We thank all reviewers for their positive comments. Below we first address common concerns among the reviewers, and then respond to questions raised by individual reviewers.

1. Response to common concerns

- "Numerical experiments": Our paper focuses on theoretical aspects of risk-sensitive RL. It is an excellent suggestion to conduct numerical experiments to support our theoretical results. We will follow up on this.

2. Response to individual reviewers

Reviewer #2

- "Numerical support": Please see our responses in the previous section.
- "Step 10 of Alg 1": Yes.
- "Non-linear functional approximator": At this point it is unclear how non-linear functional approximation would affect our results, and we believe that it is a very interesting and important future research direction.

Reviewer #3

- "Practical relevance": Risk-sensitive RL finds applications in practical and strategic decision-making scenarios where risk consideration is crucial. Examples of such scenarios include, but are not limited to, autonomous driving, medicine prescriptions and financial investment. Our regret analysis provides a critical insight that under the risk-sensitive setting, the number of samples required to learn optimal policies scales exponentially in risk sensitivity, which serves as a guideline for practitioners on data collection and algorithm deployment. Our algorithms provide a way to achieve the (almost) best possible convergence rate and sample complexity for the risk-sensitive RL problem, and they are both easy to implement. Since this work focuses on theory, we leave numerical studies for our algorithms to future work.
- "Universal constant": The universal constants in bonus terms are artifacts of standard concentration inequalities, and setting them to a large value such as 100 would suffice in practice.
- "Empirical demonstrations": Please see our responses in the previous section.

Reviewer #4

- "Challenges of non-linearity": The non-linearity of the Bellman equations poses several challenges. (1) Algorithmic design: it is unclear a priori how Q-functions should be updated given the non-linear Bellman equations, and how bonus terms should be designed to enforce "optimism in the face of uncertainty" in a principled way; (2) Regret analysis: previous regret analysis of value iteration and Q-learning algorithms depends crucially on the linearity of Q-functions wrt value functions and bonus terms. It is unclear a priori how the existing proof techniques could be adapted to analyze our algorithms.
- "Lemma 1": The purpose of Lemma 1 is to demonstrate a surprising contrast between the range of value functions and our regret bounds: while risk-sensitive value functions are on the same scale as their risk-neutral counterparts, which is independent of $\beta$, the regret bounds under the risk-sensitive setting have exponential dependency on $|\beta|$.
- "Key contributions": Another key contribution of our work is that we provide a regret lower bound that scales exponentially in $|\beta|/H$, which certifies the near optimality of our upper bounds.
- $b_h$ "We have defined $b_h$ in Line 9 of Alg 1.
- "Experiments": Please see our responses in the previous section.

We appreciate the minor issues pointed out by the reviewers, and we will fix them in our final paper.