We thank the reviewers for their helpful and constructive comments and we are happy they found the paper interesting.

Reviewer 1 Reviewer 1 suggested we investigate the influence of top-down vs. bottom up attention — we agree this is one of the more interesting aspects of the model. In Section 4.4 of the paper we provide some quantitative studies of the advantage of the top-down attention over the bottom up, but we treated it more as an ablation study to demonstrate that the top-down queries were playing an important role . This also influenced the style of bottom-up attention we chose: the one in the paper keeps most of the model the same and isolates actual top-down influence. We agree that a full analysis of the strategies that a bottom-up agent learns, including a detailed comparison with a top-down agent, would be very interesting, but length constraints prevented further analysis here.

Reviewer 2 Reviewer 2 was concerned about novelty of the model. We agree that similar models have been proposed in the past and we note this in the paper. We feel that the novel contribution of the paper is the application of an attention model to RL and the interpretability of the rich behaviors and strategies that emerge from this.

Another concern was about what happens when the agent doesn't learn well (such as in chopper command and battle zone) and why uniform attention doesn't help in that case. Uniform attention, like having an average pooling layer over the entire conv-net output, causes the information to be blurred across the whole image due to the spatial reduce sum (Equation 5. in the paper) — this will prevent the agent from extracting local information from images and often fail at the task. We also think that some long-term credit assignment issues may be exacerbated by using this attention, since the agent has to be already be attending correctly to even assign the reward to the right place. This would be an interesting avenue for further research.

Reviewer 3 Like Reviewer 1, Reviewer 3 was also interested in the effects of top-down influence and raises an interesting separation of the two salient features of the top-down mechanism - one is that it has to do with the macro/task level features, and the other is the time dependency. We argue that the it is indeed the macro level influence (as R3 suggests) and that the experiments in Section 4.4 show that. The reason time dependence is not the key factor here is that both non top-down baselines can be time-dependent - the ConvLSTM, which has a state, can carry information forward in time. What it can not do is to carry that information globally, but just locally through convolutions. So whether it's creating key/values for the fixed query as in the first baseline, or explicit attention maps, both **can** be time dependent, but can't be macro/task dependent. While this is different than the test that R3 suggests, we believe one can draw similar conclusions from it. The ConvLSTM could learn to sequentially promote different features to the learned, fixed query set which would have a similar effect to learning sequential masks.

We apologize that the videos were not available, we're not sure why (R1 was able to view them), but we suggest trying again now.

We will clarify the text to make the last point R3 raises clearer.