Thank you all for your helpful comments on our Comp Neuro paper. Due to space constraints we restrict attention to major comments below.

Thanks to Reviewer #1 and #4 for pointing out that behavioral work in cognitive science suggests that people indeed use GPs, and for directing us to some relevant references, Reviewer #4. We will include the behavioral Cog Sci work in the revised version of the paper and expand on it in the supplement. Missing to acknowledge this work has been to our own disadvantage as this strengthens the premise of our paper that the brain employs GPs for function learning. It is not merely a principled approach of what the brain should do, but indeed accounts for a wide variety of experimental results [Lucas et al., *Psychon Bull Rev*, 2015]. We are hopeful this dispels the doubts Reviewer #2 had regarding the basic assumption of this work.

We regard our paper primarily as a contribution to Comp Neuro, hence the focus on the BioNN as how a biological brain could implement Bayesian inference for regression. Although some of you apparently would have preferred a pure machine learning paper and greater focus on the ANN, we consider the finding that the ANN can produce superior results than VFE, PBP and MC Dropout on the UCI datasets as a surprising secondary finding, that is however of potential interest to the broader NeurIPS community. We thank Reviewer #3 for pointing out the superior results of deep GPs, which is, given that our ANN was derived from a standard "flat" GP, not unexpected. Whether a deep GP can similarly be mapped to an ANN and trained using standard deep learning techniques to further improve performance could be an interesting question for future work.

Our experiments follow the didactics used in the methods section, moving from the initial case of known hyper-parameters, to learning the noise variance σ^2 and signal strength s, to finally learning all hyper/variational parameters including the length scales $\boldsymbol{l}$ and inducing point locations $\boldsymbol{Z}$. Reviewer #3 correctly noticed, or assumed, that if the latter were not learned, they were taken from the VFE method (we will clarify this in the revised version), because we were interested in how the approximations needed to render the network biologically plausible affect its performance compared to VFE. We will test the learning of all hyper parameters in the larger scale experiments and add it to the supplement of the revised paper. If the results of Fig. 5 are indicative, this could further improve the results.

The stochastic online learning rule is used because this is the biological scenario, not because of a large number of $(\boldsymbol{x}_i, y_i)$ training pairs. The datasets we considered consist of fixed finite size samples from some generative process, therefore we performed multiple passes over a dataset, revisiting the same data points, cf. Fig. 2. Our update rules for the BioNN (that implements sparse GP regression) readily apply to the more biological setting where every data point is a new sample from the underlying generative model. We will illustrate this in the revised version, either using the generative model of the Snelson data, or using a 1D dataset with non-uniform input distribution, as suggested by Reviewer #3. The supervised training phase is depicted in the somewhat busy Fig. S2. The learning rule has access to the training outputs only in the form of the prediction error that the activity of the postsynaptic neuron encodes.

While we disagree with Reviewer #2's opinion that the connection between neural regression and GPs is completely arbitrary, we agree that it is important to confirm that the approximations introduced for sake of biological plausibility still result in a BioNN that approximates the full GP well. We do so by considering the KL divergence to the full GP in Fig. 3 and Table 2, which shows that our BioNN approximates the full GP not quite as well as VFE, the starting point of BioNN's derivation, but better than FITC.

We appreciate that Reviewer #4 values the application to a water maze task, the biological task that due to lack of space did not make it into the main text. We will greatly expand on this section in the revised version to provide more information and better explanation.

For reproducibility and potentially increased impact of out work, we will release the code to reproduce all figures and tables of the paper. We actually have cleaned up the code for public release already, but failed to do so in time for the supplementary material deadline.