We thank the reviewers and AC for their thoughtful comments and thorough review. For this response, we have identified a few common themes that have been raised by several reviewers, and address them in turn. We begin by discussing a few major issues we believe are key for the reviewers' evaluation of the main contribution of the paper:

1. Reviewers #1, #3, and #4 request a more thorough evaluation against standard RL and reward-seeking agents. A comparison against a reward-seeking, exploitation-only agent is provided in Fig. 3b, showing that it performs similarly to the full Active Inference (AIF) agent (but with less effective exploration, as expected). We also tested agents from OpenAI's baselines repository on the Animal-AI environment, and found that (given the same number of training episodes) our agent performs considerably better than DQN and A2C, and comparable to PPO (all baselines with default settings). We will include detailed comparisons in the camera-ready version of the paper.
2. Reviewer #1 urges us to describe our calculation of Eqs. 8b and 8c. For 8b, $H(s_\tau \mid \pi)$ is estimated sampling from the transition network, and $H(s_\tau \mid o_\tau, \pi)$ from the encoder network (both parameterised with Gaussians, so entropies can be calculated from log-variances). For the first term in 8c we sample several θ from the MC-dropouts and several s_τ from the transition network; then average the entropies $H(o_\tau \mid s_\tau, \theta, \pi)$ (which are closed-form since o_τ is Bernoulli-distributed) over the (θ, s_τ) samples. For the second term, we fix the θ and sample multiple s_τ (so that, effectively, $p(o|s) = \sum_\theta p(o|s,\theta)p(\theta)$ is approximated with a single MC sample) and repeat the procedure. (We also tried sampling several θ and averaging the distributions over o_τ, which is possible because o_τ is Bernoulli-distributed. Although noisier, the estimator described in the paper was faster and more suitable for training.) We agree with the reviewer's statement that the entropy of the average is not the same as the average of the entropies – and the difference between the two is the mutual information, which is known to be part of the EFE. We will describe this calculation in detail in the appendix.
3. Reviewers #2, #3 and #4, raise concerns about the clarity of our exposition, which we group in 3 items:
   - Regarding $P(o_\tau)$: At all times it should be conditioned on π, i.e. $P(o_\tau|\pi)$. This should not have appeared in lines 97 and 113 and we will update accordingly.
   - Regarding $\log P(o_\tau)$ as reward: We would like to clarify that in AIF the reward is not differentiated from other types of observations. Certain (future) observations (e.g. green color in Animal-AI) are considered more desirable given a task, so in practice rewards can be encoded as observations with higher prior probability using $\log P(o_\tau)$. We will make this conceptual point explicit in the camera-ready version.
   - Regarding $\tilde{Q} = Q(o_\tau, s_\tau, \theta \mid \pi)$: The expansion is shown in Eq. 8a under the expectation, although we agree it could benefit from being presented separately. We will mention it explicitly in the camera-ready version.
4. Reviewer #1 claims "scaling active inference has been done before." We agree with the reviewer that prior work has been done on this topic, but our contribution represents a technical (and qualitative) improvement over previous approaches. This is achieved by: 1) estimating all summands of EFE (line 41) and, 2) for the first time successfully training AIF agents on full-fledged, complex environments with visual input, multiple actions, and sparse rewards. We believe this constitutes a substantial improvement over the state of the art in AIF applications.
5. Reviewers #1 and #3 claim we do not provide enough details to reproduce our results. We would like to remind the reviewers we have uploaded our code to a public repository, which will be linked in the camera-ready version of this paper (line 198). Additionally, for clarity we will include pseudo-code for the algorithm in the appendix.

In addition, we would like to address a few minor comments:

6. Reviewers #1, #2 and #3 have suggested additional references for amortised action with planning, disentanglement and model-based RL. We will add these to the discussion in the camera-ready version.
7. Reviewer #2 suggests we could change the notation to be more in line with the variational inference literature. Although we agree with the reviewer's aims, given the space constraints and how much we rely on the AIF literature, we believe it would make the exposition denser and the links with prior AIF literature harder to track. Nonetheless, to make the paper more accessible to non-neuroscientists, in the camera-ready version we will add glossary to the appendix describing in detail what each symbol and probability distribution represent.
8. Reviewer #4 argues 700 trials "seems high," given that "one of the hallmarks of biological intelligence is few-shot learning." We agree with the reviewer, but emphasise that our agent starts 'from scratch' (i.e. with randomly initialised networks) each run, while biological organisms are fantastically able to form good priors that generalise and transfer between tasks. The extension of AIF to transfer learning remains an exciting avenue for future work.
9. The reviewers have identified a few grammatical errors, like occasional missing articles, misplaced sentences, or acronyms (like 'MC') that should be defined more explicitly. Additionally, reviewer #2 has identified redundant hyper-parameterisation (i.e. γ in Eq. 3). We will address all of these in the camera-ready version of the paper.

We thank again the reviewers and AC for their work, which we are sure will improve the next version of this paper. We hope this response addresses the core issues raised during the review process.