Paper ID 3012: Inverse Rational Control with Partially Observable Continuous Nonlinear Dynamics We would like to thank the reviewers for their valuable and important comments. We here submit our responses. All the issues addressed in this document will be included in the camera-ready version.

1. Sample efficiency [All reviewers] We agree it is important to investigate the relationship between the number of data points and the accuracy of the parameter recovery to guide the experimentalists about how much data they need to collect for recovering a subject's internal model. The results presented in the paper were from 500 state-action trajectories (500 fireflies), each with about 5–15 state-action time points. The amount of data is reasonable since the subjects repeat the task hundreds of times.

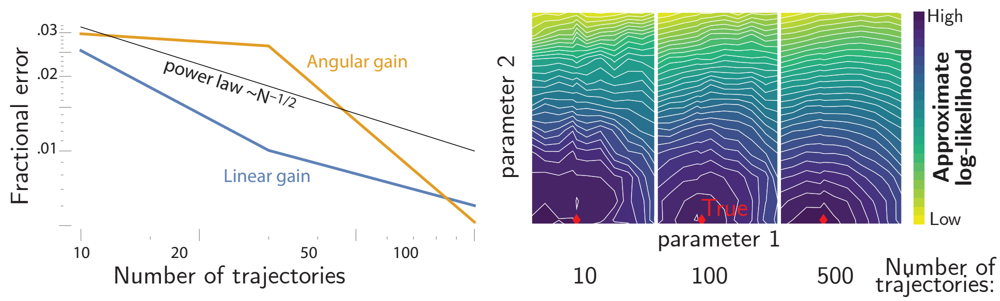

Figure 1: (Left) Recovery error vs the number of data points. (Right) Log-likelihood surface with different numbers of data points. Red diamonds indicate true parameter values.

We ran additional experiments to show the relationship between the number of state-action trajectories and the fractional error (absolute error divided by true parameter). We ran inverse rational control (IRC) for 20 agents with different model parameters, using 10 samples (L=10 in Algorithm 2) and 400 gradient ascent steps. Figure 1 (left) shows the performance given 10, 30 and 100 state-action trajectories. The error falls off with the square root of the number of trials as expected. These parameters can be readily identified with relatively few trials; for other tasks we expect the same scaling with numbers of trajectories, but with different scaling factors depending on the how much the actions vary with the task.

Figure 1 (right) explains why recovery accuracy grows with data volume: the surface of log-likelihood becomes smoother and the peak moves closer to the agent's true parameters.

2. References in psychology [R2] We thank the reviewer for suggesting related works in psychology. We have thoroughly reviewed the related works in psychology and added the following references with discussion in Section 2. Related works of the camera-ready version: Lieder et al., *Behavioral and Brain Sciences* (in press), Bourgin et al., *ICML* (2019), Krueger et al., *CogSci* (2018), Baker et al., *Nature Human Behavior* (2017) Rafferty et al., *Cognitive Science* (2015), Lewis et al., *Topics in cognitive science* (2014), Walsh et al., *Psychological Bulletin* (2014), Howes et al., *Psychological review* (2009).

3. Hyperparameters for reproducibility [R2] To increase reproducibility, we here provide details of hyperparameters for Algorithm 1 and Algorithm 2.

4. Standard benchmarks [R1,3] Now that we have a workable framework for IRC we do plan to apply it to new neuroscience and ML tasks (especially partially observed versions of standard continuous control tasks), since standard benchmarks for continuous POMDPs do not yet exist.

5. Comparison to previous models [R2] We do hope to access the behavioral data in [38,39] to directly compare our model to past work. However, our model essentially *contains* those models, which did not address control at all. So IRC will give identical findings when restricted to the older models but will provide fundamentally new explanations when including control.

Algorithm 1			
Batch size	64	Discount factor	0.99
Replay memory size	10^6	Actor learning rate	10^{-4}
Critic learning rate	10^{-3}	Optimizer	Adam
Number of units per hidden layer	128	Activation function of hidden layer	ReLU
Activation function of Actor output layer	Tanh	Activation function of Critic output layer	Linear
Algorithm 2			
Length of trajectory (T)	500	Number of samples (L)	50
Optimizer	Adam	Learning rate	10^{-3}