We thank the reviewers for thoughtful reviews and encouraging comments. We respond only to questions and concerns.

(R1) "Although the paper is generally clearly written … I really enjoyed this paper, so my comments mostly have to do with making the derivations a bit more readable.": Thanks for the helpful feedback. We will use it to *further* improve clarity. We will, e.g., include more in-words descriptions of definitions and results and be consistent on (s, s') vs (j, k).

(R1) "Finally, a question about identifiability …": Great question. If we impose additional assumptions on M, π_b in the definition of Θ in Sec 3.3 then the set shrinks. Tennenholtz et al. study a particular set of assumptions (that we discuss on page 8) that would make the set shrink to a *point*. The assumptions imply in a certain sense a view on every confounder; we work without these assumptions where policy value is *not* point-identifiable. Note that bounds (i.e., Θ) would only collapse if you impose the assumptions *a priori* on M, π_b – no method can automatically detect the validity of identifying assumptions as they must be imposed on the distribution of *unobserved* data. Will add this discussion.

(R2) "1. Why …": Asn. 2 is misnamed; we should rather attribute the term "memoryless confounding" to the special setting of Lemma 1. Asn. 2 is an assumption that it is sufficient to estimate a density ratio that is constant in s. For baseline UCs, marginalized occupancy distributions are understood to be marginalized over an initial state distribution on the baseline UC.

(R2) "2. While …": Lemma 1 (to be renamed "memoryless confounding") is just one simple setting where one can ensure Asn. 2. A practical example may be blood glucose control for diabetic patients, where s_t is blood glucose, a_t is insulin, and u_t are unobserved eating/exercise events reasonably modeled by a random arrival process (e.g., Poisson).

(R3) "One, …": In the paper we reference work that discusses how to choose a reasonable range of Γ; we will instead flesh out this discussion into the text for completeness. An analyst would have to justify an upper bound on how informative of selection an unobserved confounder can be; this can be benchmarked relative to the informativeness of observed covariates by dropping covariates and looking at the distribution of odds ratios for each covariate.

(R3) "Two, …": Most approaches to sensitivity analysis require making some untestable assumptions. Instead of assuming the most unrealistic untestable assumption of *no* unobserved confounding, we handle a case where *there is* unobserved confounding but with *structural restrictions*. Asn. 1 is a structural assumption of ergodicity and is necessary to make sense of infinite-horizon RL, whether with or without confounding. Asn. 2 assumes structure on how unobserved state variables interact with observed state and actions. Violations of Asn. 2 also violate Asn. 1. If, for example, nonstationary unobserved confounding (e.g., a single time point) is more plausible for the domain, then our approach (and other approaches based on stationarity) may be inapplicable. Will mention this and cite the suggested Namkoong et al. reference regarding single-time-point nonstationary/finite-horizon confounding.

(R3) "Do these put significant constraints on what the evaluation policy can be?": Not if the MDP is ergodic as is often assumed for infinite-horizon RL (meaning induced chain is ergodic under any deterministic policy). In infinite-horizon RL, we usually do not deal with MDPs that induce ergodic chains under one policy but not another. We stated our Asn. 1 in a minimal way since we only really need this for π_e, π_b but the spirit is that the MDP is ergodic as common for infinite-horizon RL. Will add this explanation and the stronger version of ergodic MDP.

(R3) "Three": This is a **mischaracterization**: we provide ***both*** globally optimal and heuristic approaches. We will clarify this in the final text. Prop. 3 provides a disjunctive program formulation that, as we say on line 184, can be solved directly using branch-and-bound (e.g., Gurobi). In the experiments, following our conclusion in line 184, we solve Eq. (10) directly in Gurobi (via branch-and-bound with global optimality certificates on bilinear variables). We further discuss this in appendix line 735. Alg. 1 is provided as a heuristic to tackle large state spaces, and in Fig. 8 of the appendix we compare the bounds computed by Alg. 1 vs. Gurobi. We will better advertise these results and clarify.

(R3) "empirical results would benefit from verifying both Assumptions": Definitely; we'll comment on this and explain. Asn. 1 and 2 both hold by construction of the experimental settings. The chains are ergodic for π_e, π_b and the confounders satisfy the sufficient condition in Lemma 1.

(R3) "The related works …": We'll clarify Zhang & Bareinboim and cite Namkoong et al.

(R4) "relies on the discrete nature of $\mathcal{S}$ (which might be okay) …": We focus on tabular because it is most illustrative and is very central to RL, but as Remark 2 and Appendix D.1 show all of our results still apply if $w(s) = \theta^T s$ (where s can be embedded arbitrarily). Tabular is the special case where $\mathcal{S} = \{(1, 0, \ldots, 0), \ldots, (0, \ldots, 0, 1)\}$. Indeed going beyond tabular in RL always requires some function approximation. Rather than further complicate the text, we propose to more explicitly flesh out the (mostly straightforward) generalization in the appendix.

(R4) "state more clearly the computational complexity": Each step of Alg 1 requires solving two LPs that have size $|\mathcal{S}|$. LPs are generally considered very easy. We will cite generic theoretical worst-case complexity bounds for LPs, which while polynomial are not considered representative of their practical difficulty. The branch-and-bound procedure used by Gurobi is finite-time but not guaranteed to be polynomial. In practice it does very well, solving in seconds for examples in the paper. We will cite and point to work on the *practical* tractability of integer programming.