We thank the reviewers for their thoughtful and valuable feedback. We appreciate their time and effort, especially given the current uncertain times.

There are no factual inaccuracies in the reviews that we would like to correct.

We agree with the reviewers that the contributions of this paper are mostly theoretical in nature. However, the class of algorithms that we study, top-down induction of decision trees, is of significant practical relevance, and we believe that it is interesting and important to put these algorithms on firm theoretical footing. The decision tree heuristics ID3, C4.5, and CART (and related ensemble methods) all enjoy empirical success in machine learning practice, but there are still relatively few works giving rigorous guarantees on their performance. A natural first step is to analyze these heuristics and their variants within the PAC model and under feature and distributional assumptions.

We are hopeful that our results and techniques point to concrete avenues for future work that will further close the gap between theory and practice. Regarding our feature and distributional assumptions in particular, there are analogues of noise sensitivity for categorical features and general (non-product) distributions, and therefore one could design and analyze natural extensions of our algorithm to such settings. For future work, it would be interesting to extend our provable guarantees to these more general settings, and to experimentally evaluate the performance of such generalizations of our algorithm on practical datasets.

Finally, we thank the reviewers for their specific suggestions for improving the presentation of our paper. We agree with their suggestions and will implement them: we will add more details to Remark 1, and we will also specify the choice of δ in the statement of Theorem 1.