Modeling the Mistakes of Boundedly Rational Agents Within a Bayesian Theory of Mind

Arwa Alanqary*, Gloria Z. Lin*, Joie Le*, Tan Zhi-Xuan*†
Vikash K. Mansinghka, Joshua B. Tenenbaum
Department of Brain and Cognitive Sciences, MIT
Cambridge, MA 02139
{alanqary,gzlin,joiele,xuan,vkm,jbt}@mit.edu
*Equal Contribution †Corresponding Author

***Abstract*— When inferring the goals that others are trying to achieve, people intuitively understand that others might make mistakes along the way. This is crucial for activities such as teaching, offering assistance, and deciding between blame or forgiveness. However, Bayesian models of theory of mind have generally not accounted for these mistakes, instead modeling agents as mostly optimal in achieving their goals. As a result, they are unable to explain phenomena like locking oneself out of one's house, or losing a game of chess. Here, we extend the Bayesian Theory of Mind framework to model boundedly rational agents who may have mistaken goals, plans, and actions. We formalize this by modeling agents as probabilistic programs, where goals may be confused with semantically similar states, plans may be misguided due to resource-bounded planning, and actions may be unintended due to execution errors. We present experiments eliciting human goal inferences in two domains: (i) a gridworld puzzle with gems locked behind doors, and (ii) a block-stacking domain. Our model better explains human inferences than alternatives, while generalizing across domains. These findings indicate the importance of modeling others as bounded agents, in order to account for the full richness of human intuitive psychology.**

I. INTRODUCTION

A key aspect of human intuitive psychology is our understanding that other agents are fallible: they may possess false beliefs [1], lack knowledge [2], fail to plan ahead [3], or act unintentionally [4]. This capacity is crucial to social life, allowing us to teach others [5], or forgive harms that we take as unintended [6], [7]. Remarkably, even 18-month old infants seem to account for such errors when inferring the goals of others, enabling them to offer assistance [8].

What are these errors, and how do we understand them in a way that allows us to infer the goals of others? In the Bayesian Theory of Mind (BToM) framework, goal inference is explained as *inverse planning*, where observers infer a posterior distribution over goals by modeling agents as rational planners [9]. However, while prior BToM models explain how we can infer others' goals, desires, and intentions [10], [11], as well as how we might infer mistaken beliefs [12], [13], little attention has been paid to mistaken goals, plans, or actions. With some exceptions [14], [15], most BToM models only account for low-level action mistakes [16]. This fails to capture higher-level mistakes, and has been challenged as a model of sequential decision making [17], [18].

In this paper, we build upon a recently proposed model of agents as boundedly rational planners [19]. Unlike earlier BToM agents which plan via exhaustive computation of expected value over the entire state space [20], [9], these agents do not always plan optimally, but, like ourselves, only plan several steps ahead before executing that partial plan and replanning. This is resource rational in many cases [3], [21], but can also lead to failure: you might lock yourself out of the house, because you neglect to bring your keys. We extend this model with goal mistakes, due to confusion of goals with semantically similar specifications, and action mistakes, due to occasional execution of unplanned actions. Our model thus accounts for sub-optimality at three distinct levels of human decision-making. We hypothesize that human goal inferences given sub-optimal action sequences are better explained by Bayesian inference in this model than previously introduced BToM approaches. We test this hypothesis by eliciting human goal inferences in two experiments. By comparing human judgements against predictions from each computational model, we evaluate the fidelity of these models to our intuitive theory of mind.

II. COMPUTATIONAL MODEL

To account for mistakes at multiple levels of decision making, we model agents and their environments as generative processes of the following form:

$$\textit{Goal prior:} \quad g_0 \sim P(g_0) \tag{1}$$
$$\textit{Goal transition:} \quad g_t \sim P(g_t | g_{t-1}, g_0) \tag{2}$$
$$\textit{Plan update:} \quad p_t \sim P(p_t | s_{t-1}, p_{t-1}, g_{t-1}) \tag{3}$$
$$\textit{Action selection:} \quad a_t \sim P(a_t | s_t, p_t) \tag{4}$$
$$\textit{State transition:} \quad s_t \sim P(s_t | s_{t-1}, a_t) \tag{5}$$
$$\textit{Observation noise:} \quad o_t \sim P(o_t | s_t) \tag{6}$$

where g_0 is the agent's original intended goal, and g_t, p_t, a_t, s_t are the agent's current (potentially corrupted) goal, the internal state of the agent's plan, the agent's action, and the environment's state at time t respectively. These generative processes are specified as probabilistic programs (Figure 2), and their corresponding mistakes are described below.

