Social inferences from physical evidence via Bayesian event reconstruction

Michael Lopez-Brau, Joseph Kwon, Julian Jara-Ettinger

Department of Psychology, Yale University

Abstract

Humans can make remarkable social inferences by watching each other's behavior. In many cases, however, people can also make social inferences about agents whose behavior they cannot see, based only on the physical evidence left behind. We hypothesized that this capacity is supported by a form of mental event reconstruction. Under this account, observers derive social inferences by reconstructing the agent's behavior, based on the physical evidence that revealed their presence. We present a computational model of this idea, embedded in a Bayesian framework for action understanding, and show that its predictions match human inferences with high quantitative accuracy. Specifically, Experiment 1 shows that people can infer where an agent came from and which goal they pursued in a room, all from a small pile of cookie crumbs. Experiment 2 shows that people can explicitly reconstruct the actions that the agent took, and these reconstructed trajectories can predict the entry point and goal inferences from Experiment 1. Finally, Experiment 3 shows that people can also infer whether one or two agents were in a room based on the position of two piles of cookie crumbs. Our results shed light on how people extract social information from the physical world.

Key words: Computational modeling, Event reconstruction, Social cognition, Theory of Mind

1. Introduction

As social animals, humans possess a specialized cognitive system to process, understand, and predict each other's behavior, known as a *Theory of Mind* (Gopnik et al., 1997; Wellman, 2014). Theoretical and empirical work suggests that human Theory of Mind is instantiated as a mental model that specifies the causal relation between other people's unobservable mental states and their observable actions. That is, Theory of Mind captures how we expect other people's thoughts, preferences, and feelings to guide what they do. Equipped with this intuitive theory, people can infer the mental states that causally give rise to other people's observed behavior.

A rapidly growing body of work suggests that the causal model within Theory of Mind is structured around an assumption that agents act to maximize their utilities—the difference between the subjective costs they incur and the subjective rewards they obtain—capturing the idea that we intuitively expect others to act rationally and efficiently (see Jara-Ettinger, 2019, for review). Consistent with this view, computational models of mental-state inference via utility maximization reach human-level performance on simple social tasks (Baker et al., 2017; Jern et al., 2017; Jern & Kemp, 2015; Jern et al., 2011; Jara-Ettinger et al., 2020), they capture richer forms of social behavior including pedagogy (Bridgers et al., 2020; Ho et al., 2019) and moral reasoning (Ullman et al., 2009), they explain social reasoning in early childhood and infancy (Gergely & Csibra, 2003; Jara-Ettinger et al., 2016; Liu et al., 2017; Lucas et al., 2014), and they have identifiable neural correlates (Collette et al., 2017).

Despite its success, this approach implicitly posits that mental-state inference requires access to someone's observable behavior, as it is these observed actions that enable us to evaluate the plausibility of different mental states. In many cases, however, people can make social inferences about agents whose behavior we did not get the opportunity to see. For example, imagine walking into an office building and finding a vacant receptionist desk with a chewed-up pencil, a half-filled crossword puzzle, and a cellphone. From this arrangement of objects, we can immediately infer that the receptionist might have been experiencing anxiety or restlessness (as the pencil was chewed-up), that they were likely procrastinating or had few tasks to complete at the moment (as they were working on a crossword), and that they expected to be gone only momentarily (as they chose to leave their valuable belongings unattended).

As the examples above show, human social inference is not limited to an ability to extract social information from observable actions—we can also make social inferences from physical scenes with no direct social or temporal information. How do we achieve this and how fine-grained are these inferences? Here we propose that social inferences about unobservable agents are supported by a basic form of *event reconstruction*, where, upon seeing indirect evidence of an agent's presence, we reconstruct what actions they likely took, enabling us to reason about the agent's behavior in a similar way to how we would if we had seen them act first-hand.

While it has long been known that the ability to infer social information from observed actions emerges early in infancy (Gergely & Csibra, 2003; Onishi & Baillargeon, 2005; Woodward, 1998), recent studies suggest that social reasoning from physical events also emerges early in childhood. By preschool, children can estimate the difficulty associated with building different physical arrangements of objects (Gweon et al., 2017); they understand which kinds of actions leave physical traces in the environment and which kinds of actions do not (Jacobs et al., 2021); they can infer what someone knew based on physical evidence for how they searched an area (Pelz et al., 2020); and they can even detect the transmission of ideas by comparing artifacts created by different agents (Pesowski et al., 2020).

This past research suggests that the capacities needed to perform social inference via event reconstruction might be in place from childhood. However, to our knowledge, no work has formally explored the event reconstruction hypothesis that we propose here. Specifically, we hypothesized that people can causally reason about how goals lead to actions, and how actions leave traces in the environment. Combining these two causal models enables people to understand how goals lead to observable traces in the environments, connected by an inferred internal variable consisting of the actions that the agent took, which we call an event reconstruction. Here we present a computational model of this idea, testing social reasoning from agent-less physical scenes. Given indirect evidence that someone was present, our model infers what the agent was doing (i.e., reconstructs their actions) and why (i.e., infers their goals) through a generative model of how goals produce actions, and how actions leave observable evidence.

1.1. Connection to related proposals in social psychology

Consistent with our proposal, research in social psychology has found that people leave "behavioral residues" in their environments: physical cues that support rich inferences about their personality traits. For example, by looking at a picture of someone's messy desk, people can infer that the inhabitant is likely disorganized. From similar displays, people can also infer the inhabitant's degree of extraversion, conscientiousness, and even openness to new experiences (Webb et al., 1966; Gosling et al., 2002, 2008).

These inferences have been proposed to stem from a two-stage process, where people first use physical cues (such as a desk's cleanliness, the amount of books in the room, or the cheerfulness of the décor) to infer someone's behavior, and then use this behavior to infer the underlying dispositions (Gosling et al., 2002; Brunswik, 1956). In this model, *cue utilization* captures how people transform these cues into social inferences, and *cue validity* captures whether these are accurate. Our hypothesis is consistent with this model, and it can be thought of as proposing that *cue utilization* consists of a form of Bayesian event reconstruction. From this standpoint, our work can be thought of as proposing a mechanism for how people associate different physical traces to the underlying behavior. Our work contributes to this literature by proposing a fully specified computational theory behind event reconstruction, grounded in the expectation that agents act rationally and efficiently in their environment, given their goals. Critically, however, previous models also account for inferences that people make based on stereotypes—a process that is outside of the scope of our work. We return to this point in the Discussion.

1.2. The current work

In Experiment 1, we first tested whether our model matched human inferences in a task where participants had to infer an agent's entry point into a room and their goal, all from a single pile of cookie crumbs that revealed their presence (see Figure 1). In Experiment 2, we then explicitly tested people's ability to reconstruct the actions they believe different agents took based on indirect physical evidence of their presence, lending further support to the idea that the inferences in Experiment 1 were supported by an ability to reconstruct events. Finally, if social reasoning from physical scenes is supported by event reconstruction, people should be able to also infer how many agents might have been present in a room, based on how many paths they need to reconstruct to explain the scene. We tested this prediction in Experiment 3. Combined, our results suggest that people have a nuanced capacity to infer social information from indirect evidence, and that these inferences are based on a basic capacity to "enhance" physical scenes by inferring agents' spatiotemporal behavior based on the indirect evidence that they leave behind. All studies were approved by the Yale University Institutional Review Board (protocol: "Online reasoning" #2000020357).

