We thank all our reviewers for their feedback! We will respond to (R2, R3) separately to R1 due to different concerns.

R2 and R3: We thank R2 and R3 for their vote of confidence and giving this work at high score of 9 and 8 respectively. It means a lot to us – to see our ideas accepted by our peers at NeurIPS who also believe that our "work opens many new avenues for research and is sure to be widely cited" (R2). We agree with R2 that the importance behind random single weight choice, rather than single weight should be explained clearly in the intro, and we will update the draft to do so.

We experimented with setting all weights to a single fixed value, e.g. 0.7, and saw that the search is faster and the end result better. However, if we then nudge that value by a small amount, to say 0.6, the network fails completely at the task. By training on a wide range of weight parameters, akin to training on uniform samples weight values, networks were able to perform outside of the training values. In fact, the best performing values were outside of this training set.

The work that R2 pointed out, Zador2019, is indeed an inspiring critique of deep learning from the neuroscience community, and our work shares similar motivations. We will cite and discuss this work in our revised paper. As an aside, the *Animal AI Olympics* contest and related *Learning Transferable Skills* workshop, from same organizers, at NeurIPS2019 will discuss similar themes and we are excited to see more ideas in this direction from both communities.

We agree with R3 that scaling up is the next step. Currently we are exploring hybrid approaches (in the direction of *arxiv:1904.01569* R3 mentioned that employs 3x3 convs and random graphs) and indirect encodings (i.e. HyperNEAT, Stanley2009) to scale WANNs architectures to scales able to compete on benchmarks such as ImageNET and Atari. We wish to take the time to conduct this investigation thoroughly, and plan to report the findings in a follow up paper. Qualitative analysis of sample networks was done in Section 4, and in future work we plan to develop a method of automated quantitative analysis, as suggested by R3, to identify promising structures, modules, and motifs. The 'delete' operation suggested by R3 would have interesting effects but in the end we tried to present the simplest algorithm we thought could still work – though have no reason to think that a delete operation couldn't make even more compact WANNs. We would also like to thank R3 for the other minor suggestions, we will clarify the labels and information.

R1: We thank R1 for their critical review – that the contribution of this work is not well-defined by performance metrics makes it all the more important that its value is clearly expressed. We believe it is important for us to communicate and emphasize the merits of this work beyond benchmark scores.

The point of the paper was *not* to propose a new RL algorithm which meets or beats SOTA results, but to explore the importance of biases in neural network architectures. Our goal is to provide an existence proof that neural networks can be automatically designed to encode such biases. The motivation was not to outperform weight training algorithms but to explore an orthogonal approach – "how well can we perform *without* weight training?" We will emphasize the exploratory nature of the paper more heavily, in the hopes that it will be read in this frame of mind.

In the spirit of this extreme experiment the algorithm used was purposefully kept simple. It is not a lack of understanding from R1 that the technical contribution of the EA is minimal – this was precisely the intention. The innovation introduced in our method is to evaluate networks with sampled shared weight values rather than optimizing the weights, an approach we have not encountered in previous work. In the interest of clarity we will add pseudocode of the procedure used to search for WANNs in the appendix and emphasize in the text that we are presenting a purposefully minimal method.

The car racing experiment was conducted to show how WANNs can work with existing deep learning approaches, to help bridge different fields. As for concerns raised about the effectiveness of the network architecture discovered, we demonstrate comparable performance to [32] that used both a VAE and RNN, while in our work we used the VAE alone. In [32], it was shown that a VAE-only approach does not perform well with a 1 or 2 layer controller, where we have shown that the WANN was effective for this particular task using VAE-only.

Our original intention was to focus only on continuous-control RL experiments, and decided to run MNIST "for fun" near the end of the project. We could have confined the paper to only RL experiments (most RL papers don't run MNIST experiments), but chose to report the MNIST results to highlight both the benefits and limitations of the approach. We also think that a result of 80-90% (whether good or bad) with a randomly initialized network is interesting, especially compared to chance accuracy. Even with the MNIST section omitted, we believe the paper with only RL experiments still warrants a score above 3, a clear rejection score, as the contributions are valuable to the NeurIPS community.

Finally, we do believe there is a connection to the neuroscience field. In addition to the literature already cited in the paper, we see similar ideas circulating in the neuroscience community: R2 pointed out a recent neuroscience paper, "What Artificial Neural Networks can Learn from Animal Brains" (Zador2019) whose central theme is that "The first lesson from neuroscience is that much of animal behavior is innate, and does not arise from learning. Animal brains are not the blank slates, equipped with a general purpose learning algorithm ready to learn anything, as envisioned by some AI researchers; there is strong selection pressure for animals to restrict their learning to just what is needed for their survival." Our work is strongly motivated towards these goals of blending innate behavior and learning, and believe it will help bring neuroscience and machine learning communities closer together to tackle these challenges.