Response to reviewers

We would like to thank the reviewers for their positive feedback and suggestions.

In the final version of the paper we plan to provide additional explanation for the figures, as reviewer 1 and 2 have suggested.

As our primary focus is on biological plausibility, we will expand our discussion on links to existing hippocampal literature, as well as earlier biological reinforcement learning models that considered partial observability in much more restrictive settings (with discrete variables, exact inference).

In particular, distributional SRs provide a link between various intriguing and seemingly disparate experimental observations. The link between hippocampal physiology and (non-distributional) SRs has been explored previously (e.g., Stachenfeld 2017) providing an interpretation for phenomena such as "splitter" cells, which show spatial tuning that depends on the whole trajectory (i.e. policy) traversed by the animal not just on its current position (Grieves et al., 2016). However, as discussed in the paper, relevant states for a given reinforcement learning problem (in this case states over which the SR should be learned) can not be assumed to be directly available to the agent but must be inferred from observations. The hypothesis that hippocampal place cell activity encodes *inferred* location, with its concomitant uncertainty, has also been linked to experimental data (Madl et al., 2014). Thus, our approach connects these two separate threads in the literature and thereby encompasses both groups of experimental results.

Furthermore, the framework helps to link simulation of an internal model to learning. Acquisition of the inference model in our framework requires simulating experience (sleep samples) from the agent's current model of the environment which are then used to update the recognition model. The sleep samples reflect the agent's knowledge of the environmental dynamics but they don't necessarily correspond to a previously experienced trajectory exactly. This is reminiscent of hippocampal replays which do not just recapitulate previous experience, but often to correspond to novel trajectories not previously experienced by the animal (Gupta et al., 2010; Olafsdottir et al., 2015; Stella, 2019). Relatedly, Liu et al. (2019) recently observed that replay events in humans reflect abstract structural knowledge of a learned task. Our model suggests a novel functional interpretation of these replayed trajectories, namely, that they may play an important role in *learning to infer* relevant latent states from observations. This accords with the observation that experimental interference with replay events impedes learning in contexts where optimal actions depend on history-based inference (Jadhav et al., 2012).