2. Computational Framework

Our model builds on a growing body of work showing that mental-state attribution is instantiated as Bayesian inference over a generative model of utility-maximizing action plans (Baker et al., 2009, 2017; Jara-Ettinger et al., 2020; Jern et al., 2017; Jern & Kemp, 2015; Jern et al., 2011; Lucas et al., 2014). In our model, however, rather than evaluating unobservable goals against observable actions, we model how people might use physical evidence to reconstruct the actions that an agent took, and use these reconstructed actions to attribute goals.

To make our focus concrete, consider a situation like the ones shown in Figure 1a. 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 bottom and left, respectively), a set of walls (shown in dark gray), and a small pile of cookie crumbs that reveals that someone was previously in this room. Although we cannot see where this agent came from, what actions they took, or what goal they were pursuing, the cookie crumbs nonetheless contain information that we might be able to extract. In Figure 1a (left), the cookie crumbs intuitively reveal that the agent entered through door 1 and that they were likely pursuing goal A or C, but not goal B. In Figure 1a (right), the cookie crumbs intuitively reveal that the agent was pursuing goal C, but it is unclear whether they entered through door 1 or door 2. Our computational model aims to explain how we performed these inferences.

Formally, we model the environment as a gridworld, where the possible states of the world are given by the different positions in space that agents can occupy. At each time step, we assume that agents can move in any of the four cardinal directions and that these actions successfully move them in their intended direction (except when attempting to cross a wall, in which case the agent remains in the same position as they were before).

Given an observed static scene s (a gridworld with a set of goals, doors, walls, and a pile of cookie crumbs), the objective is to infer where the agent entered the room from (a door d) and which goal they pursued (a goal g), formally

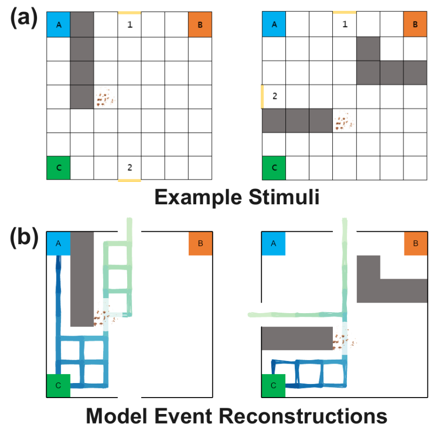

Figure 1: (a) Example stimuli from Experiment 1. Potential goals are positioned in the corners, labeled alphabetically, and color-coded. Doors are shown in yellow and coded numerically. Walls are shown in dark gray. Each trial included a pile of cookie crumbs positioned in a part of the room. (b) Visualizations of the underlying event reconstruction performed by our computational model for the examples above. Each line represents an inferred possible path, color-coded to indicate time, moving from light green to dark blue.

expressed as

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

where $\ell(s|d, g)$ is the likelihood of encountering scene s if an agent had indeed pursued goal g after entering through door d, and $p(d, g)$ is the prior over doors and goals.

According to our proposal, the ability to compute the likelihood function is mediated by a capacity to reconstruct the agent's actions. Under this view, if we can reconstruct the actions that the agent took, then judgments about the agent's entry point and goal are immediately revealed, as these are part of the reconstructed behavior (i.e., if we have access to the full reconstructed behavior, we can "see" where the agent entered from and where they were going). Formally, this idea can be implemented by expressing the likelihood function as

$$\ell(s|d, g) = \sum_{t \in \mathbb{T}} \underbrace{p(s|t)}_{\substack{\text{how do actions} \\ \text{leave traces?}}} \times \overbrace{p(t|d, g)}^{\substack{\text{how do agents} \\ \text{pursue goals?}}}. \tag{2}$$

Here $t = (\vec{s}, \vec{a})$ is a trajectory (from the set of all possible trajectories $\mathbb{T}$), which consists of an ordered sequence of pairs of states and actions that the agent took.

$p(s|t)$ is the probability that an agent who took trajectory t would produce the observed scene s, and $p(t|g,d)$ is the probability that the agent would take trajectory t if they entered from door d with the intention to pursue goal g. This equation reveals the two components critical to our theory: an expectation of how agents navigate to complete their goals ($p(t|d,g)$), and an expectation of how agents' actions leave observable traces in the environment ($p(s|t)$).

To compute the expectations for how agents complete their goals, we used the standard framework previously developed in computational models of goal inference (Baker et al., 2009, 2017; Jara-Ettinger et al., 2020) through Markov Decision Processes (MDPs)—a planning framework that makes it possible to compute the action plan or *policy* that maximizes an agent's utility function (Bellman, 1957). Classical MDPs are designed to produce a single trajectory that fulfills the agent's goal as efficiently as possible. In the cases that we consider, however, there are often multiple trajectories that can be equally efficient. As such, using a simple MDP can erroneously treat an efficient trajectory as unlikely if it is not an exact match to the solution that the MDP produced. To solve this problem, we built a probabilistic MDP that creates a probability distribution over all possible action plans, assigning higher probability to trajectories that are more efficient. Formally, we achieved this by softmaxing the MDP's value function when building the probabilistic policy. We used a low temperature parameter to identify all possible action plans that are equally (or approximately equally) efficient, enabling us to implement the expectation that agents navigate efficiently towards their goals. Using a probabilistic MDP, the probability that an agent would take trajectory t, starting from door d with the intention to fulfill goal g is given by

$$p(t|g,d) = \prod_{i=1}^{|t|} p(a_i|s_i,g), \tag{3}$$

where $p(a_i|s_i,g)$ is the probability of taking action a_i in state s_i, and the state sequence is given by trajectory t.

Finally, in our paradigm, we assume that the agent has a uniform probability of dropping the pile of cookie crumbs at any point in their path. The probability of observing scene s if the agent took trajectory t is therefore given by $p(s|t) = 1/|t|$ if the pile of cookie crumbs lies within the trajectory and 0 otherwise.

2.1. Implementation Details

To generate testable predictions, we set a number of parameters in our model prior to data collection. These choices are all reflected in our pre-registered model predictions (see `https://osf.io/q3ct5/`). We began by setting a uniform prior distribution over doors and goals, such that agents were equally likely to enter through any of the doors and equally likely to pursue any of the goals. Next, to model the forces that shape agents' actions, we assumed that agents incur a constant cost of 1 for any action that they take, and that goals produced numerical rewards over the range $0-100$. Finally, to make our MDP probabilistic, we applied a temperature parameter $\tau = 0.15$ to the value function. This parameter was set *a priori* to ensure that the model would give equal probability to all paths that were equally efficient, while only placing a negligible probability on erroneous and inefficient trajectories.

Model inferences were obtained via Monte Carlo methods, sampling 1000 combinations of doors and goals and 1000 trajectories conditioned on the selected door and goal. Figure 1b visualizes our model's inferred trajectories for the examples shown in Figure 1a, with each line corresponding to a sample from the posterior distribution, color-coded to indicate time, moving from light green to dark blue. These visualizations show how our model reconstructs the agent's probable spatiotemporal behavior, which in turn reveal the agent's entry point and goal, matching the intuitive inferences associated with these examples in the introduction.

3. Experiment 1a

In Experiment 1a, 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. If people's ability to infer goals from physical evidence is mediated by event reconstruction, then their judgments should show a quantitative fit to our model predictions, including fine-grained patterns of uncertainty. This study was pre-registered; all study materials can be found at `https://osf.io/q3ct5/`.

3.1. Participants

40 U.S. participants (as determined by their IP address) were recruited using Amazon Mechanical Turk ($M = 37.02$ years, $SD = 11.20$ years).

