Mental state inference from indirect evidence through Bayesian event reconstruction

Michael Lopez-Brau, Joseph Kwon, Julian Jara-Ettinger
{michael.lopez-brau, joseph.kwon, julian.jara-ettinger}@yale.edu
Department of Psychology, Yale University
New Haven, CT 06511 USA

Abstract

From childhood, people routinely explain each other's behavior in terms of inferred mental states, like beliefs and desires. In many cases, however, people can also infer the mental states of agents whose behavior we cannot see, such as when we infer that someone was anxious upon encountering a chewed-up pencil, or that someone left in a hurry if they left the door open. Here we present a computational model of mental-state attribution that works by reconstructing the actions an agent took, based on the indirect evidence that revealed their presence. Our model quantitatively fits participant judgments, outperforming a simple alternative cue-based account. Our results shed light on how people infer mental states from minimal indirect evidence, and provides further support to the idea that human Theory of Mind is instantiated as a probabilistic generative model of how unobservable mental states produce observable action.

Keywords: Theory of Mind; Computational modeling; Social cognition

Introduction

As social creatures, people have a natural capacity to detect other agents (Neri et al., 1998; Scholl & Tremoulet, 2000; Heider & Simmel, 1944), and infer their goals, beliefs, and desires based on how they act (Woodward, 1998; Repacholi & Gopnik, 1997; Aboody et al., 2018). In some cases, however, people can even infer the mental states of agents whose behavior we did not get the opportunity to see. Imagine, for instance, hiking through a national park and encountering a stack of rocks, about two feet tall. Even if you had never seen something like this before, you would quickly realize that someone must have built it, you would have a sense of how they achieved it (perhaps picking up rocks that they found nearby and throwing them haphazardly onto a pile), and you would even have a reasonable guess as to why they did so (to help hikers know that they are on the right path).

Our ability to make inferences about others' minds is supported by an abstract theory-like understanding of how unobservable mental states relate to observable behavior—a *Theory of Mind* (Gopnik et al., 1997; Wellman, 2014). Recent work suggests that Theory of Mind is structured around an assumption that agents act to maximize utilities. That is, we expect agents to maximize their subjective rewards while minimizing the costs that they incur (Jara-Ettinger et al., 2016; Lucas et al., 2014). Under this account, mental-state inference is a process of identifying the costs (agents' competence and dispreferences) and rewards (agents' desires) under which the observed behavior maximizes the underlying utility function. This approach thus formulates Theory of Mind as transforming observable actions into unobservable mental states. Yet, as the example above shows, we can often infer the mental states of agents whose behavior we cannot see. What principles guide these types of inferences?

Computational theories of social cognition have often argued that Theory of Mind is instantiated as a generative model that allows us to sample utility-maximizing action plans as a function of different hypothetical mental states (Jara-Ettinger et al., 2019, 2016; Baker et al., 2017, 2009; Jern et al., 2017). Building on this work, we propose that inferences about the mental states of unobservable agents are supported by a type of event reconstruction, where, upon seeing indirect evidence of someone's presence, we jointly infer the mental states and corresponding actions that explain how the observable evidence arose.

While related research has found that people can infer each other's personality based on indirect evidence (such as seeing someone's bedroom; Gosling et al., 2002), to our knowledge, no work has analyzed people's capacity to infer mental states from indirect evidence. In this paper we present a computational model of mental-state attribution from agent-less physical scenes. Given indirect evidence that someone was present, our model infers what the agent was doing through a generative model of how mental states produce actions and how actions leave observable evidence. In Experiment 1, participants saw a single pile of cookie crumbs left in a room and were asked to infer the agent's entry point and goal (see Fig. 1). We show that our model predicts participant inferences with quantitative accuracy. We contrast our model with a simple cue-based account trained to fit participant judgments through superficial features of the environment (such as the relative distance between the indirect evidence and the possible goals). We find that our model significantly outperforms the cue-based account, even though the latter is directly trained on participant data. Our model predicts that people should not only be able to infer the goals and actions of an absent agent, but also the number of agents that may have been in a room, based on the indirect evidence. We test this prediction in Experiment 2. Once again, we find that participant's confidence about the number of agents that were in a room is quantitatively predicted by our model. Combined, our results suggest that people can quickly reconstruct agents' behavior from minimal indirect evidence that reveals their presence.