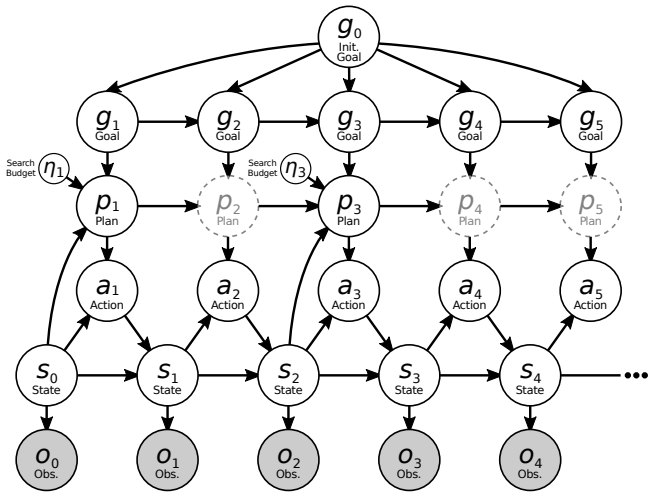


Fig. 1: A single realization of our boundedly rational agent model. At t=1, the agent samples a search budget η_1 and searches for a plan p_1 that is two actions long. At t=2, no additional planning needs to be done, so p_2 is copied from p_1, as denoted by the dashed lines. The agent then replans at t=3, sampling a new search budget η_3 and an extended plan p_3 with three more actions.

```
1: model GOAL-TRANSITION(t, g_{t-1}, g_0)
2:     parameters: ε_g (goal noise)
3:     if BERNOULLI(ε_g) = TRUE then
4:         return if g_0 = g_{t-1} then CORRUPT(g_0) else g_0 end if
5:     else
6:         return g_{t-1}
7:     end if
8: end model
```

(i) Samples goals from $P(g_t|g_{t-1}, g_0)$

```
1: model PLAN-UPDATE(t, s_t, p_{t-1}, g)
2:     parameters: PLANNER, r, q, γ, h
3:     if t > LENGTH(p_{t-1}) or s_t ∉ p_{t-1}[t] then
4:         η ~ NEGATIVE-BINOMIAL(r, q)
5:         p̃_t ~ PLANNER(s_t, g, h, γ, η)
6:         p_t ← APPEND(p_{t-1}, p̃_t)
7:     else
8:         p_t ← p_{t-1}
9:     end if
10:    return p_t
11: end model
```

(ii) Samples plans from $P(p_t|s_t, p_{t-1}, g)$

```
1: model ACTION-SELECTION(t, s_t, p_t)
2:     parameters: ε_a (action noise)
3:     if BERNOULLI(ε_a) = TRUE then
4:         return UNIFORM(ACTIONS(s_t) \ p_t[t][s_t])
5:     else
6:         return p_t[t][s_t]
7:     end if
8: end model
```

(iii) Samples actions from $P(a_t|s_t, p_t)$

Fig. 2: Generative subroutines specifying (i) goal transition noise, (ii) plan updates, and (iii) action selection.

A. Mistaken goals via temporary goal confusion

Consistent with previous BToM literature [9], we restrict our scope to inference over a fixed set of context-relevant goals with associated prior probabilities (Eq. 1), leaving aside how people come up with these contextual hypotheses. However, within a fixed set of goals, complex goals can still get confused. For example, when stacking blocks to spell a word, one might have in mind a misspelling: Is it "firey" or "fiery"? To account for these mistakes, we introduce goal transition noise (Eq. 2), specified by the procedure in Figure 2(i). At each step t with probability ϵ_g, the original goal g_0 may be corrupted to produce a similar *temporary* goal g_t, or the corrupted goal corrected to the original one (Line 4). Goal noise can be domain-specific, e.g. in Blocks World, CORRUPT might correspond to a random permutation.