3.2. Stimuli

Stimuli consisted of 23 gridworld images, like those in Figure 1a. 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 trials. In addition to these three features, a subset of trials included walls (shown by the dark gray squares in Figure 1a) 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 (e.g., 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 the location of the cookie crumbs). We began by considering four different possible inference patterns: assigning probability close to 1 to a hypothesis (HIGH CERTAINTY trials), assigning probability close to 0 to a hypothesis, while also not having full certainty over two remaining hypotheses (HIGH NEGATIVE CERTAINTY trials), assigning a higher probability to one of the hypotheses (PARTIAL CERTAINTY trials), and assigning an approximately uniform distribution to the hypothesis space (UNCERTAIN trials).

We first designed seven single-door trials that captured each of these inference patterns in goal inference (two HIGH CERTAINTY, HIGH NEGATIVE CERTAINTY, and PARTIAL CERTAINTY trials, and one UNCERTAIN trial; schematic versions shown in Figure 3a). We then designed 16 additional trials with multiple doors by combining every possible inference pattern for the goal the agent was pursuing and the entrance that they took (schematic versions shown in Figure 3b).

3.3. 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 the room from any of the doors with the intention of going directly to one of the three goals (to remove the possibility that agents pursue multiple goals, or wander aimlessly before selecting one). After the introduction, participants completed a questionnaire that ensured they read and understood the instructions. Participants that failed at least one question were redirected to the beginning of the instructions and given a second chance to participate in

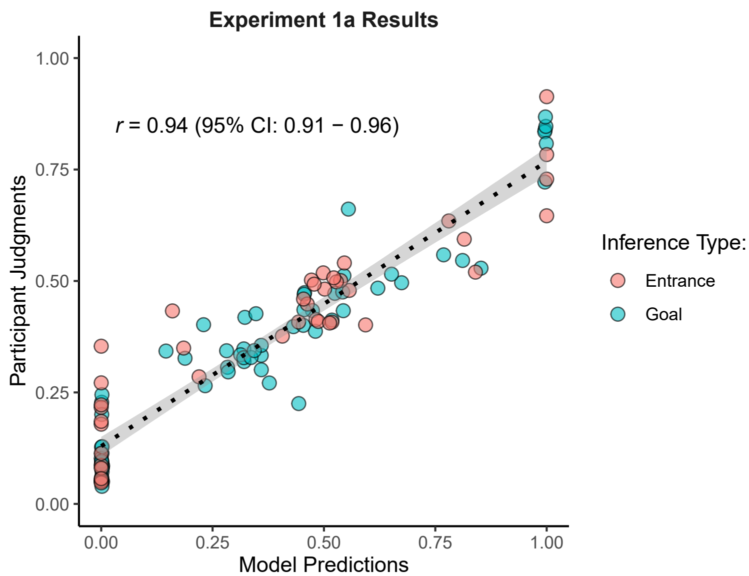

Figure 2: Results from Experiment 1a. Each point corresponds to a judgment, with model predictions on the x-axis and mean participant judgments on the y-axis. Color indicates inference type and the dotted line shows the best linear fit with 95% confidence bands (in light gray).

the study. Participants that failed the questionnaire twice were not permitted to participate in the study.

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

3.4. Results

Each participant's judgments were first normalized within-trial (such that every distribution over goals or doors added up to 1) and then averaged across

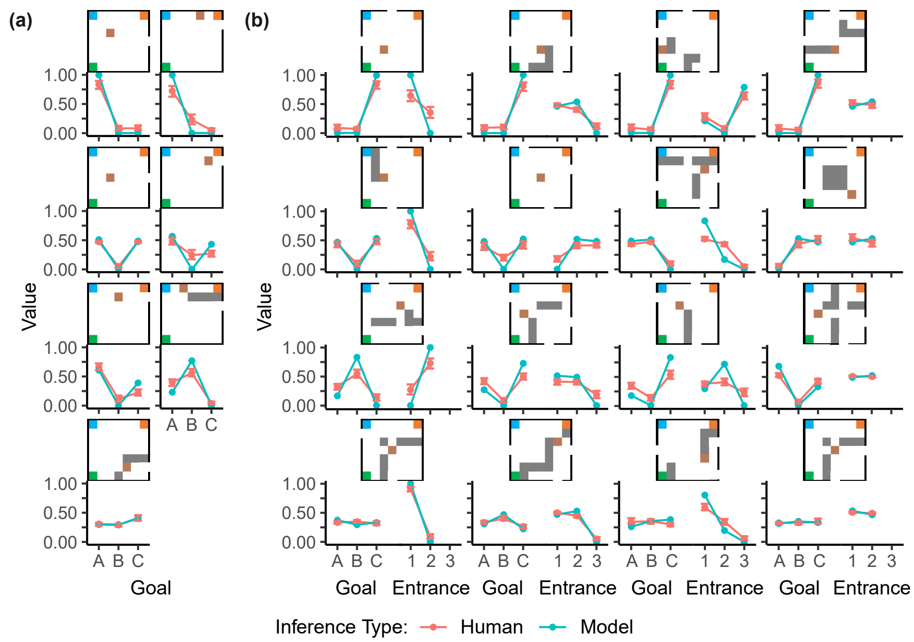

Figure 3: Detailed results from Experiment 1a. From top to bottom, each row of subplots corresponds to the HIGH CERTAINTY, HIGH NEGATIVE CERTAINTY, PARTIAL CERTAINTY, and UNCERTAIN trials for goal inferences, respectively. (a) Results for trials that only had one door. (b) Results for trials that had more than one door. From left to right, each column of subplots corresponds to the HIGH CERTAINTY, HIGH NEGATIVE CERTAINTY, PARTIAL CERTAINTY, and UNCERTAIN trials for door inferences, respectively. 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 top door). Walls are marked as dark gray squares and the pile of cookie crumbs are indicated by the brown squares. Red lines represent mean participant judgments and blue lines represent our model's predictions. Error bars on participant judgments represent 95% bootstrapped confidence intervals.

participants. Figure 2 shows the results from Experiment 1a. Overall, our model showed a correlation of $r = 0.94$ (95% CI: $0.91 - 0.96$) with participant judgments, and the strength of the model fit was similar when looking only at goal inferences ($r = 0.95$; 95% CI: $0.92 - 0.97$) or door inferences ($r = 0.92$; 95% CI: $0.86 - 0.95$).

Figure 3 shows our model's results as a function of trial. In each subplot, the image at the top shows an abstract schematic of the trial, with the pile of cookie crumbs marked as a brown square. This figure reveals how our model not only predicted participant judgments in situations where the agent's entry point and goal were clear, it also matched participant judgments in its expression of uncertainty. Critically, our model produced nuanced patterns of uncertainty across trials, which reflect how well it was able to reconstruct the event, becoming less confident as a function of how much conflict there is in entry points and goals across different hypothetical reconstructions. The fact that this event-based uncertainty matched participant judgments with quantitative accuracy suggests that participants may have also been performing these inferences via some form of event reconstruction.

One possibility is that the underlying goals or entry points of the agent correlate with superficial features of the stimuli, such as the proximity of the cookie crumbs to different doors or goals. If this is the case, then participants may have been able to infer agents' entry points and goals without performing any form of event reconstruction. We tested this possibility through a multinomial logistic regression trained to predict participant goal inferences 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, the number of doors, and all of their interactions. To train this model, 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. We generated the alternative model's predictions in a leave-one-out fashion—that is, the predictions for each trial consisted of the output of a regression trained on all remaining trials.

