	Humans	**RL Agents**	**Persistence** (r)			**Persistence** (q)			**Search Temp.** (T)			**Heuristic** (h)			
Domain	n=5	α=50	1	2*	4	0.8	0.9	0.95*	2	10*	50	Manhat.*	Maze.Dist.	Goal.Count.	h_{add}*
Doors, Keys, Gems	0.79	0.58	0.60	0.73	0.73	0.53	0.60	0.73	0.67	0.73	0.77	0.73	0.90	–	–
Block Words	0.73	0.82	0.90	0.87	0.90	0.70	0.83	0.87	0.83	0.87	0.87	–	–	0.43	0.87

Table 1: Robustness to model mismatch. Top-1 accuracy of SIPS at the third time quartile (Q3), evaluated on data generated by humans, RL agents, and mismatched models. We ran SIPS assuming r=2, q=0.95, T=10, and a Manhattan (h_{add}) heuristic for Doors, Keys, Gems (Block Words). Matched parameters are starred (*).

We thank the reviewers for engaging carefully with our paper, and for providing helpful and constructive feedback. We commit to addressing the minor issues raised and adding the suggested references.

R1, R2, R3, R4 raised concerns about whether our model of boundedly-rational planning is robust to plans generated by humans and other models. We appreciate these concerns, and agree that it best to perform a user study as R2 suggests, and to avoid the circular evaluation pointed out by R3. In response, we have performed a series of robustness experiments (**Table 1**), showing that SIPS is robust to data from 5 pilot human subjects (30 trajectories per subject), a Boltzmann RL agent, mismatched parameters r, q, T, and h. While performance can degrade with mismatch, this is partly due to the difficulty of inference from highly random behavior (e.g. q=0.8, h=Goal Count). In fact, when mismatched parameters are *more* optimal, performance can *improve* (e.g.h=Maze Dist.). Importantly, SIPS does well on human data (Top-1=0.78 / 0.73), showing robustness even when the planner is unknown. **We have also conducted preliminary experiments showing that human goal inferences correlate highly with SIPS** (Pearson's $r = 0.85$). We will expand on these experiments in the final paper with more domains and cross-method comparisons.

R1 asked about the motivation for using a negative binomial for search budgets $\eta \sim \text{NB}(r, q)$. To explain, r and q model the persistence of a planner who may give up after expanding each node. 1-q is the probability that the planner considers giving up, while r is the number of times the planner has to consider giving up before actually giving up. This captures the intuition that people are more likely to give up the longer we plan, while still exhibiting some persistence. **R1 also expressed concerns about our relaxation of A* search**, where a node s is sampled from the open list for expansion according to $P_{\text{expand}}(s) \propto \exp(-f(s, g)/T)$. To explain, this captures how agents may fail to choose the best cognitive action (i.e. node expansion) during search, following a Boltzmann distribution. T controls how informed the search is, with T=0 (least informed) equivalent to breadth-first search, T=∞ (most informed) equivalent to standard A*. Search is sound, because all nodes are eventually expanded, and path costs updated. Similar modifications are made in both [1] and [2]. We will add these references in the final paper.

R2 and R3 were also concerned whether our modeling assumptions are overly specific *(R2. Q1, R3. Weaknesses)*. We agree that there are many types of planning that go beyond the scope of this paper. As we note in Future Work, this motivates extensions to domains like task and motion planning. To R3's point on optimal RL actions without explicit planning, we agree this is a good model when humans are well-practiced. Still, when inferring plans for novel tasks with sparse rewards, we believe explicit planning is a better model, and will gladly add discussion on this point. **Planners can also exhibit many other bounds, including limited memory (R2) or inference capacity (R3).** We see modeling such bounds as important future work. While our current model no doubt approximates human planning by leaving out these bounds, humans are likely to use similarly approximate models when inferring the goals of others. As **Table 1** shows, this approximation is enough to ensure reasonable robustness in the domains we consider.

R1 asked about the importance of probabilistic programming for our research. While R1 is right that the basic particle filtering approach could be implemented manually with some additional work, our use of custom proposal programs [3] and involutive MCMC [4] in the SIPS rejuvenation moves requires computing importance weight ratios for all random choices in both the proposals and the model. Without Gen's automated weight computation, implementing this would be highly tedious and error prone, akin to implementing back-propagation for a VAE without TensorFlow or PyTorch. A manual implementation would also be much more difficult to extend to complex goal priors and planners. We are happy to emphasize this in the final paper.

R2 asked for clarification about whether our approach is tied to deterministic settings. While all the evaluated domains were deterministic, our framework also supports Probabilistic PDDL [5], and can be readily extended to stochastic domains, e.g. by determinization as R2 suggests. Our use of a replanning model also means that it is not as strongly tied to deterministic environments as R3 contends: if the agent encounters a state that it did not plan for (e.g. randomly failing to pick up a block), it simply replans from that new state, which we believe to be a reasonable approximation of human behavior in mostly deterministic domains. We will clarify these points in the final paper.

R4 asked why unbiased BIRL performs better in the Taxi environment, and why oracle BIRL sometimes performs better for top-1 accuracy. This is because SIPS can suffer from particle collapse, whereas BIRL can perform exact inference if it has good Q-value estimates (true for both oracle BIRL and the Taxi domain's small state space). SIPS can be improved by increasing the particle count, using rejuvenation, or variance control techniques. We have since implemented some of these techniques (e.g. residual resampling), and will update our results accordingly. **R4 also asked how well SIPS performs when the agent does not replan.** This is addressed in the supplement, where we show that SIPS performs reasonably well on trajectories from an optimal full-horizon planner.