Computational Framework

To make our focus concrete, consider a situation like the ones shown in Fig. 1a-b. Each of these displays represents a room with three possible goals (A in blue, B in orange, and C in green), two different doors (1 at the top in both rooms and 2 on the left and bottom, respectively), a set of walls (shown in dark grey), and a small pile of cookie crumbs revealing that someone was previously in this room. Although we cannot see where this agent came from, what goal they were pursuing, or the actions that they took, the cookie crumbs nonetheless contain information that we readily extract. In Fig. 1a, we can infer that the agent was clearly pursuing goal C, despite not being able to tell which door they came from. In Fig. 1b, the cookie crumbs reveal that the agent entered through door 1, but is unclear whether they intended to pursue goal A or goal C. Our computational model aims to explain how we performed these inferences.

Our model builds on past work that formalizes mental-state attribution as Bayesian inference over a generative model of utility-maximizing action plans (Baker et al., 2017; Jara-Ettinger et al., 2019). In our model, however, rather than evaluating unobservable mental states against observable actions, we perform a joint inference over mental states and actions that, combined, explains the visible indirect evidence.

Formally, we model the environment as a gridworld, where the possible states of the world are given by the different positions in space that an agent can occupy. At each time step, we assume that the agent can move in any of the four cardinal directions, and that these actions successfully move the agent in their intended direction (except when the agent attempts to cross a wall). We further assume that all actions incur a cost of 1, and that reaching each reward location yields an unknown reward (set as a uniform distribution over the range $0-100$).

Given a static scene s (a gridworld with a set of goals, doors, walls, and a pile of cookie crumbs), the posterior probability that an agent entered through door d and took trajectory $t = (\vec{s},\vec{a})$ (a sequence of pairs of states and actions that the agent took) to complete goal g is given by

$$p(t,g,d|s) \propto \ell(s|t,g,d)p(t|g,d)p(g)p(d) \tag{1}$$

obtained by combining Bayes' theorem with the chain rule. Here, $\ell(s|g,t,d)$ is the likelihood of observing scene s if the agent indeed completed g by taking trajectory t, defined as $\frac{1}{|t|}$ if the pile cookie crumbs lie within the trajectory (to account for the uniform probability of the agent dropping them anywhere along the path), and 0 otherwise.

$p(t|g,d)$ is the probability that the agent would take trajectory t if they entered from door d with the intention to fulfill goal g. This probability was set to 0 if the trajectory did not begin at door d. To implement the expectation that agents are more likely to complete their goals efficiently (and hence maximize their utilities; Csibra et al., 2003), we used a Markov Decision Process (MDP)—a planning framework that makes it possible to compute the exact action plan or *policy* that maximizes an agent's utility function (Bellman, 1957). Classical MDPs produce a single utility-maximizing policy. In our model, we instead used a probabilistic MDP that creates a probability distribution over all possible action plans, obtained by softmaxing the value function (see Jara-Ettinger et al., 2019, for implementation details). Here we used the softmax parameter $\tau_{action} = 0.15$, which was set prior to data collection. The probability of a trajectory, given a goal and a door, is thus given by

$$p(t|g,d) = \prod_{i=1}^{N} p(a_i|s_i,g) \tag{2}$$

where N is the length of the action plan and $p(a_i|s_i,g)$ is the probability of taking action a_i in state s_i, where the state sequence is derived from trajectory t.

Next, $p(g)$ is the prior distribution over goals. This probability depends both on the costs and rewards associated with the goal. To compute this term, we softmaxed the expected utility associated with each goal, given by the expected reward minus the expected navigation cost (using softmax parameter $\tau_{goal} = 0.10$; also set prior to data collection). Finally, $p(d)$ is the prior distribution over which door the agent entered through, which we set as a uniform distribution.

We implemented this inference procedure through Monte Carlo sampling, where we sequentially sampled entrance points, goals, and trajectories. Each model prediction was obtained by sampling 1000 combinations of entrance points and goals, and 1000 trajectories conditioned on the selected entrance and destination. Fig. 1c-d visualize model posterior inferences. Each line corresponds to a sample from the posterior distribution, color coded to indicate time. This shows how our model reconstructs the agents' probable behavior to determine where they came from and what goal they were pursuing.