Even though this alternative model was trained on the qualitative structure of participant judgments, it nonetheless only produced a correlation of $r = 0.49$ (95% CI: $0.30-0.63$) with participant judgments, which was substantially lower than the one produced by our model ($\Delta r = 0.46$; 95% CI: $0.33-0.65$). 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, further suggesting that people's inferences were centered on a form of Bayesian event reconstruction.

4. Experiment 1b

Experiment 1a showed initial evidence for our model in a situation where people had no prior information about the agent. In many cases, however, people have prior knowledge about others, and this information affects their inferences. In Experiment 1b, we therefore tested if our model continued to capture participant inferences in a context where people were given prior information about the agent's behavior. This study was pre-registered; all study materials can be found at `https://osf.io/q3ct5/`.

4.1. Participants

160 English-speaking participants were recruited using Prolific ($M = 33.49$ years, $SD = 11.36$ years).

4.2. Stimuli

Stimuli consisted of 16 gridworld images, evenly divided across a *door prior* and a *goal prior* condition. Each gridworld was similar to those in Experiment 1a, with the difference that each trial now included prior information about an agent's behavior. In the *door prior* condition, each gridworld contained nine red 'X' markers, distributed across the doors. These markers represented the number of times the agent previously entered through each door. In the *goal prior* condition, each gridworld contained nine red 'X' markers, distributed across the three goals. These markers represented the number of times the agent previously pursued each goal.

To construct the stimuli for the *goal prior* condition, we first selected four gridworlds from Experiment 1a's PARTIAL CERTAINTY condition, and four gridworlds from Experiment 1a's UNCERTAIN condition (with respect to goal inferences). For each selected gridworld, we considered four possible prior distributions over the goals: {(3, 3, 3), (6, 2, 1), (1, 6, 2), (2, 1, 6)}. Because this condition consisted of eight gridworlds, each possible prior distribution was randomly assigned to one gridworld from the PARTIAL CERTAINTY set and to one gridworld from the UNCERTAIN set. This assignment was randomized across participants to ensure an equal amount of data for every possible combination of gridworld and prior distribution (resulting in a total of $8 \times 4 = 32$ possible combinations).

The stimuli for the *door prior* condition was designed in a parallel way. We first selected four gridworlds from Experiment 1a's PARTIAL CERTAINTY condition, and four gridworlds from Experiment 1a's UNCERTAIN condition (this time with respect to door inferences). Because all gridworlds from the PARTIAL CERTAINTY condition had three doors, we used the same set of priors and assignment procedure used in our *goal prior* condition described above. By contrast, all gridworlds from the UNCERTAIN condition had two doors. The priors for these trials were therefore sampled from the set {(5, 4), (5, 4), (7, 2), (2, 7)}.[1]

[1] The pre-registered duplication of (5, 4) in the prior set was accidental, as it was meant to

4.3. Procedure

The procedure was nearly identical to Experiment 1a, except that participants were also taught how to read the prior information. Participants were told that, in each gridworld, they would see the agent's entry point or goal (depending on condition) for the agent's nine previous visits, and their task was to infer the agent's entry point and goal for the tenth event. After the introduction, participants completed a questionnaire that ensured they read and understood the instructions. Participants that failed at least one question were redirected to the beginning of the instructions and given a second chance to participate in the study. Participants that failed the questionnaire twice were not permitted to participate in the study.

Participants completed all 16 trials in two experimental blocks, one for the *door prior* condition and another for the *goal prior* condition. Experimental block order and within-block trial order were randomized across participants. The prior information on each trial was determined by one of four distributions (see Stimuli). On 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 were asked to 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, labeled as "definitely"), and asked to infer the agent's entry point ("Which door did they come from?") using one slider per door (each also ranging from 0, labeled as "definitely no," to 1, labeled as "definitely"). Participants were allowed to submit their responses for each trial only when they correctly answered the attention-check question. Otherwise, participants were prompted to "please pay attention and try again."

4.4. Model Predictions

Model predictions were obtained in the same way as Experiment 1a, with the difference that the prior distribution over goals and doors was based on agents' prior behaviors. To achieve this, we began with a uniform distribution over goals and doors for every gridworld, and updated each distribution through Bayes' rule based on the prior behavior (i.e., the nine observations) shown in the gridworld, using the generative process specified in our model (i.e., by assuming

be (4, 5). This affected only 4 of the 64 possible gridworld-by-prior tests, and our experiment continues to have the necessary variability to compare participants to our model.

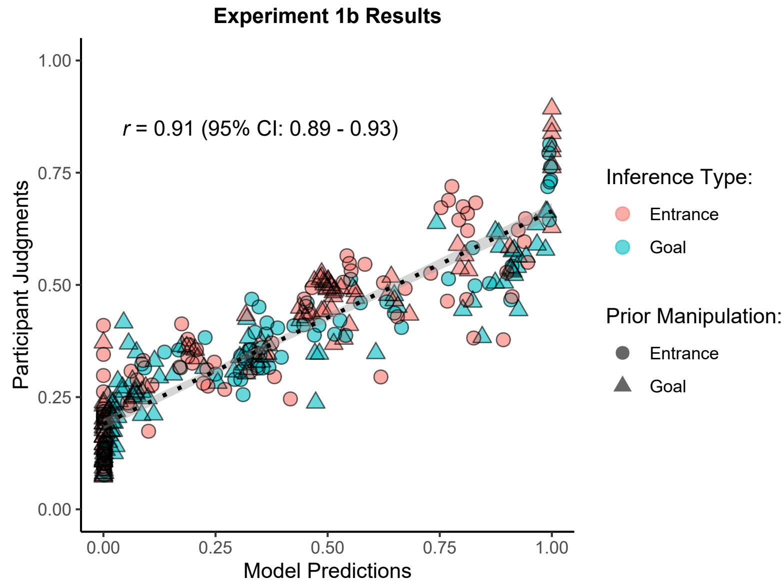

Figure 4: Results from Experiment 1b. Each point corresponds to a judgment, with model predictions on the x-axis and mean participant judgments on the y-axis. Color indicates inference type, shape indicates condition, and the dotted line shows the best linear fit with 95% confidence bands (in light gray).

that agents probabilistically choose the goal with the highest utility, subject to a softmax process with temperature $\tau = 0.1$). The resulting distributions were then set as the prior distributions in the study.

4.5. Results

Data was analyzed in the same way as Experiment 1a. Each participant's judgments were first normalized within-trial (such that every distribution over goals or doors added up to 1) and then averaged across participants for each condition. Figure 4 shows the results from Experiment 1b. Overall, our model showed a correlation of $r = 0.91$ (95% CI: $0.89 - 0.92$) with participant judgments, and the strength of the model fit was similar for the *goal prior* condition ($r = 0.91$; 95% CI: $0.89 - 0.93$) and the *door prior* condition ($r = 0.90$; 95% CI: $0.86 - 0.92$). Critically, these inferences once again revealed that participants produce graded patterns of confidence across trials, as predicted by our model. Together, these results show that people, like our model, can integrate prior information about how an agent behaved to reconstruct their actions given indirect physical evidence.

5. Experiment 2

In Experiment 1, we found that people can infer where an agent was going and where they came from, all from a single piece of indirect evidence about their presence. Participant judgments were quantitatively predicted by a model centered on an ability to reconstruct what happened. If our account is correct, then people should also be able to explicitly reconstruct the actions that an agent took in a way similar to our model. We test this prediction in Experiment 2. This study was pre-registered; all study materials can be found at `https://osf.io/q3ct5/`.

