regret curves. We also point out that the simulation indicates that the optimal policy could still be found despite of using the approximate planner, which is of some significance.