B. Mistaken plans via resource-bounded planning

We model agents that interleave resource-bounded planning with plan execution. This can cause mistakes due to failure to plan ahead. At each step t, agents already have a previous partial plan p_{t-1}, and construct an updated plan p_t (Eq. 3).

This procedure is shown in Figure 2(ii), and it captures boundedly rational planning in the following ways: First, the agent does not update it previous plan (i.e., a sequence of intended actions) if it already extends to the current step t (Line 8). Second, if the plan *does* need to be extended, the agent only spends a limited budget η to construct a new partial plan $\tilde{p}_t$. For example, if the agent plans via forward search, η is the number of steps the agent thinks ahead. We sample η from a negative binomial distribution (Line 4), encoding the assumption that neither very short nor very long plans are likely. Third, we assume that the planning procedure itself, PLANNER, is noisy (Line 5). While in principle, any planning algorithm could be used, we adopt a probabilistic version of A* search, capturing the intuition that humans often plan by thinking a few steps ahead, guided by a heuristic $h(s, g)$ that evaluates how promising a state s is relative to the goal g. In regular A* search, only the most promising state s is expanded at each iteration. However, since humans may not rank states perfectly, we instead sample s from the Boltzmann distribution:

$$P_{\text{expand}}(s) \propto \exp(-f(s, g)/\lambda) \tag{7}$$

where higher λ increases the randomness of search, $c(s)$ is the cost of reaching s from the initial state, and $f(s, g) = c(s)+h(s, g)$ is the estimated total cost of reaching the goal g by passing through s. The search algorithm terminates when either the goal state g is reached or the ηth state is expanded (i.e. the plan budget is exhausted), at which point a partial plan $\tilde{p}_t$ to last-expanded state s is returned.

C. Mistaken actions via execution errors

Humans may not always execute plans as intended. Instead, due to carelessness or lack of motor control, we may sometimes commit execution errors — for example, dropping a block by accident, or walking a step more than intended [22]. As such, we model action selection (Eq. 4, Figure 2(iii)) as a process where the agent usually executes their intended action given the current plan p_t and state s_t (Line 6), but with

probability ϵ_a executes one of the other possible actions in state s_t at random (Line 4).

D. Bayesian goal inference

We model observers as performing Bayesian inference over an agent's intended goal g_0 given observed states $o_{1:t}$, which are potentially noisy observations of the actual states $s_{1:t}$. We assume that the agent and observer have a shared symbolic understanding of the environment, specified with predicates in the Planning Domain Description Language [23]. To model observation noise, Boolean predicates are corrupted with probability ϵ_s, while numeric predicates have Gaussian noise added with variance σ_s^2. Given the complexity of our agent model, exactly computing the goal posterior $P(g_0|o_{1:t})$ is intractable. Thus, we use Sequential Inverse Plan Search, the sequential Monte Carlo algorithm developed by [19], to approximate $P(g|o_{1:t})$. We refer readers to that work for technical details.

As baselines for comparison, we compute goal inferences using lesioned agent models: **G-Lesioned**, where goal mistakes are absent, **P-Lesioned**, where resource bounded planning is absent, or **A-Lesioned**, where action mistakes are absent. We also compare with the **Boltzmann agent model** used in earlier BToM approaches [9] and Bayesian Inverse Reinforcement Learning [20]. In this model, agents precompute the expected future reward $V(s)$ of every state s, and follow a Boltzmann policy, noisily selecting actions that tend to maximize reward:

$$\pi(a|s) \propto \exp(\alpha[R(s,a,s') + V(s')]) \tag{8}$$

Here, s' is the successor state, $R(s,a,s')$ is the reward from taking action a from s to s', and higher α leads to lower noise. We set $R(s,a,s') = -1$ for all actions and treat goal states as terminal, leading agents to prefer shorter routes to goals. Because this model exhaustively plans (i.e. computes the expected reward $V(s)$) over the entire state space, it only accounts for low-level action mistakes. As such, we hypothesize that it will not explain mistaken goals or plans as well as our boundedly-rational model.