5.1. Participants

40 U.S. participants (as determined by their IP address) were recruited using Amazon Mechanical Turk ($M = 38.25$ years, $SD = 11.02$ years).

5.2. Stimuli

The stimuli were the same as those from Experiment 1a (see Figure 1a for examples and Figure 3 for schematic versions).

5.3. Procedure

Participants read a brief tutorial that explained the logic of the task. Participants were then taught how to draw their paths. After the introduction, participants completed a questionnaire that ensured they read and understood the instructions. Participants that failed at least one question were redirected to the beginning of the instructions and given a second chance to participate in the study. Participants that failed the questionnaire twice were not permitted to participate in the study.

Participants completed all 23 trials in a random order. On each trial, participants were asked to infer the path they thought the agent took, given the pile of cookie crumbs. Participants generated their paths by sequentially clicking on the squares they believed the agent walked through. Participants were only allowed to proceed when they had successfully generated a valid path, which consisted of paths that started at a door, ended at a goal, and passed through the pile of cookie crumbs. Participants were allowed to reset the drawn path as many times as they wished.

5.4. Model Predictions

To evaluate the participant-generated path reconstructions, we used our framework to calculate

$$p(t|s) \propto p(s|t)p(t), \quad (4)$$

where $p(s|t)$ is the likelihood of a trajectory t generating scene s and $p(t)$ is the prior over possible trajectories. Here, $p(s|t) = 1/|t|$ (like in Equation 2) and $p(t)$ is obtained by marginalizing over agents' potential goals and entry points, as follows:

$$p(t) = \sum_{d,g} p(t|d, g)p(d, g). \quad (5)$$

5.5. Results

Our computational framework enables us to calculate the probability assigned to each path generated by participants. However, directly interpreting these probabilities is difficult, as they are sensitive to the length of the path and to the number of competing paths that fulfill a goal efficiently. To make our results easier to interpret, we compared our model's evaluations of the participant-generated path reconstructions with that of a baseline model. This baseline model used a uniform transition function over all actions, excluding the one that would generate a transition to the previous state (to prevent infinite back-and-forth loops). For every participant, we computed the Bayes factor for each of their reconstructed paths by dividing the probability of each path, as predicted by our model (i.e., $p(t|s)$), by the probability predicted by the baseline model. A Bayes factor greater than one would indicate that our model explains a participant's reconstructed path better than the baseline model; a Bayes factor less than one would indicate that the baseline model explains a participant's reconstructed path better than our model.

Our model outperformed the baseline model on all trials. The average Bayes factor in our experiment was 16935.33 (lowest factor = 7933.79; highest factor = 84383.12), meaning that our model was, on average, much more likely to produce the participant-generated path reconstructions relative to the baseline model ($t(39) = 9.10$, $p < 0.001$ using a Bayes factor of 1 as the reference level).

Figure 5 shows trial-by-trial results from Experiment 2. Each trial is presented twice, with our model's path reconstructions on the left and participant-

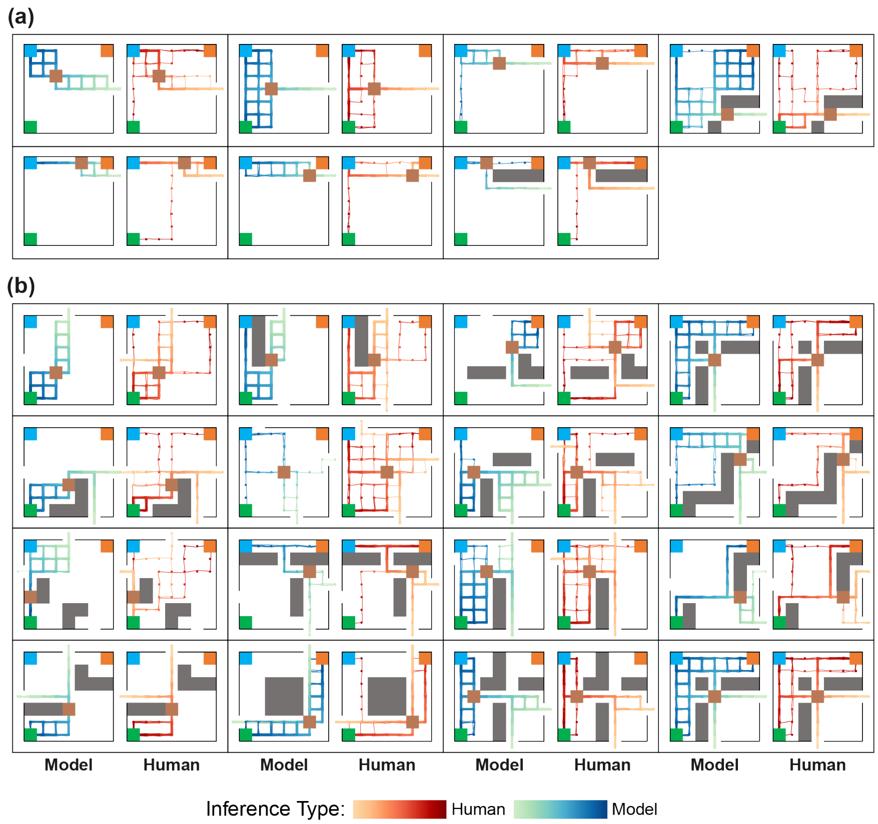

Figure 5: Comparison of reconstructed paths generated by our model and participants in Experiment 2. From left to right, each column of subplots corresponds to the HIGH CERTAINTY, HIGH NEGATIVE CERTAINTY, PARTIAL CERTAINTY, and UNCERTAIN trials for goal inferences, respectively. (a) Results for trials that only had one door. (b) Results for trials that had more than one door. From top to bottom, each row of subplots corresponds to the HIGH CERTAINTY, HIGH NEGATIVE CERTAINTY, PARTIAL CERTAINTY, and UNCERTAIN trials for door inferences, respectively. The goals A, B, and C are indicated by the blue, orange, and green squares, respectively. The doors are sequentially numbered in a clockwise order, with door 1 starting from the top (or from the right if there is no top door). Walls are marked as dark gray squares and the pile of cookie crumbs are indicated by the brown squares. Each line represents a reconstructed path, color-coded to indicate time, moving from light orange to dark red (for participants) or light green to dark blue (for the model).

generated path reconstructions on the right. All paths are color-coded to indicate time (with darker colors occurring later in time). For both our model and participants, the higher path density indicates where the majority inferred the agent to have traveled. As this figure shows, the distribution of participant-generated path reconstructions largely matched those generated by our model (although participants were more likely to generate suboptimal paths).

6. Do explicit event reconstructions in Experiment 2 predict inferences from Experiment 1?

The previous results showed that that people can not only reconstruct agents' actions, but do so in a way similar to our model. According to our proposal, this event reconstruction underlies people's capacity to infer agents' goals and entry points in Experiment 1. If this is the case, then the path reconstructions from Experiment 2 should have predictive power over the inferences that participants made in Experiment 1. To test this possibility, we extracted the goals and doors from the participant-generated path reconstructions. To achieve this, we calculated the proportion of paths that originated from each possible entrance, and the proportion of paths that reached each possible goal, and compared these values to the corresponding goal and door inferences from Experiment 1a. Figure 6 shows the results from this analysis. Overall, the goals and doors extracted from the participant-generated path reconstructions showed a correlation of $r = 0.89$ (95% CI: $0.83 - 0.92$) with the inferences participants made in Experiment 1a, and the strength of this fit was similar when looking only at goals ($r = 0.88$; 95% CI: $0.80 - 0.93$) or doors ($r = 0.90$; 95% CI: $0.82 - 0.95$). Furthermore, when we compared these extracted goals and doors against our model's predictions in Experiment 1a, we found a correlation of $r = 0.86$ (95% CI: $0.79 - 0.91$), and a similar fit when looking only at goals ($r = 0.85$; 95% CI: $0.76 - 0.91$) or doors ($r = 0.88$; 95% CI: $0.78 - 0.93$).

