We thank the reviewers for their feedback and suggestions. Below we will clarify the points raised by the reviewers.

Novelty. Our approach connects MDP homomorphisms with equivariant networks, introduces a novel way for constructing such networks and uses them to speed up learning in RL. We are encouraged by the positive feedback of the reviewers regarding the novelty and method.

Scalability. To **R1**, **R3** and **R4**: Scalability is not directly an issue, largely because the most expensive step, the equivariant basis construction, is performed only once, prior to training. To **R3**, regarding matrix inversions: The transformation matrices we used are orthogonal, so that we can take the cheap transpose. To **R1**: A truncated SVD could provide a reasonable approximation if the one-time cost of construction is prohibitive. To **R1** and **R4**: We do not encounter issues within the transformation groups we consider. For large groups such as permutation groups we note that the number of filters scales linearly with the size of G, as does the number of input channels for the filters. For very large weight matrices, finding the SVD is computationally expensive.

Data augmentation. We thank **R3** and **R4** for suggesting additional comparisons to data augmentation. Per **R3** and **R4**'s suggestion, we ran two data augmentation baselines. The first data augmentation is designed to be a direct port of supervised learning to RL, akin to **R3**'s suggestion: Each state image is randomly transformed or not. If it is transformed, the output is correspondingly transformed. The second data augmentation is an equivariant version of (Kostrikov 2020), where both state and transformed state are input to the network. The output of the transformed state is appropriately transformed, and both policies are averaged. We show results on 4 random seeds for Pong in Figure 1. While data augmentation is beneficial in RL, our approach outperforms both variants. This is consistent with other results in the equivariance literature (see e.g. Worrall 2017, Winkels 2018, Bekkers 2018, Weiler 2018). Data augmentation can benefit RL because it *encourages* symmetries by increasing the dataset, on the other hand, equivariance *enforces* them, so the network does not need to learn the symmetries. We will incorporate the comparison and a discussion in the paper.

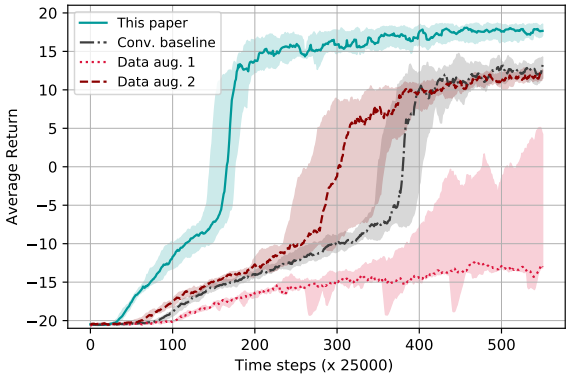

Figure 1: Data augmentation baselines for Pong.

Network construction. To **R2**, regarding ambiguity about network construction. We will improve the explanation in the appendix and add a short summary to the main paper, and examples will be included in the released code. To clarify, the representation of the group in the intermediate layers can be chosen arbitrarily. Our proposed solution works for any discrete group (as shown to work best in Weiler 2019), but other choices are definitely possible.

Other environments. To **R1**, **R2** and **R4**: We focus on CartPole, grid world and Pong, because these environments provide varying levels of complexity while still being compact enough to allow us to run a grid search and multiple baselines across environments with different observation spaces and symmetry groups. Our approach is in theory applicable to any RL problem that exhibits discrete group symmetry. Thus, this method is certainly applicable to more complex Atari games that exhibit symmetry. Based on the suggestion by **R1**, **R2**, **R4**, we are currently evaluating on Breakout, a more challenging Atari game. Experiments are currently running but exceed the length of the rebuttal period. To **R1**, our method is indeed also applicable to DM control for vision, as it exhibits flip symmetry.

Clarifications. To **R2**: We use nullspace/random baselines to show that equivariance is key to improving performance. To **R1**: While nullspace/random perform similar to the regular baseline for the other two environments they perform better on Pong. We expect that this may be related to different gradient dynamics when using basis networks, which could influence learning. In all cases, equivariance performs best. To **R2**: The range we considered for Pong was chosen as the baseline performed much worse at other learning rate ranges. We therefore searched in only this range to optimize our own method. We use 6 learning rates in a larger range for grid world in Figure 5c. To **R3**: We think our approach can be useful for generalization, for example by learning in a state and directly generalizing to its transformed versions. To **R2**: The action transformation is a group representation, it therefore must have invertibility.

We thank all reviewers for their time and efforts. We will incorporate the experiments and discussions, as well as typos, references and minor issues in the paper, and release all code.