We thank all the reviewers for their careful readings and constructive comments.

Reviewer #1: Many thanks for appreciating our work.

Reviewer #2 (Re. Motivation behind the Regret definitions):
Note that following the RUM interpretation of MNL model (please see response to Rev. #3 for details), the score parameter θ_i of each item $i \in [n]$ essentially represents the mean utility/reward of item i, which in turn governs its preference relation w.r.t. the other items (based on the "Feedback type", Sec. 2.1). Thus our regret definitions (i.e. both *Winner* and *Top-k regret*, Sec 2.2), simply penalize the learner for pulling any suboptimal set in terms of the sub-optimality in its average utility-score w.r.t. that of the optimal set (which is a^* for *Winner regret*, and $S_{(k)}$ for *Top-k regret*) — an intuitive quantification of loss/value of a subset in terms of the underlying utility scores of its items.

Re. Applications: As discussed in the Introduction, some motivating applications of our problem lies in various kind of *partial monitoring frameworks*, e.g. launching new products, recommender systems, crowdsourcing etc., where value of a subset is measured in terms of the average utility-scores (θ_is) of its items, but the learner only gets to observe a preference feedback of the selected items drawn according to the MNL($\boldsymbol{\theta}$) model, $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_n)$.

Moreover, as we clarified in Rem. 1 and 2, for the special case of only two-sized subsets (i.e. when $k = 2$), our regret definition simply boils down to that of '*Dueling Bandit*' problem – an extensively studied and well accepted notion of regret in bandit-literature (Ref. [5,12,40-47]), which too is based on the concept of penalizing every subset (i.e. pair of items as $k = 2$) in terms sub-optimality of average item scores. In fact, the very few recent works that extends Dueling Bandits to subsetwise feedback (*Multi-Dueling bandits*), also use the same notion of regret as ours (see Ref. [11,39]).

Reviewer #3 (Re. Assumptions of the proposed models and practical relevance):
We have assumed Multinomial Logit (MNL) (alternatively known as Plackett Luce) McFadden and Train [2000], Luce [1959] as our subsetwise feedback model which is a widely used preference model in econometrics and social choice theory literature (Ref. [7],Soufiani et al. [2013]), specially for assortment selection problems (Refs. [2,3,4]), as well as in machine learning community, be that offline batch optimization (Ref. [23,29,39]), or online learning setting (Ref. [17,35, 40]) etc. In fact, even for the special case when subsetsize $k = 2$, the model is extensively studied as *Bradley Terry Luce* (BTL) model Negahban et al. [2012], Rajkumar and Agarwal [2014], Shah and Wainwright [2015], and its various extensions have also been considered Wen and Koppelman [2001], Yan et al. [2019] — thus MNL model is indeed one of the most well studied preference model, which has natural applications to various real world scenarios, e.g. customer preferences, recommender systems, voting methods, or more generally any application which aims to aggregate information from preferences over discrete choices. (see response to Rev. #2 for more applications).

For a more theoretical interpretation of MNL feedback model (Def. 1): MNL model belongs to the class of Random Utility Models(RUM), which assumes an underlying utility scores of the items $\theta'_i \in \mathbf{R}$ for each item $i \in [n]$, and assigns a conditional distribution $\mathcal{D}_i(\cdot|\theta'_i)$ for scoring item i. Upon receiving any subset $S \subseteq [n]$, the environment first draws a random utility score $X_i \sim \mathcal{D}_i(x_i|\theta'_i)$ for each item $i \in S_t$, and selects the winner item $J = j$ with probability of X_j being the maximum among all the scores of items in S, i.e. Winner Feedback: $Pr(J = j) \sim Pr(X_j > X_{j'} \ \forall j' \in S \setminus \{j\}) \ \forall j \in S$. Now it can be shown that when $\mathcal{D}'_i$s are Gumbel$(\theta_i, 1)$ distributions (Ref. [7],Soufiani et al. [2013]), i.e. $\mathcal{D}_i(x_i|\theta'_i) = e^{(x_j-\theta'_j)}e^{-e^{(x_j-\theta'_j)}}$, then $Pr(i|S_t) := Pr(X_i > X_j \ \forall j \in S_t \setminus \{i\}) = \frac{e^{\theta'_i}}{\sum_{j \in S_t} e^{\theta'_j}}$ — which precisely gives rise to the MNL choice model. (We used $\theta_i = e^{\theta'_i}$, $\forall i \in [n]$. Unfortunately due to space constraints we could not include this RUM interpretation of MNL model, which really sheds light into its specific mathematical form.)

We sincerely request the reviewers to kindly reconsider their scores based on the above clarifications.

References

R Duncan Luce. *Individual Choice Behavior: A Theoretical Analysis*. Wiley, 1959.

Daniel McFadden and Kenneth Train. Mixed mnl models for discrete response. *Journal of applied Econometrics*, 2000.

Sahand Negahban, Sewoong Oh, and Devavrat Shah. Iterative ranking from pair-wise comparisons. In *Advances in Neural Information Processing Systems*, 2012.

Arun Rajkumar and Shivani Agarwal. A statistical convergence perspective of algorithms for rank aggregation from pairwise data. In *Proceedings of 31st International Conference on Machine Learning*, 2014.

Nihar B Shah and Martin J Wainwright. Simple, robust and optimal ranking from pairwise comparisons. *arXiv preprint arXiv:1512.08949*, 2015.

Hossein Azari Soufiani, David C Parkes, and Lirong Xia. Preference elicitation for general random utility models. In *Uncertainty in Artificial Intelligence*, page 596. Citeseer, 2013.

Chieh-Hua Wen and Frank S Koppelman. The generalized nested logit model. *Transportation Research Part B: Methodological*, 35 (7):627–641, 2001.

Yongnan Yan, Xiangdong Xu, and Anthony Chen. Is it necessary to relax the iid assumptions in the logsum-based accessibility analysis? *Transportation Research Record*, page 0361198119839972, 2019.