We also note that the Boltzmann-rational model is algorithmically implausible for humans in the compositional domains we consider [19]. This is due to the sheer number of states s over which expected rewards $V(s)$ have to be computed (e.g. 400,000 feasible arrangements of 8 blocks). In practice, we circumvent this by computing $V(s)$ only for states within the observed trajectories.

III. EXPERIMENTS

To demonstrate the generality of our model, we conducted experiments in two domains: (i) a gridworld puzzle called Doors, Keys & Gems, and (ii) a Blocks World variant called Block Words, where an agent spells words out of lettered blocks. These domains exhibit the compositional structure that humans encounter in daily life, making them tractable to plan in, but also complex enough for mistakes to arise.

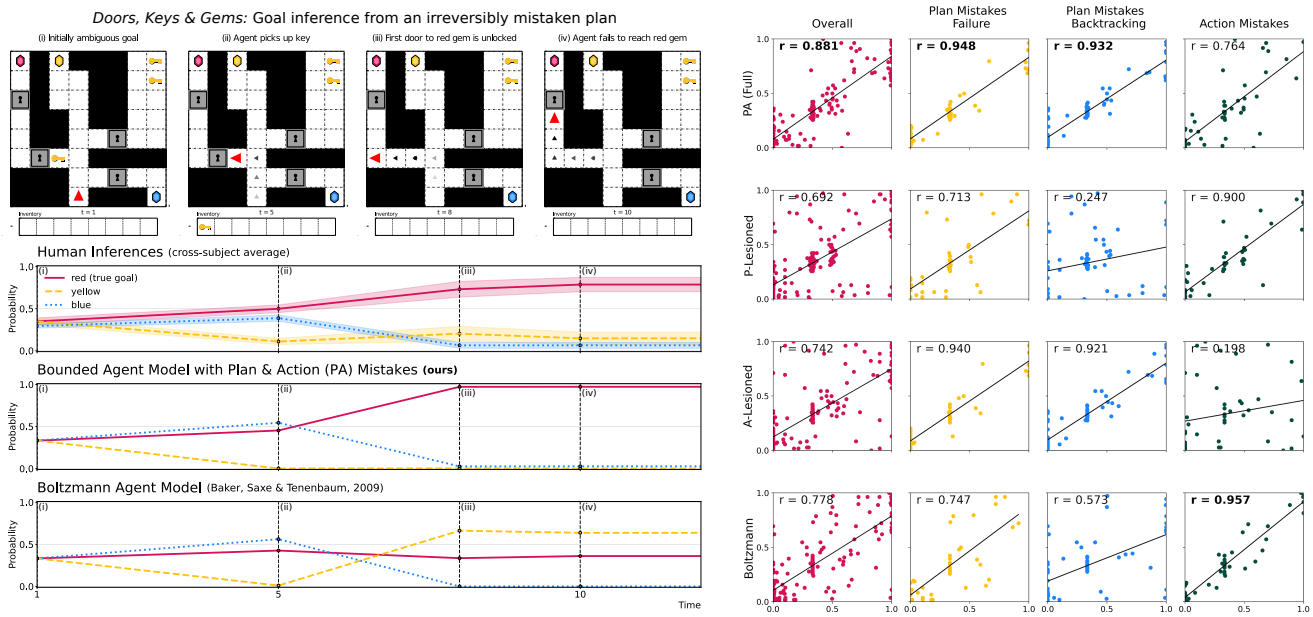
In each domain, we elicited goal inferences from human observers as they watched a variety of optimal and sub-optimal agent trajectories unfold. Given the complexity of our model, sub-optimal trajectories might admit multiple interpretations: "Was that a mistaken action, or a mistaken plan?" As such, we designed trajectories to make some mistakes more likely than others. For example, if someone walked a step out of the house before turning to get their keys, an observer might take that as a mistaken action, but if they walked all the way to the bus stop before turning around, a mistaken plan would seem more likely. In the Block Words domain specifically — which has a richer goal space — we also designed trajectories with mistaken goals: agents would sometimes stack a misspelled word, before correcting themselves.