Critically, participants in Experiment 2 could only generate a single path per trial. By combining the paths of multiple participants, we were able to reveal distributions over goals and doors that quantitatively resembled the inferences participants made in Experiment 1a. The fact that these distributions predicted inferences from Experiment 1a suggests that generated paths were samples from the posterior distribution (rather than maximum likelihood or maximum *a posteriori* estimates, which would not contain enough information to reconstruct the full probability distribution over inferences). This analysis suggests that participants in Experiment 2 had access to and sampled paths in accordance to these goal and door distributions.

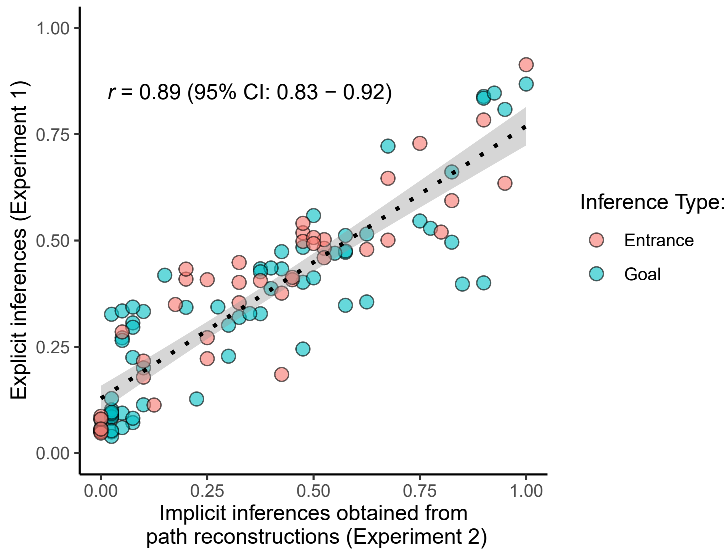

Figure 6: Comparison between the extracted goals and doors from Experiment 2 and the participant inferences from Experiment 1a. Color indicates inference type and the dotted line shows the best linear fit with 95% confidence bands (in light gray).

7. Experiment 3

Experiment 1 showed that people can infer an agent's goals and origins, and that these inferences exhibit the quantitative structure predicted by a model of event reconstruction. Experiment 2 further showed that people could explicitly reconstruct the paths in a way similar to our model. In Experiment 3, 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 a *single* agent's probable actions and goals, but also be able to estimate how many agents might have been in a room, based on how many path reconstructions are needed to explain a given scene. This study was pre-registered; all study materials can be found at `https://osf.io/q3ct5/`.

7.1. Participants

40 U.S. participants (as determined by their IP address) were recruited using Amazon Mechanical Turk ($M = 37.62$ years, $SD = 11.94$ years).

7.2. Stimuli

Stimuli consisted of 15 gridworld images that were similar to those in Experiment 1, with the difference that each trial now has two piles of cookie crumbs instead of one (see Figure 7 for examples). Our stimuli set was designed to capture different types of inferences that our model supports. Specifically, we designed three different trials for each of the following possible inference patterns: high certainty that one agent was in the room (DEFINITELY ONE trials), partial certainty that one agent was in the room (PROBABLY ONE trials), uncertainty whether it was one or two agents in the room (UNCERTAIN trials), partial certainty that two agents were in the room (PROBABLY TWO trials), and high certainty that two agents were in the room (DEFINITELY TWO trials).

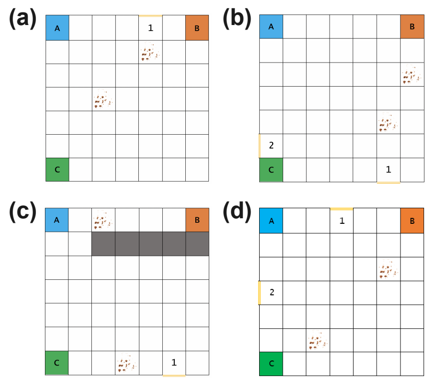

Figure 7: (a-d) Example stimuli from Experiment 3 for DEFINITELY ONE, PROBABLY ONE, PROBABLY TWO, and DEFINITELY TWO trials, respectively (see Experiment 3 Stimuli for details). Potential goals are positioned in the corners, labeled alphabetically, and color-coded. Doors are shown in yellow and coded numerically. Walls are shown in dark gray. Each trial included two piles of cookie crumbs positioned in various parts of the room.

7.3. Procedure

The procedure was nearly identical to Experiment 1a, except that participants were instead shown two piles of cookie crumbs and were told that their task was to infer if one or two agents had been in the room. After the introduction, participants completed a questionnaire that ensured they read and understood the instructions. Participants that failed at least one question were redirected to the beginning of the instructions and given a second chance to participate in the study. Participants that failed the questionnaire twice were not permitted to participate in the study.

Participants completed all 15 trials in a random order. On 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 were asked to infer how many agents were in the room ("How many people were in the room?") using a continuous slider (ranging from 0, labeled as "definitely one," to 1, labeled as "definitely two"). 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."

7.4. Model Predictions

To predict how many agents might have been in a scene we computed the probability that a agents were in scene s, through

$$p(a|s) \propto p(s|a)p(a), \tag{6}$$

where $p(a)$ is a prior over the number of agents that could have been present. In natural contexts, this prior should reflect the statistics of how often different agents might interact in different environments. To model our experiment, however, we used a simple uniform prior over the possibility of having one or two agents. This prior was then weighted by the likelihood of a particular number of agents a generating scene s, given by

$$p(a|s) \propto \begin{cases} \sum_{t \in \mathbb{T}} p(s|t)p(t) & a = 1 \\ \sum_{t_1, t_2 \in \mathbb{T}} p(s|t_1, t_2)p(t_1)p(t_2) & a = 2 \end{cases} \tag{7}$$

To compute the likelihood that two trajectories explain the scene (i.e., $p(s|t_1, t_2)$), we modified our generative model to sample two sets of entry points, goals, and trajectories at a time instead of one, where the likelihood is defined as $1/(|t_1| + |t_2|)$ if there was a scene match (i.e., both piles of cookie crumbs lie within both trajectories, and each trajectory was responsible for one of the piles) and 0 otherwise.