Experiment 1

In Experiment 1 we tested our model in a task where people had to infer which goal an agent was pursuing and where they came from, all from a single piece of indirect evidence about their presence.

Participants

40 U.S. participants (as determined by their IP address) were recruited using Amazon Mechanical Turk ($M = 37.03$ years, $SD = 11.20$ years).

Stimuli

Stimuli consisted of 23 gridworld images, like those in Fig. 1a-b (see `https://tinyurl.com/syt4q5h` for full stimuli). Each gridworld was 7-by-7 squares in size and represented a room that contains three goal squares (A in blue, B in orange, and C in green), up to three doors (labeled 1, 2, and 3), and a pile of cookie crumbs. The goals were always in the same corners, but the position of the doors and the pile of cookie crumbs varied between gridworlds. In addition to

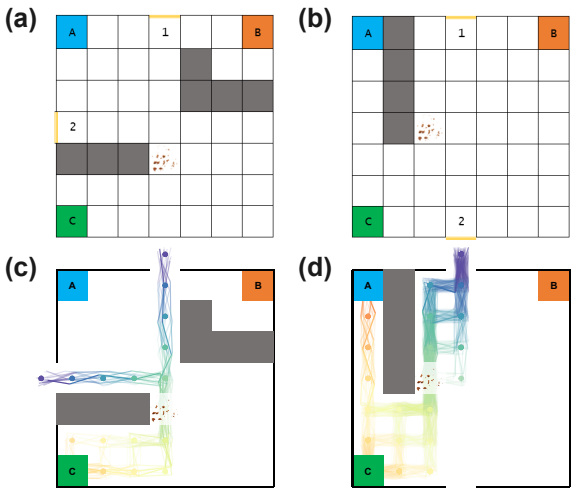

Figure 1: (a-b) Example stimuli from Experiment 1. Potential goals are positioned in the corners, labeled alphabetically, and color coded. Doors are shown in yellow and coded numerically. Walls are shown in dark grey. Each trial included a set of cookie crumbs positioned in a part of the room. (c-d) Visualization of the underlying event reconstruction performed by our computational model for (a) and (b), respectively. Each line represents an inferred possible path, color coded in time, moving from cool to warm. For visibility purposes, only paths with probability greater than 0.001 are shown (although our model is probabilistic, inefficient paths do not appear here because their probability did not surpass the visualization threshold).

these three features, a subset of trials included walls (shown by the dark grey squares in Fig. 1a-b) that agents could not walk through.

Our stimuli set was designed to capture different types of inferences while also controlling for features that simple heuristics could exploit (i.e., ensuring that the target goal was not always the one closest to the cookie crumbs, and that it could not be determined by projecting a straight line that intersected the entrance and cookie crumb location). We began by considering four different possible inference patterns: full certainty (assigning probability close to 1 to a hypothesis; *D* trials), full negative certainty (assigning probability close to 0 to a hypothesis, while also not having full certainty over two remaining hypotheses; *N* trials), partial certainty (assigning a higher probability to one of the hypotheses; *P* trials), and no certainty (assigning a uniform distribution to the hypothesis space; *U* trials).

We first designed seven single-door trials that captured each of these inference patterns in goal inference (two *D*, *N*, and *P* trials, and one *U* trial; Fig. 3a). We then built 16 additional trials by combining every possible inference pattern for the goal the agent was pursing and the entrance that they took (Fig. 3b).

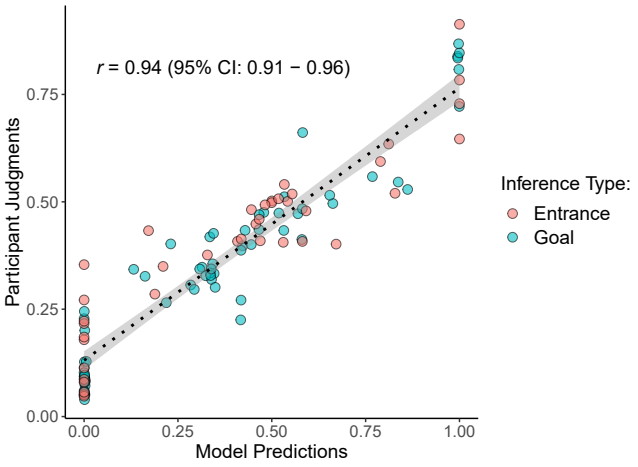

Figure 2: Experiment 1 results. Each point corresponds to a judgment, with model predictions on the *x*-axis and average participant judgments on the *y*-axis. Color indicates inference type and the dotted line shows the best linear fit with 95% confidence bands (in light grey).