A. Experiment 1: Doors, Keys, & Gems

In this domain, an agent must navigate a maze in order to collect one of three colored gems, which may be locked behind doors (Figure 3(a)). Keys are required to unlock doors, and can only be used once, leading to the possibility of irreversible failure if the agent does not plan ahead. We assume that planning is guided by a maze distance heuristic that ignores the presence of doors, leading occasionally to myopic plans which neglect the need for keys. Because the goals in this domain are simple, we do not model mistaken goals.

1) Stimuli: We designed 16 agent trajectories as stimuli for human participants, organized into four subsets: (1) A control set of optimal trajectories. (2) Trajectories with irreversibly mistaken plans, where the agent myopically uses up all obtainable keys, such that the goal gem is locked out of reach (Figure 3(a)). (3) Trajectories with mistaken actions, where the agent takes a few false steps, then corrects their behavior. (4) Trajectories with short-sighted plans, leading the agent to backtrack to obtain keys. Participants accessed a web interface which presented all 16 stimuli in random order. Stimuli were presented as animated videos which paused at selected judgement points. At each point, participants provided goal inferences by selecting the gem(s) they believed to be the agent's most likely goal. Participants could select multiple gems if more than one seemed equally likely, and these responses were converted to probability distributions.

2) Participants: We recruited 20 US participants (mean age 36.4, SD 10.4; 8 women, 12 men, 0 non-binary/other) via Amazon Mechanical Turk (AMT), restricting to those with a HIT approval rate of 99% and above. Participants went through a tutorial and answered four comprehension questions before viewing the stimuli. Participants also earned points proportional to the probability they gave the true goal (mean score 28, SD 10), and were paid \$1 per 10 points, incentivizing accurate guesses. Two participants were excluded from our analysis, either for failing two or more comprehension checks, or for guessing indiscriminately and failing to reach a threshold of 10 points.



(a) Goal inferences over time for a trajectory with an irreversible failure.

(b) Human (y-axis) vs. model (x-axis) inferences.

Fig. 3: Results for the Doors, Keys & Gems domain. In **(a)**, we show goal inferences of humans (avg., n=18, w. standard error), our model (PA) and the Boltzmann agent model for an illustrative trajectory. In **(b)**, we compare human vs. model inferences across models (full model, lesioned models for plan/action mistakes, Boltzmann model) and stimulus types.

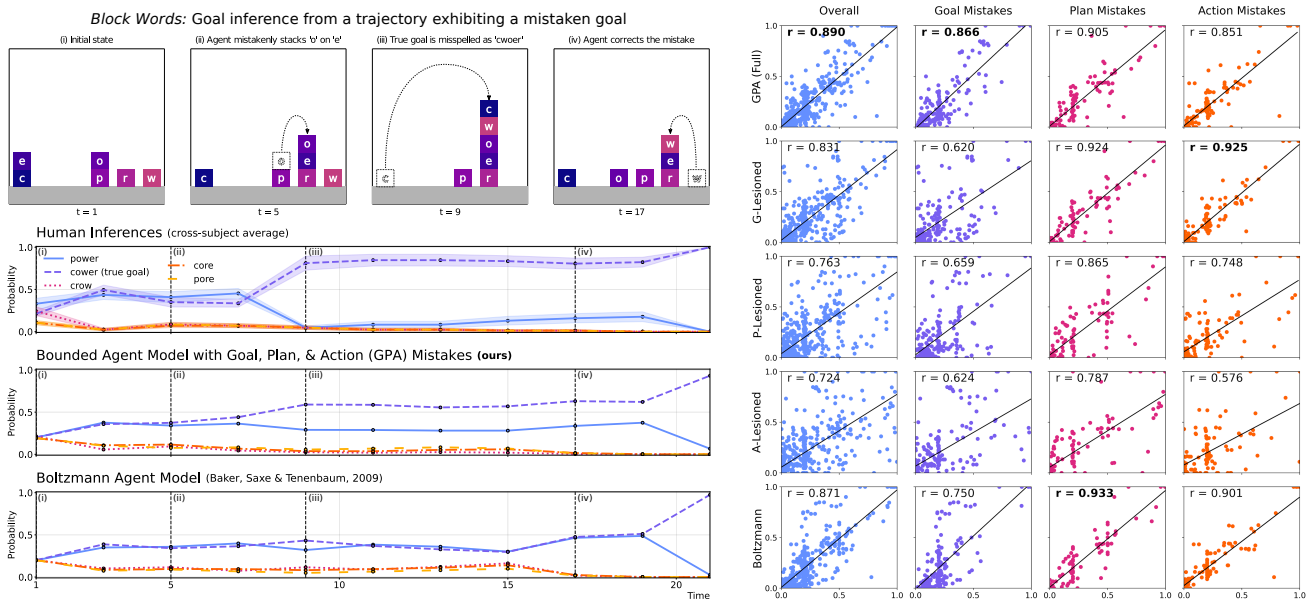
3) Results: After collecting human data, we fit the parameters of each model to maximize correlation between human and model inferences for every stimulus, embedding the assumption that humans flexibly adjust their inferences about different individuals so as to best explain their behavior. Figure 3 shows the resulting inferences. We present an illustrative example from the Doors, Keys, & Gems domain in Figure 3(a), where an agent locks their desired gem out of reach due to myopic planning. The panels show how the agent mistakenly uses up a key to unlock the first door to the red gem, instead of going to collect the other two keys. They then approach the second door, and are stuck. Below these panels, we show average human goal inferences over time (with stardard error ribbons), alongside the inferences produced from our boundedly-rational agent model and the baseline Boltzmann agent model. Humans are able to recognize the true goal as soon as the first door is unlocked (t=8) despite knowing that the agent won't be able to reach it. Our model, which accounts for plan and action (PA) mistakes, exhibits highly similar behavior to humans, placing high confidence in the red gem once the door is unlocked (t=8). For the Boltzmann agent model, however, the probability of the red gem decreases at t=8, and more weight goes to the yellow gem, since it becomes the only gem reachable without any keys.