7.5. Results

Participant judgments were averaged across trials and compared against our model's predictions. Figure 8 shows the results from Experiment 3. Participant's relative confidence about the number of agents in the scene was quantitatively similar to our model's predictions, yielding a correlation of $r = 0.76$

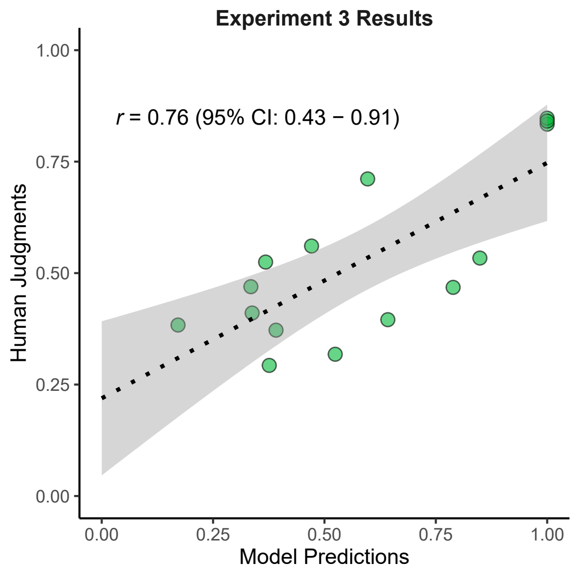

Figure 8: Results from Experiment 3. Each point corresponds to a judgment, with model predictions on the x-axis and mean participant judgments on the y-axis. The dotted line shows the best linear fit with 95% confidence bands (in light gray).

(95% CI: 0.43 − 0.91). As before, participants' pattern of data did not only qualitatively identify the best inference, but also revealed a graded pattern of confidence that is broadly consistent with event reconstruction.

Figure 9 shows our model's results as a function of each trial. In each subplot, the image at the top shows an abstract schematic of the trial, with both piles of cookie crumbs marked as brown squares. From left to right, each column corresponds to the DEFINITELY ONE, PROBABLY ONE, UNCERTAIN, PROBABLY TWO, and DEFINITELY TWO trials, respectively. This figure reveals how our model quantitatively predicts participant judgments across the various trials and levels of uncertainty.

Interestingly, the model fit in Experiment 3 was lower relative to Experiment 1. Under our account, this difference may arise because Experiment 3 requires reconstructing paths for a single agent, reconstructing paths for multiple agents, and weighting their relative probability of generating the observed scene. Consistent with this, we found higher mismatches between our model and participants in the PROBABLY trials ($MSE = 0.053$) over the DEFINITELY ($MSE = 0.021$) and UNCERTAIN trials ($MSE = 0.019$). That is, participants struggled more in trials that relied on a capacity to make precise comparisons between the number of single-agent reconstructions and two-agent reconstructions.

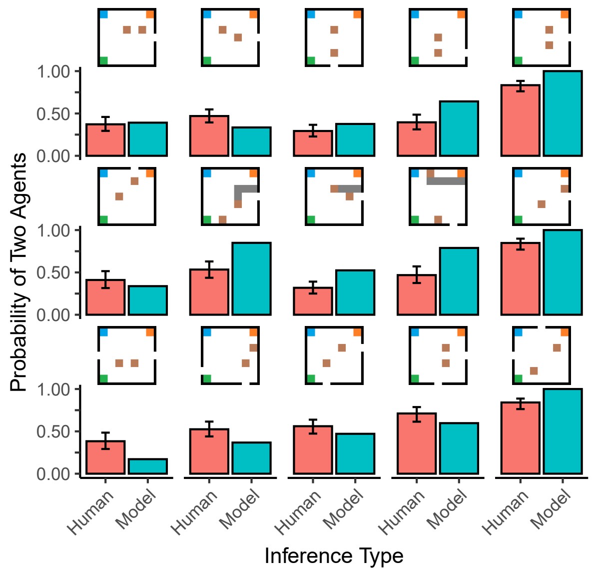

Figure 9: Detailed results from Experiment 3. From left to right, each column corresponds to DEFINITELY ONE, PROBABLY ONE, UNCERTAIN, PROBABLY TWO, and DEFINITELY TWO trials, respectively. Red bars represent mean participant judgments and blue bars represent our model's predictions. Error bars on participant judgments represent 95% bootstrapped confidence intervals.

As in Experiment 1a, we also evaluated whether participant judgments could be explained by superficial features of the stimuli rather than via event reconstruction. We tested this possibility through a logistic regression trained to predict 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 of their interactions. We trained and tested this alternative model in the same way as the one described in Experiment 1a.

Even though this alternative model had access to the qualitative structure of participant judgments, it nonetheless produced a correlation of $r = 0.19$ (95% CI: $-0.30 - 0.66$) with participant judgments, which was substantially lower than the one produced by our model ($\Delta r = 0.58$; 95% CI: $0.12 - 1.17$). These results extend our findings from Experiments 1 and 2, suggesting that people can not only infer an agent's goals and origins based on indirect evidence of their presence, but also whether multiple agents may have been present in a given scene.

8. 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 and includes nuanced social inferences from simple physical scenes. In Experiment 1, we showed that people can infer an agent's goals (i.e., where an agent was going) and past actions (i.e., where an agent came from) from a single piece of indirect evidence about their presence. The tight correspondence between our model's predictions and the fine-grained structure of participant judgments suggested that these inferences were structured around a form of mental event reconstruction: people infer the actions that an agent took, and use this reconstructed behavior to make richer social inferences. Experiment 2 showed further support for our proposal, revealing that people can explicitly reconstruct the actions that someone took based on indirect physical evidence, in a way similar to our model. Furthermore, these explicit reconstructions predicted participant inferences in Experiment 1, showing a direct link between people's ability to reconstruct behavior from physical evidence, and the corresponding social inferences that they make. Finally, in Experiment 3, we found that people can also infer how many agents were in a given scene, based on the number of paths they needed to reconstruct to explain the scene.

8.1. What cognitive capacities are required for event reconstruction?

Our computational model formalized social inferences as the process of reconstructing behaviors that explain the observed physical evidence. Our model's quantitative fit with participant judgments, and the failure of our alternative models (despite being trained on participant judgments), suggests that people were performing similar computations. In particular, the similarity between the paths generated by our model and those drawn by participants (see Figure 5) suggests that social inferences from physical evidence are tied to a form of event reconstruction.

The heart of our proposal—expressed in Equation 2 (see Section 2)—posits that event reconstruction depends on two different cognitive capacities. The first is a model of how agents act in the world. The second is a model of how agents' actions leave observable traces in the environment.

In our model, the first capacity consisted of a simple expectation that agents navigate towards their goals as efficiently as possible, given the environmental constraints. This expectation, known as a *teleological stance* (Gergely, 2003; Gergely & Csibra, 1997), has been hypothesized to be a precursor to mental-state reasoning, supporting simple social inferences without requiring active representations of other people's minds (Gergely & Csibra, 2003). From this standpoint, our computational model shows that a full-fledged Theory of Mind is not necessary for performing social reconstructions from physical evidence, and a teleological stance can suffice.

At the same time, agents with a Theory of Mind might be able to derive richer social inferences. To illustrate this, imagine that a valuable object that was hidden in a closet in someone's house has gone missing. Suppose also that drawers and cabinets throughout the house were left open, but nothing else had been taken. In this situation, a pure teleological stance could reveal that the thieves navigated through the house opening drawers and cabinets. However, a teleological stance alone would end there, failing to reveal *why* the thieves pursued these goals. This event, analyzed through a Theory of Mind, however, would reveal that the thieves knew that the valuable object was in the house, did not know its exact location, and therefore searched the house to find it.

This example raises the possibility that a non-mentalistic teleological stance enables people to reconstruct the actions that an agent took, by assuming that they navigate efficiently in space. Once these actions have been reconstructed, our Theory of Mind might enable us to extract the complex mental states that can explain why the agent took the actions that they did. This is a direction that we hope to explore in future work.