Procedure

Participants read a brief tutorial that explained the logic of the task. After learning how to interpret the images, participants were told that agents were equally likely to enter from any the possible doors (to prevent people from assuming that agents intentionally pick doors that are closest to their desired goal) with the aim of going directly to one of the three goals (to remove the possibility that the agent pursued multiple goals, or wandered aimlessly before selecting one). After the introduction, participants completed a questionnaire that verified that they had read the instructions. Participants that failed at least one question were redirected to the beginning of the instructions and those that failed the questionnaire twice were not permitted to participate in the study.

Participants completed all 23 trials in a random order. In each trial, participants had to answer a multiple-choice attention-check question ("Which corner is farthest from Door 1 (there may be more than one)?") and infer the agent's goal ("Which corner is the person going for?") using three continuous sliders, one for each goal (each ranging from 0, labeled as "definitely no", to 1, labelled as "definitely"). Trials with at least two doors included a third question asking participants to infer the agent's entrance point ("Which door did they come from?") using one slider per door. Participants were allowed to submit their responses for each trial only when they correctly answered the attention-check question. Otherwise, participants were told to "please pay attention and try again."

Results

Participant judgments were first normalized within each trial (such that every distribution over goals or doors added up to 1) and then averaged across participants. Fig. 2 shows

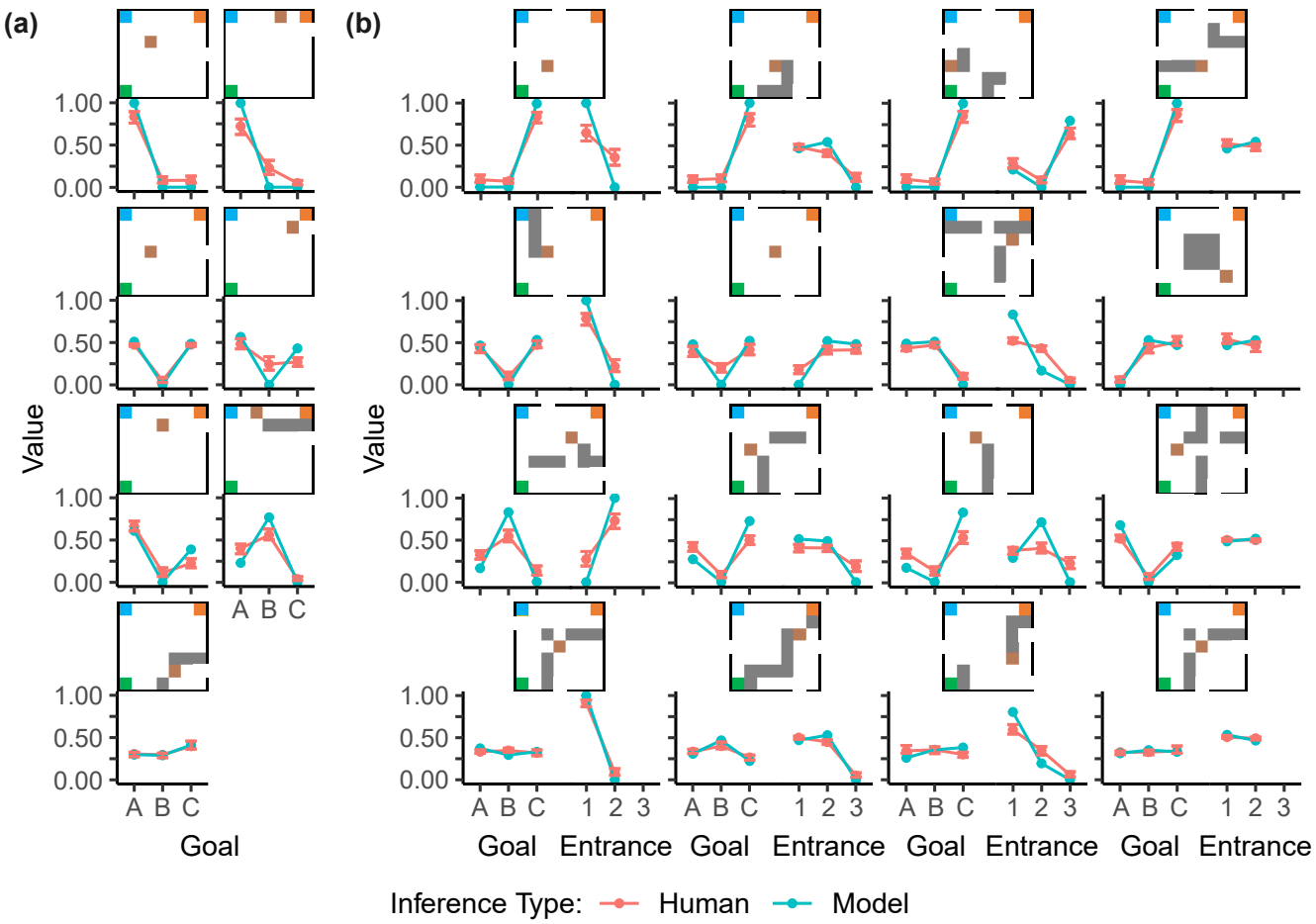

Figure 3: Detailed results for Experiment 1. From top to bottom, each row corresponds to the *D*, *N*, *P*, and *U* trials for goal inferences. (a) Results for trials that only had one door. (b) Results for trials that had more than one door. From left to right, each column of plots corresponds to the *D*, *N*, *P*, and *U* trials for door inferences. The goals A, B, and C are indicated by the