We present the correlations between human and model inferences in Figure 3(b). The left column shows correlations across all 16 stimuli, with our full model (PA) best explaining overall human inferences (r=0.88). The next two columns show correlations for stimuli with mistaken plans, leading to failure and backtracking respectively. In both cases, our full model fits the human data best. The final column shows correlations only for stimuli with action mistakes, and here our model performs slightly worse than models that account only for action noise (P-Lesioned, Boltzmann). Notably, however, these action-only models perform much more poorly on mistaken plans, mirroring how the A-Lesioned model performs much worse with mistaken actions.

B. Experiment 2: Block Words

In this Blocks World variant adapted from [24], blocks are labeled with letters, and an agent may pick up a block, put it down, or stack it on top of another. The agent's goal is to stack a block tower that spells (top-down) one out of five English words, which are provided to the observer in advance. We assume that planning is guided by a domain-general relaxed distance heuristic [25], leading agents to occasionally neglect how stacking one block on another will prevent the block underneath from being reached. Due to the complexity of goals in this domain, we account for temporary goal confusion, where the agent tries to spell a random permutation of the original word, for example, "cwoer" instead of "cower" (Figure 4(a)).

1) Stimuli: As in Experiment 1, we designed 16 stimuli, organized into four subsets: (1) A control set of optimal trajectories. (2) Trajectories with mistaken goals (misspellings of the intended block tower). (3) Trajectories with mistaken plans, where the agent stacks a block on top of other block(s) it will later need. (4) Trajectories with mistaken actions, where the agent drops a block in an unintended location, or picks up a block adjacent to the intended one. Each participant accessed a web interface which presented 10



(a) Goal inferences over time for a trajectory with a mistaken goal.

(b) Human (y-axis) vs. model (x-axis) inferences.

Fig. 4: Results for the Block Words domain. In **(a)**, we show goal inferences of humans (avg., n=27, w. standard error), our model (GPA) and the Boltzmann agent model on an illustrative trajectory. In **(b)**, we compare human vs. model inferences across models (full model, lesioned models for goal/plan/action mistakes, Boltzmann model) and stimulus types.

stimuli in random order, with at least two stimuli from each subset. Stimuli were presented as animated videos which paused every two actions (picking and placing a block). At these pauses, subjects provided goal inferences by selecting which word(s) they believed to be the most likely goal. Participants could select multiple words if more than one seemed equally likely, and these responses were converted to probability distributions.

