We thank all the reviewers for acknowledging the contributions of our work and providing insightful comments and suggestions. In the following, we address all the concerns of each reviewer in detail.

[Reviewer #1]

Reflect on the pseudo-label algorithm. We agree that a poor self-training procedure might reinforce errors. However, as we demonstrated both theoretically and empirically, given a fairly good base classifier, more relevant unlabeled data is helpful (cf. discussions in Section 2.1, L93, and results in Table 1), regardless of the data imbalanceness as well as base SSL methods. We remark that here, our main aim is to explore how semi-supervised techniques help in imbalanced learning. Further ablation studies on different SSL algorithms are also provided in appendices to give a complete picture. Altogether, we believe that these justify the value of SSL and our claims. We plan to include more discussions in the main paper to reflect on the usage of different SSL algorithms.

Compare to one other algorithm and ignore others. We would like to clarify that we actually did a comprehensive study (main paper and appendices) on the effect of the proposed techniques through **5 dimensions**: different (1) datasets, (2) imbalance ratio, (3) unlabeled data imbalance ratio, (4) balanced (imbalanced) training strategy, and (5) SSL method. The key point is to demonstrate that SSL as a technique is generally beneficial for both representative non-imbalanced methods (CE in Table 1) and imbalanced methods (LDAM-DRW). We believe the extensive tests over the above dimensions, together with consistent improvements in performance, confirm the impact of our contributions.

More detailed related work. We agree with you that the section on related work is relatively short. While we include most of the relevant literature, the discussions are not elaborated enough. This is largely due to the space limit. We will definitely include more details to provide an informed picture for the audience in the revised version.

[Reviewer #2]

Definition of relevance score. Thank you for pointing this out. We did not define it clearly in the main text, but have elaborated it in Appendix D.2 (from L541). We will bring the detailed setups to the main paper in the revised version.

Theorems w.r.t. relevance score. This is a good point. In our paper, to motivate our study and provide insights, the theorems mainly consider the imbalanceness for labeled & unlabeled data. When data relevance comes into the picture, it might not be easy to obtain a clean and rigorous mathematical formalization. We believe the current theorems are of their own interests, and should be insightful enough in understanding key components in imbalanced learning. On the other hand, including other components such as relevance, is certainly an interesting future direction to consider.

Self-supervision vs. irrelevant unlabeled data. Section 3 mainly motivates SSP by presenting the potential practical issues of unlabeled data. For self-supervised methods, we would like to clarify that our key theme is to investigate the two perspectives on the dilemma of imbalanced labels. Hence, we decompose the usage of semi- & self-supervised learning, and analyze each of them *separately* on how they improve imbalanced learning. As such, in our experiments of SSP, we do not consider extra unlabeled data. However, we do appreciate the idea of combining both to further improve results. As you noted, one should be always aware of the potential risk of irrelevant unlabeled data in practice.

Title & Theorems towards complex models. We intended to let the title reflect the main perspectives of our investigation on the imbalanced labels: the positive and the negative side of the dilemma. We certainly welcome any suggestions. For the theorems, these are excellent points. We would be interested in a fully general theorem. A thorough analysis of every aspect is likely hard and beyond the scope of current manuscript, but is definitely an important future direction.

[Reviewer #3]

More analysis on different label distributions. This is a great suggestion, and we believe a comprehensive study is broadly valuable in real-world applications. In fact, as a first step, we have made efforts to address it by considering different imbalance type of the labeled data. The main paper focuses on long-tailed distribution which is most common in literature, but we also further study the step imbalance distribution in Appendix F.3. We remark that within both imbalance types, we investigate various distributions by changing the imbalance ratio. The conclusions are consistent across those studies. We will include a more clear summary of our results in the revised version, and remark on the practical importance of further analysis on different imbalanced distributions.

[Reviewer #4]

Claim on SSL demonstrating the value of labeled data. This is an interesting perspective. To clarify, let us elaborate our viewpoint provided in the framework. We regard the labeled data as providing useful information and positive guidance for the unlabeled data, i.e., offering pseudo-label. This in return, as we consistently demonstrated, helps the overall learning. In this sense, the labels indeed contain positive value. Further, we remark that the labeled data may not be viewed solely or exactly as "unchanged controlled variable". In our experiments, we also actively study the amount of labeled & unlabeled data (Appendix E.3), i.e., the amount of "value" or supervision the label provides. We will provide a clear summary to distinguish our perspective on the positive value of the imbalanced label.

Discussion between cRT and the proposed SSP. Your general feeling is correct: in a sense, SSP also aims to learn better representations that are more agnostic to the imbalanced label bias. As mentioned in our response to Reviewer #1, with more space in the next revision, we will definitely include a more detailed discussion in the related work.