blue, orange, and green squares, respectively. The doors are sequentially numbered in a clockwise fashion, with door 1 starting from the top (or from the right if there is no door there). Red lines represent average participant judgments and blue lines show our model's predictions. All participant judgments have 95% bootstrapped confidence intervals.

the results from Experiment 1. Overall, our model showed a correlation of $r = 0.94$ with participant judgments (95% CI: $0.91 - 0.96$), and the strength of the model fit was similar when looking only at goal inferences ($r = 0.95$, 95% CI: $0.93 - 0.97$) and door inferences ($r = 0.92$, 95% CI: $0.86 - 0.95$). Fig. 3a-b shows our model results split by trial, showing the tight correspondence between our model's predictions and participant judgments.

One alternative possibility is that participants judgments were driven by superficial features of the stimuli, rather than by performing Bayesian inference over a generative model of event reconstruction. We tested this possibility through a multinomial logistic regression that predicts participants' distribution over goals as a function of the distance between the pile of cookie crumbs and each goal, the average distance between the pile of cookie crumbs and each door (as opposed to using the distance to each door, allowing us to use the data from all trials since our stimuli varies on door count), the number of doors, and all of their interactions. To train this regression, we transformed participant judgments into a one-hot vector, marking 1 for the goal with the highest probability and 0 for the rest, and implemented LASSO regularization (Tibshirani, 1996) to avoid overfitting. To test this regression, we performed leave-one-out cross-validation (LOOCV).

Even though this alternative model had access to the qualitative structure of participant judgments, it nonetheless produced a correlation of $r = 0.48$ (95% CI: $0.29 - 0.63$), which was substantially lower than the one of our model ($\Delta r = 0.46$; 95% CI: $0.33 - 0.66$). These results show that, while superficial features can capture the broad structure of participant judgments, they fail to do so at our model's level of granularity.