2) Participants: We recruited 32 US participants (mean age 40.8, SD 12.5; 13 women, 19 men, 0 non-binary/other) via AMT, restricting to those with a HIT approval rate of 99% and above. Participants went through a tutorial and answered five comprehension questions before proceeding to the stimuli. Following Experiment 1, we awarded points proportional to the probability they assigned to the true goal (mean score 37, SD 7). Five participants were excluded from our analysis, either for failing two or more comprehension checks, or for failing to reach a threshold of 20 points.

3) Results: As in Experiment 1, we fit the parameters of each model to maximize correlation between human inferences and model inferences for every stimulus. The resulting inferences are presented in Figure 4. We show an illustrative stimulus with goal confusion in Figure 4(a). The panels show the agent intending to spell the word "cower" but misspells it as "cwoer" instead. Below these panels, we show average human inferences over time (with standard error ribbons), alongside the results for our bounded agent model and the baseline Boltzmann agent model. Humans are able to identify the true goal as soon as all the letters in "cower" are stacked, despite the wrong order (t=9). When the agent corrects their mistake (t=17), humans remain confident in this inference. Our model, which encompasses goal, plan, and action (GPA) mistakes, exhibits similar behaviour, assigning higher credence to "cower" from t=9 onwards. In contrast, the Boltzmann model fails to account for "cwoer" as a misspelling, giving equal weight to "cower" and "power" until the very last step, when "cower" is correctly spelled.

We compare human vs. model inferences across all models and stimuli in Figure 4(b). The left column shows the overall correlation across all 16 stimuli, indicating that our full model (GPA) best explains human inferences ($r = 0.89$). The next few columns show correlations for each category of mistake. We see that our full model fits human data best when goal mistakes are present, compared to models which do not account for goal errors (G-lesioned, Boltzmann, etc.). The full model also does well with plan and action mistakes, though slightly out-performed by other models. Surprisingly, the Boltzmann model correlates highly even when trajectories exhibit short-sighted planning mistakes. We hypothesize that this is because the plan mistakes in our dataset are sufficiently short-lived (i.e. require backtracking only 1–2 actions) that they are readily attributed to action noise. In contrast, the model without action mistakes (A-lesioned), performs worst across the board, indicating the importance of modeling execution errors in this domain.

IV. DISCUSSION

Our experimental results suggest that humans are robust to observed mistakes when inferring the goals of others. By comparing collected human inferences against inferences produced by our models, we find considerable support for our hypothesis that human inferences are better explained

by a Bayesian Theory of Mind that accounts for mistaken goals, plans, and actions. In Doors, Keys & Gems, we find that models which only account for mistaken actions are not robust to mistaken plans, and vice versa, while in Block Words, we find that models which do not account for goal confusion result in poor inferences when mistaken goals are observed. Together, these findings indicate that it is important to model *distinct* errors at *multiple* levels of cognition and action in order to understand others' goals, and to build an intuitive theory of boundedly rational minds.

REFERENCES

[1] H. Wimmer and J. Perner, "Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception," *Cognition*, vol. 13, no. 1, pp. 103–128, 1983.

[2] J. Phillips, A. Morris, and F. Cushman, "How we know what not to think," *Trends in cognitive sciences*, 2019.

[3] F. Callaway, F. Lieder, P. Das, S. Gul, P. M. Krueger, and T. Griffiths, "A resource-rational analysis of human planning." in *CogSci*, 2018.

[4] C. A. Schult and H. M. Wellman, "Explaining human movements and actions: Children's understanding of the limits of psychological explanation," *Cognition*, vol. 62, no. 3, pp. 291–324, 1997.

[5] H. M. Wellman and K. H. Lagattuta, "Theory of mind for learning and teaching: The nature and role of explanation," *Cognitive development*, vol. 19, no. 4, pp. 479–497, 2004.

[6] L. Young, F. Cushman, M. Hauser, and R. Saxe, "The neural basis of the interaction between theory of mind and moral judgment," *Proceedings of the National Academy of Sciences*, vol. 104, no. 20, pp. 8235–8240, 2007.

