Please allow us to start by thanking the reviewers and area chairs for their consideration and time. Before providing rebuttals to specific concerns, we would like to emphasize a major contribution of this work, which was perhaps overlooked by some of the reviewers. Specifically, **a core strength of this work is our theoretical demonstration of an exact relationship between updates computed by backpropagation of error and those computed by a learning rule using local targets**. The demonstration of such a relationship (and the conditions under which it holds) has eluded researchers thus far, as also articulated by Reviewer 3. Considering the importance of local target-based learning for both biologically plausible learning and neuromorphic chips, this theoretical development is inherently valuable.

Let us now address the concerns of our reviewers. In order to provide an integrated response, we have grouped the major concerns into three main categories: the use of exact inversions, restrictions surrounding orthogonal weight matrices, and finally the depth/sophistication of the models tested.

The use of exact inverse operations, which allow accurate inversion of our targets from output to hidden layers, was a common concern among our reviewers. The biological plausibility of such operations is undeniably questionable and inverse models in the brain would require plausible learning mechanisms (e.g. via an auto-encoder-like training process or an approach such as that described by [1]). However, **our use of exact inverses was motivated by two main desires: a clear theoretical exposition, and a robust comparison against the baseline (target propagation)**. Exact inverses allowed us to theoretically develop the aforementioned relationship between backpropagation and GAIT-propagation. Furthermore, we could isolate the performance differences between the traditional target propagation and GAIT-propagation methods, establishing an upper bound of performance without the requirement for expensive parameter tuning of an auto-encoder for the inverse model (just as the authors of [2] make use of perfectly symmetric weight matrices when testing the ideal conditions of their biologically plausible model). Without this rigid structure, comparisons of these methods (both theoretically and empirically) would have been made opaque.

Justifications of our approach aside, we agree that, ultimately, the methods we described ought to be recreated with entirely biologically grounded components for maximum impact and utility in the computational neuroscience community. Such a biologically grounded implementation would include plausible and effective implementations of learned inverses, mechanisms for weight matrix orthogonalization, and online learning. We consider our current work to provide an ideal theoretical basis suitable for the NeurIPS community and are excited as authors to explore biologically motivated extensions in the future.

Our discussion of plausibility naturally leads us to another criticism (see reviews 1 and 4): orthogonal weight matrix requirements for GAIT-prop and their enforcement. On this account we could have been more explicit regarding the empirical observations made of GAIT-prop. **We observe empirically that a weak orthogonality regularizer is sufficient for high performance of the GAIT-prop training approach.** Our submitted Supplementary Material can be examined in order to confirm that our parameter sweeps show the stabilization of network training and achievement of competitive performance with GAIT-prop when using a relatively small orthogonality regularizer. We would therefore suppose that known decorrelation mechanisms of the brain would be sufficient for good training performance in practice. Decorrelation could be integrated in future studies by taking inspiration from existing work in computational neuroscience literature [3]. Furthermore, we observe that even without orthogonal regularization/orthogonalization shallow networks and networks with reducing width can successfully be trained (see green lines labelled "GAIT-prop" in Figures 2 and 3 of our submission). We believe that this resolves the reviewer criticism on this point.

Finally, Reviewers 2 and 4 highlighted limitations in the complexity of our datasets and depths of the networks we trained. On the account of model depth, we wish to draw attention to our Supplementary Material in which **we showed robust training of networks by GAIT-propagation up to depths of 6 and 8 layers.** We would expect that in much deeper networks, and without strictly orthogonal weight matrices, there would be impaired learning and we are happy to make this clearer in the text. However given the neuro-centric approach taken here, we must also consider that extremely deep feedforward neural networks do not reflect the structure of biological neural networks. For example, visual processing occurring between 50 and 100ms of stimulus presentation is largely feed-forward but involves a limited number of cortical 'layers'. Reviewer 4 also mentioned the extension of our approach to alternative architectures. On this note we are excited to see further theoretical developments in this direction, however our aim in this paper was to reconcile current developments in target-based learning, which are largely explored in feed-forward neural networks.

[1] Mohamed Akrout, Collin Wilson, Peter C Humphreys, Timothy Lillicrap, and Douglas Tweed. Deep learning without weight transport. *ArXiv*, 1904.05391, April 2019.

[2] Alexandre Payeur, Jordan Guerguiev, Friedemann Zenke, Blake A Richards, and Richard Naud. Burst-dependent synaptic plasticity can coordinate learning in hierarchical circuits. *bioRxiv*, 2020.03.30.015511, March 2020.

[3] Paul D King, Joel Zylberberg, and Michael R DeWeese. Inhibitory interneurons decorrelate excitatory cells to drive sparse code formation in a spiking model of V1. *J. Neurosci.*, 33(13):5475–5485, March 2013.