The second capacity implemented in our model is an understanding of how actions leave observable traces in the environment. Our model therefore posits that event reconstruction requires an ability to associate different actions with their corresponding observable traces. Our model used a highly simplified setting where the observable evidence consisted of a small pile of cookie crumbs. In more realistic situations, the types of traces that agents leave behind can be rich and variable, from unambiguous cues like foot tracks on the ground, to more subtle ones, like finding a single apple tree with no apples, in a row of trees full of ripe apples. This suggests that people's capacity to reconstruct behavior is simultaneously powered and constrained by their knowledge of the relationship between actions and physical traces.

While our work focused on adults, some recent research suggests that these capacities might emerge in early childhood. In particular, preschoolers can judge what types of physical constructions (such as different types of block towers) require more physical effort (Gweon et al., 2017), suggesting an early understanding between actions and physical outcomes. Moreover, children can also determine what actions are more likely to leave physical traces. For example, lifting an upside-down cup filled with rice will likely leave visible rice grains after the cup has been repositioned. But it is possible to lift and reposition an upside-down cup filled with a few large rocks without leaving any evidence behind (Jacobs et al., 2021). Recent research has found that children can even associate physical outcomes with the corresponding mental states of the agent who generated them (Pelz et al., 2020). Finally, and most strikingly, young children can infer the transfer of ideas by seeing how different agents create artifacts (Pesowski et al., 2020), a capacity known as "intuitive archaeology" (Hurwitz et al., 2019; Schachner et al., 2018). While these results point towards an early understanding of the relation between the social and physical world, to our knowledge, it is an open question whether these inferences are also linked to some form of explicit or implicit event reconstruction.

Finally, at the highest level, our work builds on the idea that human cognition is structured around mental models (also called intuitive theories) of the world (Tenenbaum et al., 2011), including intuitive theories of the physical world (Battaglia et al., 2013) and of others (Jara-Ettinger et al., 2020). Following this tradition, our model posits that people have (i) a causal understanding of how goals lead to actions and how actions leave observable traces, and (ii) a mechanism for inverting this causal model, enabling people to move from observed traces to the underlying goals. In our model, the inversion mechanism was implemented as Bayesian inference via Monte Carlo simulations. This approach is consistent with growing evidence that action-understanding involves some form of Bayesian inference (Baker et al., 2017; Ullman et al., 2009; Jara-Ettinger et al., 2020). Nonetheless, our work only tested our model at Marr's computational level of analysis (Marr, 1982), and it does not imply that people are specifically using a Monte Carlo based approach to implement Bayesian reasoning. Indeed, related work has found that this type of inference can be approximated via simpler strategies (Bonawitz et al., 2014), and people's inferences in our task might not have required active sampling in participants. At the same time, work in intuitive physics has found some evidence of active sampling in physical reasoning, opening the possibility that this extends to social reasoning as well (Hamrick et al., 2015). These are questions that we also hope to explore in future work.

8.2. Study limitations

Our work has three main limitations. First, our model and experiments focused on highly simplified events. In more realistic situations, the space of goals that an agent might pursue, and the physical evidence they leave behind is substantially more complex than what our two-dimensional gridworlds can capture. To reason about a chewed-up pencil, for example, our model would require a more extensive description of human behavior to compute how an anxious mental state shapes an agent's action space, and how the resulting candidate actions (e.g., chewing) leave traces in the environment. Our proposed model does not currently support social inferences at this level of complexity, and it is an empirical question whether our approach could capture human reasoning in these more naturalistic events.

One way in which our framework could tackle richer inferences is by using a full-fledged model of intuitive physics to evaluate how actions leave traces in the environment. A recent body of work in cognitive science has found that human intuitive physics is instantiated as a *physics engine* that supports rich probabilistic simulations of how objects and forces interact in the environment (Fischer et al., 2016; Battaglia et al., 2013), and that physical simulations might underlie how we reason about the interaction between agents and objects (Yildirim et al., 2019). Thus, using a physics engine to simulate how the forces that agents apply to the world leave observable traces might enable our computational framework to handle more complex physical events that contain social information.

Our second main limitation lies in the narrow range of inferences that we asked people to make: inferences about where an agent was going, where they entered from, and how many agents were involved. As noted above, all of these inferences can be explained through a *teleological stance* (Gergely & Csibra, 2003). Consequently, our work does not test the extent to which people can infer complex mental states or personality traits from physical evidence. Recent work has found that people can indeed make rich communicative inferences from physical arrangements of objects (Lopez-Brau & Jara-Ettinger, 2020; Sarin et al., 2021); however, in this work, the position of the objects unambiguously revealed the agent's actions (they positioned the objects where they were most visible to others). This work therefore leaves open whether the capacity to infer these types of mental states extends to events where people must perform more complex forms of event reconstruction. In future work, we hope to incorporate richer models of mental-state inference to test people's capacity to infer mental states such as beliefs, desires, knowledge, and intentions from physical evidence (Jara-Ettinger et al., 2020; Baker et al., 2017).

Our third limitation is that our work used simple events with minimal social context: participants had nearly no information about the agent, and the goals consisted of simple abstract squares. This enabled us to test people's capacity to reconstruct events in a controlled manner. In more naturalistic situations, however, the content of the goals often reveals important information that can help people build more nuanced inferences. Imagine, for instance, that one of the squares in our stimuli was a work desk, the second one was a stationary bicycle, and the third one was a TV. With this context, the physical trace would not only allow people to infer the agent's goal, but also richer aspects of their personality. Relatedly, when more context is available, people also rely on inferred stereotypes to attribute dispositions (Gosling et al., 2002, 2008). These richer context-based inferences were not captured by our work, and are a critical challenge towards building computational models that fully capture human social reasoning.

Our work also leaves a critical question open. Our experiments focused on situations where people were explicitly told that an agent was previously present. Our work therefore does not speak to how people use physical information to infer that an agent was present in the first place. One possibility is that people engage in a pervasive and constant social analysis of all physical scenes. Doing so, however, might be prohibitively costly and unnecessary. As such, it is likely that people are attuned to the physical signatures that reveal the presence of an agent, which then trigger social reasoning from physical evidence. Consistent with this second view, research suggests that people can infer the presence of an agent based on apparent order (Newman et al., 2010; Keil & Newman, 2015) and on a sensitivity to human-like errors that people leave behind when interacting with the world (Lopez-Brau et al., 2021). An open question is how the ability to detect the presence of an agent interacts with the ability to reconstruct their behavior and infer their mental states.

8.3. Implications and conclusions

At first glance, our computational framework appears to suggest that any creature with some form of naïve psychology and naïve physics ought to be able to perform social inferences from physical evidence (i.e., access to the two key components of Equation 2). This may not be the case, however, because our model also requires an ability to transfer information across these intuitive theories (reconstructing behavior via naïve psychology and evaluating how they compare to the environment via naïve physics). While this is an open empirical question, research suggest that intuitive physics and intuitive psychology rely on separate neural circuitry (Fischer et al., 2016; Saxe & Powell, 2006), leaving open the question of how these two intuitive theories might work in tandem to reconstruct other people's behavior from physical evidence.

One interesting case that suggests such a feat might not be simple comes from research with vervet monkeys. Vervet monkeys have an astonishing degree of social intelligence, including a nuanced repertoire of vocal calls to signal different types of predators, each associated with different escape responses (Seyfarth et al., 1980a,b). Yet, vervet monkeys routinely fail to identify predators from indirect physical evidence. For instance, vervet monkeys fail to infer that a python is hiding in a nearby bush when they encounter the distinct tracks that they leave behind. