We thank reviewers for detailed comments and suggestions. We will address all comments in the revision. In this work, we consider a novel unsupervised stochastic contextual bandits problem. In follow-on work, we will study relevant open questions like, lower bounds, private information, and real-valued feedback, pointed out by reviewers.

Novelty. The novelty of our work, as such, is a combination of novel modeling principles to account for unsupervised contextual sequential selection, as well as subsequent method and analysis, and experimentation. The earlier work (Verma et al. AIStats'19) considered the problem of learning an optimal action but ignored the contextual information. In this work, we incorporated the contextual information, which is readily available in many applications. Exploiting the *real-valued* contextual information (features) for improving the arm selection strategy is non-trivial due to the unsupervised nature of the problem where the standard analysis of contextual bandits does not apply. We made necessary modeling assumptions leveraging GLM models and extended the existing definitions to address the learnability issues in the new setup. However, the problem still requires new ideas and analysis methods to derive an efficient algorithm and poses new technical challenges for analysis.

Response to common comments of Reviewers 2 and 4:
`The idea might look incremental. What are the main challenges solved by this work? :`
We respectfully disagree. The new ideas and technical challenges addressed in this work are as follows:

New ideas: USS-UCB (Verma et al. AIStats'19) uses two-sided test derived from Eq. (6) and Eq. (7) to identify an optimal arm. In contrast, our work identifies that a one-sided test is enough to learn an optimal arm with contextual information. We exploit this idea to come up with simpler algorithms. The difference in tests used in earlier and our work becomes apparent by comparing arm selection strategy in USS-PD (line 9) with that of USS-UCB (lines 7-9) where two sets $\hat{B}_t^h$ and $\hat{B}_t^l$ and their interaction needs to be computed. However, this simplification in USS-UCB throws some challenges in the analysis that do not arise with two-sided test, but we carefully handle it (see lines 488-492).

Technical Challenges

1. GLM bandits are well studied but require reward or loss information. In the USS setup, loss of selected arm can not be observed; hence finding the optimal arm is challenging. We have shown that if problem instance satisfies contextual weak dominance (CWD) property, then the pairwise disagreement between arms can be used to estimate context-dependent disagreement probability, and that can be used to find an optimal arm for a given context.
2. Regret analysis of GLM bandits hinges on bounding the instantaneous regret in each round, which is tied to the estimation error of the GLM parameters. Due to the unsupervised setting and cascade structure, this way of bounding regret does not work in our setup. Our analysis goes by bounding the number of pulls of the sub-optimal arms. However, unlike standard bandits, we have to distinguish whether the sub-optimal arm pulled by USS-PD is on the 'left' or 'right' of the optimal arm in the cascade. It requires our analysis to carefully handle both the cases (see Lemma 5 and 6). Since USS-PD uses a similar MLE estimator for parameter estimation as in GLM bandits, we only adapt their asymptotic normality results, the other steps of bounds are new in our work.
3. Though it is not reported in our work, we did try several other models and analysis approaches to solve the USS problem with contextual information. However, due to the weak feedback structure of the problem, the other methods are not amenable for analysis. Our final presentation is a model and analysis that is clean and complete. For example, in the appendix, we point out that analysis based on Optimism in the Face of Uncertainty for Linear bandits (OFUL) method with a regularizer did not go through without making more assumptions.

Response to Reviewer 1:
`Lemma 1 should be equality, and does not require` $j > i$: Thanks for catching the typo. We will fix it.

Response to Reviewer 2:
`Experiments lacks comparison with baselines and other SOTA methods:`
To the best of our knowledge, we are first to consider the contextual USS problem, so there is no state-of-the-art (SOTA) method. In our experiments, we have considered baseline policies that select either a fixed arm or a uniformly random arm in each round (see Figure 1c). We also compare USS-PD's performance with the policy that can observe the true loss (see Figure 1b).

Response to Reviewer 3:
`In equation (2),` λ_{I_t} `is missing:` Thank you for catching the mistake. We will correct it.

Response to Reviewer 4:
`In [1], the exactly same weak dominance property is introduced and Theorem 1 is same:`
We **disagree** on both. In this paper, the WD property is context-dependent, whereas the contextual information is ignored in [1]. This new definition has nuances. First, Eq. 4 points to the fact that examples can be partitioned based on strength of CWD, a situation that does not arise in [1]. Additionally, we allow for instances to violate CWD property, and present algorithms that are agnostic to the presence of these instances. As for Theorem 1, although it bears similarities with the previous work but it requires adaptation to the contextual setting.