Experiment 2

Experiment 1 shows that people can infer an agent's goal and entrance from a single piece of indirect evidence about their

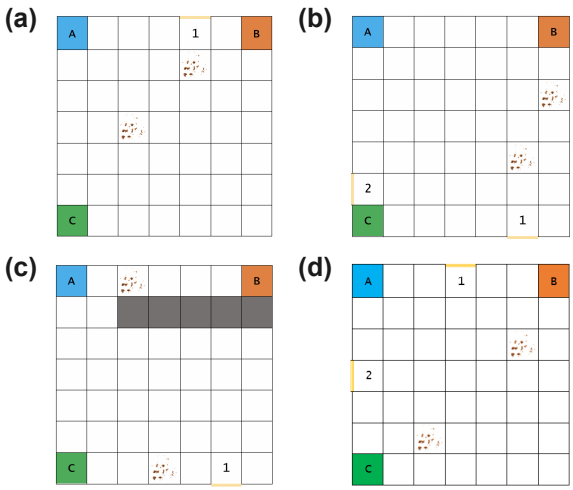

Figure 4: (a-d) Example stimuli from Experiment 2 for *D1*, *P1*, *P2*, and *D2* trials, respectively. Potential goals are positioned in the corners, labeled alphabetically, and color coded. Doors are shown in yellow and coded numerically. Walls are shown in dark grey. Each trial included two piles of cookie crumbs positioned in various parts of the room.

presence. In Experiment 2, we test a further prediction of our account. If our model of event reconstruction is correct, then people should not only be able to infer an agent's probable actions, but also infer the number of agents that may have been in a given scene.

Participants

40 U.S. participants (as determined by their IP address) were recruited using Amazon Mechanical Turk ($M = 37.63$ years, $SD = 11.94$ years).

Stimuli

Our stimuli consisted of 15 gridworld images that were identical to those in Experiment 1 with the difference that each trial has two piles of cookie crumbs (see Fig. 4 for examples; `https://tinyurl.com/syt4q5h` for full stimuli). As in Experiment 1, our stimuli set was designed to capture different types of inferences that our model supports. We considered five different possible inference patterns: full certainty that one agent was in the room (assigning probability close to 1 to a hypothesis; *D1* trials), partial certainty that one agent was in the room (assigning probability between 0.5 and 1 to a hypothesis; *P1* trials), full uncertainty whether it was one or two agents in the room (assigning probability close to 0.5 to a hypothesis; *UN* trials), partial certainty that two agents were in the room (assigning probability between 0 and 0.5 to a hypothesis; *P2* trials), and full certainty that two agents were in the room (assigning probability close to 1 to a hypothesis; *D2* trials). We built three different stimuli for each inference pattern.

Procedure

The procedure was nearly identical to Experiment 1. After reading a cover story that explained that their task was to infer if one or two agents had been in the room, participants completed a questionnaire to ensure they read the instructions. As in Experiment 1, only participants that answered all of the questions correctly were given access to the experiment. Participants that got at least one question wrong were re-directed to the beginning of the tutorial and given a second chance to complete the inclusion questionnaire.

Participants completed all 15 trials in a random order. In each trial, participants answered a multiple-choice attention-check question ("Which corner is the farthest walk from Door 1? If there is more than one correct answer, just choose one of them.") and our key question ("How many people were in the room?") using a continuous slider (ranging from 0, labelled as "definitely one", to 1, labelled as "definitely two"). As in Experiment 1, participants were allowed to submit their responses for each trial only when they correctly answered the attention-check question. Otherwise, participants were told to "please pay attention and try again."

Results

Participant judgments were averaged across trials and compared against our model's predictions. To obtain these predictions, we computed the probability that two agents were in the room by first computing (1) the likelihood that two trajectories explain the scene ($\ell(s|t_1,t_2,g_1,g_2,d_1,d_2)$) and (2) the likelihood that one trajectory explains the scene ($\ell(s|t,g,d)$; from Eq. 1) and then normalizing the first likelihood using both terms.

We computed the likelihood that two trajectories explain the scene by modifying our generative model to sample two sets of entrance points, goals, and trajectories at a time instead of one. Similar to the likelihood term in Eq. 1, this likelihood is defined as $\frac{1}{|t|}$ if both piles of cookie crumbs lie within both trajectories—specifically, the set union of the set of states that define each trajectory—and 0 otherwise.

Fig. 5 shows the results from Experiment 2. Participant's relative confidence about the number of agents in the scene was quantitatively similar to our model's predictions, yielding a correlation of $r = 0.78$ (95% CI: $0.45 - 0.92$). Like in Experiment 1, we proposed the alternative possibility that participant judgments were driven by superficial features of the stimuli rather than invoking Bayesian inference over a generative model of event reconstruction. We tested this possibility through a linear regression that predicts participants' distribution over the number of agents they thought were in the room as a function of the distance between each goal and each pile of cookie crumbs, the average distance between each pile of cookie crumbs and the doors, the number of doors, and all their interactions. We trained and tested this regression using LOOCV and implemented LASSO regularization to avoid overfitting.