[7] F. Cushman, "Crime and punishment: Distinguishing the roles of causal and intentional analyses in moral judgment," *Cognition*, vol. 108, no. 2, pp. 353–380, 2008.

[8] F. Warneken and M. Tomasello, "Altruistic helping in human infants and young chimpanzees," *Science*, vol. 311, no. 5765, pp. 1301–1303, 2006.

[9] C. L. Baker, R. Saxe, and J. B. Tenenbaum, "Action understanding as inverse planning," *Cognition*, vol. 113, no. 3, pp. 329–349, 2009.

[10] J. Jara-Ettinger, H. Gweon, L. E. Schulz, and J. B. Tenenbaum, "The naïve utility calculus: Computational principles underlying commonsense psychology," *Trends in cognitive sciences*, vol. 20, no. 8, pp. 589–604, 2016.

[11] S. Liu, T. D. Ullman, J. B. Tenenbaum, and E. S. Spelke, "Ten-month-old infants infer the value of goals from the costs of actions," *Science*, vol. 358, no. 6366, pp. 1038–1041, 2017.

[12] C. L. Baker, J. Jara-Ettinger, R. Saxe, and J. B. Tenenbaum, "Rational quantitative attribution of beliefs, desires and percepts in human mentalizing," *Nature Human Behaviour*, vol. 1, no. 4, pp. 1–10, 2017.

[13] O. Evans, A. Stuhlmüller, and N. Goodman, "Learning the preferences of ignorant, inconsistent agents," in *Thirtieth AAAI Conference on Artificial Intelligence*, 2016.

[14] O. Evans and N. D. Goodman, "Learning the preferences of bounded agents," in *NIPS Workshop on Bounded Optimality*, vol. 6, 2015.

[15] M. Kryven, T. Ullman, W. Cowan, and J. Tenenbaum, "Outcome or strategy? a bayesian model of intelligence attribution." in *CogSci*, 2016.

[16] B. D. Ziebart, A. L. Maas, J. A. Bagnell, and A. K. Dey, "Maximum entropy inverse reinforcement learning." in *AAAI*, vol. 8. Chicago, IL, USA, 2008, pp. 1433–1438.

[17] T. Otter, J. Johnson, J. Rieskamp, G. M. Allenby, J. D. Brazell, A. Diederich, J. W. Hutchinson, S. MacEachern, S. Ruan, and J. Townsend, "Sequential sampling models of choice: Some recent advances," *Marketing letters*, vol. 19, no. 3, pp. 255–267, 2008.

[18] A. Bobu, D. R. Scobee, J. F. Fisac, S. S. Sastry, and A. D. Dragan, "Less is more: Rethinking probabilistic models of human behavior," in *Proceedings of the 2020 ACM/IEEE International Conference on Human-Robot Interaction*, 2020, pp. 429–437.

[19] T. Zhi-Xuan, J. Mann, T. Silver, J. Tenenbaum, and V. Mansinghka, "Online bayesian goal inference for boundedly rational planning agents," *Advances in Neural Information Processing Systems*, vol. 33, 2020.

[20] D. Ramachandran and E. Amir, "Bayesian inverse reinforcement learning." in *IJCAI*, vol. 7, 2007, pp. 2586–2591.

[21] M. E. Bratman, D. J. Israel, and M. E. Pollack, "Plans and resource-bounded practical reasoning," *Computational intelligence*, vol. 4, no. 3, pp. 349–355, 1988.

[22] J. Lee, J. Fong, B. C. Kok, , and H. Soh, "Getting to know one another: Calibrating intent, capabilities and trust for human-robot collaboration," in *IEEE/RSJ International Conference on Intelligent Robots and Systems*, October 2020.

[23] D. McDermott, M. Ghallab, A. Howe, C. Knoblock, A. Ram, M. Veloso, D. Weld, and D. Wilkins, "PDDL - the Planning Domain Definition Language," 1998.

[24] M. Ramírez and H. Geffner, "Probabilistic plan recognition using off-the-shelf classical planners," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 24, no. 1, 2010.

[25] H. Geffner, "Heuristics, planning and cognition," *Heuristics, Probability and Causality. A Tribute to Judea Pearl. College Publications*, 2010.