We would like to thank all the reviewers for their time and helpful feedback.

All reviewers mention that the current version of the paper is a little dense. We will work on improving clarity and accessibility and try to convey more intuition in the main text. We would like to thank reviewers 1 and 2 for their helpful suggestions on how to make the paper more accessible.

We also plan to add more simulation results in the final version. To reviewer 1, we will add simulation results for the conjecture. To reviewer 2, we completely agree that it would be very interesting to see simulated examples of information gain and regret bounds. For the linear-Gaussian bandit, the regret bound derived in the paper is comparable to the best bounds in the literature up to logarithmic factors. It will be interesting to see the simulated total information gain and plot the corresponding regret bounds. For MDPs, we suspect that the upper bounds on total information gain are loose and the actual information gain might be much smaller.

Several reviewers ask about how the paper might lead to new algorithms. This is definitely an important direction for future research. As a starting point, it may be interesting to experiment with UCB algorithms that use information gain in their upper confidence functions. They come with regret guarantees for problems discussed in the paper, and it would be interesting to see how they perform in practice. UCB algorithms that are studied previously usually construct confidence sets around the model parameter based on past observations, and they do not consider how much information would be revealed about the model when we see a new observation. By explicitly taking into account information gain, these new UCB algorithms tend to give a higher bonus to actions that reveal more information about the system compared to existing UCB algorithms. It would be interesting to see whether these new algorithms would work better for some classes of problems.

Finally, we will work on clarifying the notations and proofs.

Below are additional responses to individual reviewers.

Reviewer 1

– *Line 26: please add references for "Most analyses.."*
Thanks for pointing it out. We will add the references.

– *Line 69: should a be A_ℓ here?*
We use A_ℓ to denote a possibly random action selected by the algorithm, while a is used to denote some fixed action. We will clarify this in the final version.

– *It seems the θ variable is overloaded at different places. For example, Line 111 θ refers to the true model but elsewhere it refers to the random variable.*
Under our Bayesian framework, the true model is a random variable. We will emphasize this in the final version.

Reviewer 2

– *The paper offers novel regret bounds for Thompson sampling and UCB algorithms but the key idea is mostly a straightforward extension of Russo and Van Roy (JMLR'16)...*
It is true that our main idea bears similarity to Russo and Van Roy (JMLR'16), but a major difficulty of applying their approach to complex environments is that it is unclear how to analyze the mutual information between optimal actions and observations, as optimal actions themselves are fairly complicated objects in complex environments. Our approach considers a more natural information gain between the model and observations, which sidesteps their issue and allows us to obtain information-theoretic bounds for more complex problems like MDPs and factored MDPs that are not achieved in their paper.

– *There are some problems with the notation in the paper...*
We will clarify our notation on probability measures. We use T to denote the total number of time periods for the linear bandit problem as it is more consistent with other bandit literatures, but I see how it can cause confusions and we will make it clear in the final version.

– *...presenting less material (e.g., leave the factored MDPs for the appendix)...*
We include factored MDPs in the main text as an example to demonstrate that our approach is able to handle complex structured environments that could not be handled using tools from previous work such as Russo and Van Roy (JMLR'16).

Reviewer 3

– *...more exhaustive descriptions of the proof will help the readers and the reviewer to understand the paper...*
We will polish the proofs for the final version.