Even though this alternative model had access to the qual-

itative structure of participant judgments, it nonetheless produced a correlation of $r = -0.06$ (95% CI: $-0.47 - 0.61$), which was substantially lower than the one of our model ($\Delta r = 0.83$; 95% CI: $-0.13 - 1.34$). These results extend our findings in Experiment 1, suggesting that people can not only infer what an agent may have done from indirect evidence, but also the number of agents that may have been present in the area.

Discussion

Research on human action understanding has historically focused on how we infer the goals and mental states of agents whose behavior we are observing. Our results show that our capacity to reason about others goes beyond face-to-face interactions. In Experiment 1, we showed that people can infer an agent's desires (where an agent was going) and past actions (where an agent came from) from just a single piece of indirect evidence about their presence. In Experiment 2, we showed that people can even infer the number of agents that were in a room, all from indirect patterns that reveal what may have happened.

Here we focused on inferences we make upon recognizing that an agent was present. Thus, our work does not speak to how we recognize this evidence in the first place. We do not know whether our ability to identify evidence that reveals the presence of an agent is guided by similar inferences to those from our model, or by more superficial visual features (such as structure or statistical rarity). We are currently investigating this question.

Our computational model formalized these inferences as the process of reconstructing the goals and behavior that can explain the indirect observable evidence. Our model's quantitative fit with participant judgments, as well the failure of our alternative models in both experiments (despite being trained on participant judgments), suggests that people were performing similar computations. Nonetheless, neither of our experiments directly tested participants' ability to explicitly reconstruct an agents' path. We are currently testing this ability to see if participants do so in a way similar to our model.

While our model was able to quantitatively predict participant judgments in both experiments, the model fit was notably higher in Experiment 1 (Fig. 2) relative to Experiment 2 (Fig. 5). Interestingly, the amount of computation necessary for Experiment 2 is substantially larger than the one necessary for Experiment 1, as it requires reconstructing multiple possible paths that combined explained the observed scene. In this sense, the lower fit in Experiment 2 may be additional evidence that participant inferences are indeed supported by some form of event reconstruction. In current work, we are testing if participant errors are predicted by the number of Monte Carlo samples necessary to approximate our model's normative inference.

While our focus here was on human adults, one open question is when and how this ability develops. Related research has shown that even infants can reason about other agents' goals, preferences, and desires (Wellman, 2014), and that they can even infer the presence of an agent from indirect evidence (Saxe et al., 2005; Newman et al., 2010). Nonetheless, it is unknown if infants can combine these abilities to simultaneously detect the presence of an agent from indirect environmental evidence, and infer the corresponding mental states that explain how this evidence arose.

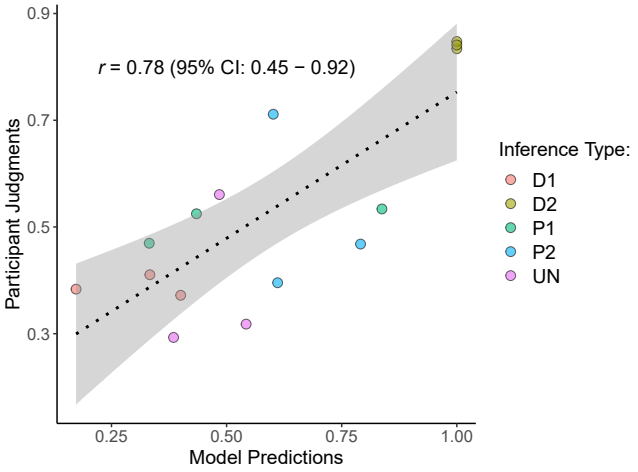

Figure 5: Experiment 2 results. Each point corresponds to a trial, with model predictions on the x-axis and average participant judgments on the y-axis. Color indicates inference type and the dotted line shows the best linear fit with 95% confidence bands (in light grey).

