We thank the reviewers for their constructive comments. Responses to the questions raised follow.

Reviewer #2:

1. **Going beyond the pattern:** While most of our applications indeed follow the pattern described by the reviewer, it is certainly not necessary for hypothesis set stability analysis. The distillation application in section 5.4 is one example which does not follow the pattern. Similarly, one can modify the application in section 5.3 by setting $\mathcal{H}_S$ to be any γ-Lipschitz function class (i.e. not necessarily a linear class as currently written), and the bounds follow verbatim. Similarly, the algorithm that selects the base hypotheses does not need to be uniformly stable: for example, in the bagging application of section 5.1, we get non-trivial generalization guarantees even if the base algorithm is *not* uniformly stable, as noted in line 235 of the paper.
2. **Related work comparison to PAC-Bayes bounds.** Thank you for elucidating the nuances of the prior work on PAC-Bayes bounds. We will certainly elaborate more on the points you raised in the next version of the paper.
3. **Typo corrections:** Thank you, we will fix the typos in the next version. We will also define sensitivity – it is indeed essentially the same as bounded differences. We will also add more details on the comparison of our bagging bounds with those of Elisseeff et al (2005).

Reviewer #4:

1. **Impact:** The impact of this paper can be judged from the contributions listed in section 1.1; additionally, we believe that our paper provides foundational work on analyzing generalization in data-dependent hypothesis sets.
2. **Audience:** We expect that any theoretically-minded ML researcher working with data-dependent hypothesis sets would find our paper interesting, and engineers may be able to use insights from our paper to design learning algorithms with good generalization properties. We therefore expect our paper to appeal to a large audience.

Reviewer #6:

1. **Invariance under sample permutation:** We indeed implicitly assume that $\mathcal{H}_S$ is invariant to permutation of samples. We will mention this explicitly in the next version.
2. $\hat{\mathfrak{R}}_T(\mathcal{H}_{S,T})$**:** Although the standard definition of empirical Rademacher complexity is for hypothesis sets that are not data-dependent, the definition still remains valid for data-dependent hypothesis sets. In particular, $\hat{\mathfrak{R}}_T(\mathcal{H}_{S,T})$ indeed coincides with the empirical Rademacher complexity of the hypothesis set $\mathcal{G}$ for the empirical sample T, with $\mathcal{G} = \mathcal{H}_{S,T}$, and we will further clarify this in the next version of the paper.
3. **Sensitivity defintion:** We will define sensitivity at an appropriate spot before section 5.2.