We thank the reviewers for their valuable suggestions. Please find our answers for each reviewer below.

Reviewer 1

We thank the reviewer for the positive assessment of our work. Below, we provide a concrete plan of incorporating reviewer's feedback in the updated version of the paper.

Extended experimental analysis. As suggested by the reviewer, we will add a more detailed analysis about the experimental results in the paper. In particular, we will add the following experiments/details/results: *(i)* evaluation of learner-aware teaching under unknown constraints for L3-L5 (the findings are similar as for the already presented experiments); *(ii)* experiments illustrating the effect of C_r and C_c in soft preference constraints; *(iii)* additional details and discussion of parameter choices in our experiments; *(iv)* reporting the run time of our algorithms, and illustrating scalability w.r.t. the problem size; and *(v)* reporting standard errors in Figure 3 (b) (the currently reported results in the paper are significant at significance level 0.1).

Ideas for outlook and future work. To the best of our knowledge this paper is the first to consider IRL with preference constraints. Hence, we primarily focused on developing the theoretical framework and algorithms (for both the known and unknown constraint settings). Nevertheless, we agree that the directions suggested by the reviewer (more complex domains and human subject experiments; suboptimal demonstrations and implications on performance; addressing the problem from a learner's perspective) are important. We will add a discussion on these directions in the revised paper.

Technical clarifications. Below we answer the technical questions raised by the reviewer.

- The values C_r and C_c describe a learner's relative importance to mimic the teacher's demonstrations and following its own preferences, respectively, and are thus properties of a learner and not the parameters of a teacher.
- The performance of AWARE-BIL decreases for increasing learner's constraints because the learner's preferences to avoid certain cells conflicts with the goal to go to certain cells to accumulate rewards. Note that this decrease is due to the experimental setup and not due to limitations of AWARE-BIL.
- δ_r^{hard} and δ_r^{soft} are used to characterize a learner's reward feature matching behaviour as part of the learner's optimization objective: While a mismatch of up to δ_r^{hard} between the learner's and teacher's reward feature expectations incurs no cost regarding the optimization objective, a mismatch larger than δ_r^{hard} incurs a cost of $C_r \cdot \|\delta_r^{\text{soft}}\|_p$. Please also note that δ_r^{hard} is a fixed parameter, while δ_r^{soft} is an optimization variable. In equation 1, m is the number of preference constraints of the learner. In general, $m \neq d_c$. Note that we have a typo in the paper in line 108 which might have caused some confusion: we incorrectly wrote $\delta_c^{\text{soft}} \in \mathbb{R}^{d_c}$ but we wanted to say $\delta_c^{\text{soft}} \in \mathbb{R}^m$. We will correct this typo and elaborate on the notation in the revised paper.
- $\delta_r^{\text{soft,low}}$ and $\delta_r^{\text{soft,up}}$ are auxiliary variables used to rewrite the constraints on the absolute value of the mismatch in a form more convenient for optimization. We will add clarification and more details to the revised paper.

Reviewer 2

We thank the reviewer for providing useful suggestions and high-level comments on the paper structure.

As suggested, we will remove the auto-pilot example from the introduction and elaborate more on the other two examples. We will also emphasize that the learner-aware teacher with full-knowledge of the learner allows us to formalize the problem and introduce a theoretical/algorithmic framework to study the limitations of learner-agnostic teaching. The real use-case of learner-aware teaching is for incomplete knowledge of the learner. We believe that in this paper we consider an important new direction for inverse reinforcement learning which we would like to make available to the community in a timely manner by a conference publication. However, we will revise the paper to include more details on the algorithms in Section 5.

Reviewer 3

We thank the reviewer for appreciating the novelty of the problem setting and providing suggestions for improvements.

Regarding linearity of the reward function. It is true that our results are currently for the linear setting. However, we believe that it is worthwhile to first thoroughly understand this setting. Moreover, as we don't constrain the feature maps ϕ_r and ϕ_c, the features we consider can be nonlinear functions of a set of "basic" features, which in principle makes it possible to accommodate quite general situations in our setting. Nevertheless, we agree that a natural next step is to investigate to what extent our ideas can be extended to nonlinear reward settings.

Regarding experimental evaluation on more realistic tasks. Generally, we agree with the reviewer's suggestions and believe that evaluating our algorithms on more realistic tasks is a natural direction for future work. We would like to reemphasize that the paper's primary focus is on introducing an important problem setting for IRL, developing algorithms for the problem, and empirically understanding the performance of these algorithms. We will further extend the experimental analysis in the paper as outlined in our response to Reviewer 1.