We thank the reviewers for their overall positive and encouraging feedback.

Responding to the main comments raised by the reviewers:

- Regarding the simplified nature of the target class of tree-structured Boolean circuits: our main goal in the paper was to suggest a crisp property of “natural” distributions that turns them “easy” to learn. We chose the class of tree-structured Boolean circuits since this family is rich enough to reveal interesting properties about learnability using gradient-based algorithms, yet is simple enough to give thorough theoretical analysis. Showing that the local correlation property plays a role in learning broader families of functions is an exciting topic for future research.

- On the assumption of known tree structure: we believe that in the context of learning neural networks, knowing the structure of the tree is a realistic assumptions. Neural networks in practice are often designed to have a structure that reflects the prior knowledge of the problem (for example, convolutional networks, LSTMs, transformers etc.).

- Regarding read-once formulas: we will add a comparison to the literature of learning read-once formulas. However, we note that our focus is not on PAC learnability of read-once formulas in general, but rather on understanding distributional properties that allow gradient-based algorithms to converge to an optimal solution.

- Regarding the analysis of layerwise training algorithm: we believe that similar results may be shown for end-to-end optimization, at the expense of making the theoretical analysis much more complicated. End-to-end gradient-descent is notoriously hard to analyze in the case of deep networks, due to the highly non-convex nature of the optimization. We chose the less popular layerwise optimization to achieve a simpler theoretical analysis, and yet show guarantees that reflect the behavior of gradient-descent in practice.

We will correct the final version of the paper according to additional comments and suggestions raised by the reviewers.