We thank the reviewers for both their favorable and their critical feedback. We have used the latter to improve and clarify the paper. Below we group and respond to key issues, highlighting the changes we've made:

1. Tasks 1-3 are perhaps too simple (R1). We agree that tasks 1-3 are simple in that they are easy to solve with existing methods. We included them not to showcase the real-world potential of HNNs but simply to allow for a scientific investigation of their basic properties. Task 1 was used to show that optimizing the symplectic gradient of a neural network is a viable and stable strategy. We then asked whether an HNN could learn a nonlinear vector field, and Task 2 provided a sanity check that it could. We then asked whether an HNN would learn something sensible from real-world data, and Task 3 showed that it did. Once these sanity checks were complete, we applied our approach to the more complex and exciting Tasks 4 and 5 which we believe better showcase the power of HNNs. While we could remove tasks 1-3 due to their simplicity, we feel the rigor and clarity are improved thanks to their inclusion.

2. Hamiltonian is not generally proportional to $q^2 + p^2$ (line 19) (R1). This part was poorly written: we had not intended to claim that the Hamiltonian will be generally proportional to $q^2 + p^2$. We've updated the text as shown in a in the figure below.

3. Fix issues with structure: Table 1 is before tasks 4 and 5; Figure 6 not found (R1). We agree that these parts were a mess. Table 1 has been moved to page 7 of the paper so that the reader can view it after reading the explanations of Tasks 4 and 5. Table 1 was poorly formatted; we have added confidence intervals and updated it to be much more readable. These changes are reflected in c in the figure below. As for Figure 6, it has been moved to the Appendix and is now more clearly referenced as such as shown in b in the figure below.

4. Add references for optimizing gradients of neural networks (R2) Thanks for the five references. All are great additions to the paper and we have added them along with a discussion of the overlap (see part d in the figure below).

5. Show a simple control example, especially with the latent coordinate space (R3). Great idea! We believe that applying HNNs to control tasks will produce compelling results. However, to limit the scope of this paper to that of introducing the core concept of an HNN, we decided to leave this application to future work.

6. Model systems with (a) friction/damping or (b) contact modeling (R3). Thanks for these suggestions. When we were writing the paper we discussed some related experiments: how to model (a) a damped harmonic oscillator and (b) balls bouncing in a box. Ultimately, we decided to focus this paper on a thorough and principled investigation of the key effect and leave these more complicated scenarios to future work.

7. Outline applications where the equations of motion are not sufficient/where HNNs are needed (R2). Address significance of HNNs, especially for solving real-world problems (R1). One key difference between HNNs and potential energy-based approaches is that the latter generally require a reference energy obtained via electronic structure calculations (Equation 10 in Behler 20ll and Equations 1,2 in Pukrittayakamee 2009). Meanwhile, HNNs are trained in an unsupervised manner: we do not require reference energies. Because of this, HNNs are promising for datasets that have unusual coordinate systems (e.g. Task 5: Pixel Pendulum) such that the reference energies are not trivial to compute. We are very excited about real-world applications of this technique. Since releasing a preprint of this paper, we have heard from one group that is currently running experiments with HNNs to calibrate quantum computers and another group that is considering using HNNs to learn/calibrate the gaits of a biped robot.