We thank the reviewers for their thoughtful feedback. We are pleased that the reviewers found the multi-criteria preference learning (MCPL) formulation interesting and the theoretical analysis insightful. We begin by clarifying some common concerns, including how the the user study is more nuanced than it seems, eliciting target sets (and how even pre-specified ones are useful, as our study shows), and why the passive sampling model is a natural first step towards an understanding of our framework. We would also like to highlight to R4 that our game-theoretic approach to the MCPL problem provides a systematic way to select amongst Pareto-optimal alternatives by using target sets.

**Scope of user study** [R1, R2, R3, R4]. While the reviewers appreciated the inclusion of a real-world user study, they raised some valid concerns regarding its scope and design. While we agree that the study with 5 base policies, 5 criteria and 50 participants is a small-scale study, we would like to highlight that it was already complex enough to capture a wide variety of behaviors: for example, a strict preference ordering in the overall comparison, as well as circularity in preferences along criterion C4 (conservativeness). Moreover, in spite of the limited sample size, our statistical tests demonstrate that our hypotheses are indeed supported within a $95\%$ confidence interval. Lines 912-917 further highlight some of the challenges that we faced in designing such a user study along with our adopted solutions. As mentioned by R2, we do not know of any existing benchmarks to evaluate MCPL methodologies. Our user study is a first step towards establishing a systematic benchmark for this domain and we shall make our anonymized data publicly available.

**Choice of target sets** $S$ [R1, R3]. Our framework currently puts the onus of the selection of the target set $S$ on the designer. This is applicable for a range of tasks like medical diagnosis wherein domain experts are required to assess these multi-criteria trade-offs (as we did in our user study). However, we agree with R1, R3 that several applications would benefit from a user-elicited target set. Classically, such elicitations have been studied in the economics literature and as stated in our Discussions section, studying such mechanisms is an important future direction.

**Passive sampling model** [R1, R3]. Our statistical analysis focuses on the passive sampling framework wherein each query is sampled uniformly at random. This framework captures several scenarios (including our self-driving case-study) wherein each user completes a questionnaire comprising *all* the comparison queries. In contrast, an active framework is useful for scenarios where some queries are easier to obtain, and understanding it from a theoretical perspective would require techniques (e.g from the Bandits literature), which is an interesting direction for future work.

**Reviewer R1.** – (**Novelty**) Our work formalizes the MCPL framework, proposes a game-theoretic solution concept (called the Blackwell winner) and provides a thorough theoretical analysis to understand the statistical and computational properties of this winner; the framework and the analysis together comprise our novel contributions. We will take R1's advice and add more algorithmic content from the appendix into Section 3.3.
– (**Theoretical Results**) Note that our upper bounds in Theorem 1 are for general $\ell_q$ norms and general target sets. While our lower bounds focus on the $\ell_\infty$ norm, Proposition 2 establishes an asymptotic lower bound for a wide class of target sets. In addition, we believe that our lower bound constructions are quite informative and are likely to be of independent interest, for instance, in understanding the stability of the Nash equilibrium to sampling errors.
– (**Details**) (i) We chose the horizon $H = 18$ to be three times the planning horizon $= 6$ for our MPC policies, allowing them to exhibit a varied behaviour over the complete trajectory without affecting their planning capabilities. (ii) The std dev for the weights were quite small (order 1e-3). (iii) The Condorcet and Borda winners have been cited in Line 29.

**Reviewer R2.** We thank R2 for their encouraging remarks. We shall include the missing references in the updated draft.

**Reviewer R3.** – (**General learning distributions**) Our sampling framework, which considers uniform distribution over the entries, can be extended to *any* distribution with complete support. The statistical guarantees will worsen by a condition number factor $\kappa = {}^{\text{max. prob}}/_{\text{min. prob}}$ where max (min) prob is the largest (smallest) probability of sampling.
– (**Access to target set**) Our algorithmic framework requires access to a pre-specified target set. One way to accomplish this is to write these target sets as an intersection of a finite number of half-spaces that in turn capture different trade-offs between criteria. For general convex sets, the distance oracle may be computationally intensive.
– (**Details**) (i) The numbers in Figure 1 are not from an actual study but are illustrative of results from our MTurk user-study; we shall clarify this. (ii) We used cvx optimization package in Matlab to compute the Blackwell winner.

**Reviewer R4.** – (**Comparison with Pareto-optimal solutions**) We agree that this is an important comparison. As a solution concept for multi-criteria preference learning, the Pareto-optimal set is insufficient since it does not provide a mechanism for selecting amongst those alternatives. Our framework with target sets provides a systematic way to specify preferences amongst these various Pareto optimal solutions. Further, if such preferences are indifferent between the Pareto-optimal solutions, the set of Blackwell winners will coincide with the Pareto-optimal set: in particular, one can show that for any preference tensor $\mathbf{P}$, there exists a target set $S$ such that the Blackwell winners for $(\mathbf{P}, S, \|\cdot\|_\infty)$ can recover the complete Pareto-optimal set. We will add a discussion comparing our approach with those for finding Pareto-optimal sets. In addition, we would like to point that our idea of using target sets as a selection mechanism is not arbitrary; it arises as a natural extension from real to vector-valued games, dating back to Blackwell's work.
– (**Distance metrics**) Since our example focuses on the $k = 1$ setup, all $\ell_q$ norms are equivalent to the absolute distance function $|\cdot|$. In comparison, our upper bounds in Theorem 1 concern general $\ell_q$ norms.
– (**User study parameters**) We tried to extensively cover different trade-offs through our choice of linearization weights and target sets. We shall add an ablation study on the robustness of the solutions to variations in weights and target sets.