Overall, our results show the sophistication of human social intelligence. Beyond being able to read the mental states of agents that we are interacting with, we can also infer the mental states of agents we have never encountered, just from minimal indirect evidence that reveals their presence. Researchers have long argued that humans are unique in their ability to reason about and navigate the social world (Hermann et al., 2007). Our work shows that this ability is not confined to social interactions, but fundamentally affects how we reason about the physical world, allowing us to see meaning embedded in physical structures, like a pile of rocks, where other animals see merely just that: a pile of rocks.

Acknowledgments

We thank Alan Jern for useful comments. This work was supported by a Google Faculty Research award. This material is based upon work supported by the Center for Brains, Minds, and Machines (CBMM), funded by NSF-STC award CCF1231216.

References

Aboody, R., Huey, H., & Jara-Ettinger, J. (2018). Success does not imply knowledge: Preschoolers believe that accurate predictions reveal prior knowledge, but accurate observations do not. In *Cogsci.*

Baker, C. L., Jara-Ettinger, J., Saxe, R., & Tenenbaum, J. B. (2017). Rational quantitative attribution of beliefs, desires and percepts in human mentalizing. *Nature Human Behaviour*.

Baker, C. L., Saxe, R., & Tenenbaum, J. B. (2009). Action understanding as inverse planning. *Cognition*.

Bellman, R. (1957). A markovian decision process. *Journal of mathematics and mechanics*, 679–684.

Csibra, G., Bíró, S., Koós, O., & Gergely, G. (2003). One-year-old infants use teleological representations of actions productively. *Cognitive Science*, *27*(1), 111–133.

Gopnik, A., Meltzoff, A. N., & Bryant, P. (1997). *Words, thoughts, and theories.* Mit Press Cambridge, MA.

Gosling, S. D., Ko, S. J., Mannarelli, T., & Morris, M. E. (2002). A room with a cue: personality judgments based on offices and bedrooms. *Journal of personality and social psychology*, *82*(3), 379.

Heider, F., & Simmel, M. (1944). An experimental study of apparent behavior. *The American journal of psychology*.

Herrmann, E., Call, J., Hernàndez-Lloreda, M. V., Hare, B., & Tomasello, M. (2007, September). Humans have evolved specialized skills of social cognition: the cultural intelligence hypothesis. *Science*, *317*(5843), 1360–1366.

Jara-Ettinger, J., Gweon, H., Schulz, L. E., & Tenenbaum, J. B. (2016). The naïve utility calculus: Computational principles underlying commonsense psychology. *Trends in cognitive sciences*, *20*(8), 589–604.

Jara-Ettinger, J., Schulz, L., & Tenenbaum, J. (2019, Dec). The naive utility calculus as a unified, quantitative framework for action understanding. Retrieved from psyarxiv.com/e8xsv doi: 10.31234/osf.io/e8xsv

Jern, A., Lucas, C. G., & Kemp, C. (2017). People learn other people's preferences through inverse decision-making. *Cognition*, *168*, 46–64.

Lucas, C. G., Griffiths, T. L., Xu, F., Fawcett, C., Gopnik, A., Kushnir, T., ... Hu, J. (2014). The child as econometrician: A rational model of preference understanding in children. *PloS one*, *9*(3), e92160.

Neri, P., Morrone, M. C., & Burr, D. C. (1998). Seeing biological motion. *Nature*, *395*(6705), 894–896.

Newman, G. E., Keil, F. C., Kuhlmeier, V. A., & Wynn, K. (2010). Early understandings of the link between agents and order. *PNAS*.

Repacholi, B., & Gopnik, A. (1997). Early reasoning about desires: Evidence from 14-and 18-month-olds. *Developmental Psychology*, *33*(1), 12–20.

Saxe, R., Tenenbaum, J., & Carey, S. (2005). Secret agents: inferences about hidden causes by 10- and 12-month-old infants. *Psychological Science*, *16*(12), 995–1001.

Scholl, B. J., & Tremoulet, P. D. (2000). Perceptual causality and animacy. *Trends in cognitive sciences*, *4*(8), 299–309.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, *58*(1), 267–288.

Wellman, H. M. (2014). *Making minds: How theory of mind develops*. Oxford University Press.

Woodward, A. L. (1998). Infants selectively encode the goal object of an actor's reach. *Cognition*, *69*(1), 1–34.