To area chairs,

Thank you for your time. Reviewer 3 has given us quite valuable comments and we have accordingly improved our paper. But we still have a few concerns on his comments.

We believe reviewer 3 is an expert in related fields, but it seems that he has relative less knowledge on the reinforcement learning. Because he has raised a main issue that why our method predicts both state value and policy, when we already state in the main text that our method is an advantage actor-critic method, similar as A3C [35], and Eq. 6 is clearly using advantage to update policy. Although we have included more derivations of the reinforcement learning part in our response and revised paper, we are still worrying that this aspect may affect a fair comment on our work.

Again, many thanks for your time.