We thank all reviewers for their efforts in reviewing our paper, and for the helpful comments and suggestions.

Reviewer 2:

- *"Id like to see some statistical properties of these new loss functions, and how they are compared to the existing ones"*:

  We show that our Bregman-based approach leads to proper bi-tempered losses, whereas the previous Tsallis-based losses are not proper.

- *"The authors are encouraged to present more theoretical & empirical analysis on the robustness of these loss functions... If this new non-convex function is used, can we still theoretically analyze the performance of the learning algorithm?"*:

  We provide strong empirical evidence of the efficacy of our method, i.e. for dropping the convexity in the activations of the last layer. Proving theoretical convergence under our non-convex losses for a multi-layer neural net is formidable and beyond the scope of the paper.

  However a plausible next goal would be to prove theoretical bounds for a single neuron using our non-convex tunable loss. We leave this up to future work.

Reviewer 3: We will add the additional references you suggested. Thanks!

- *"There are many possibilities for combining heavy-tailed distributions and robust/consistent divergence. It is nice to discuss the advantage of the proposed bi-tempered loss: computational cost? mathematical beauty?"*:

  Our method leads to proper losses which is important for many applications that require a probability distribution as output. The computational cost is only negligibly larger than the cost of logistic regression. The foundation of our method is based on generalizations of the exponential family of distributions. The crux of our method is the fact that even in the "many class" case we only have two additional parameters to tune.

Reviewer 4: Thanks for your excellent feedback. We will expand on the relationship to the additional references and more clearly contrast our losses with the previous Tsallis-based versions.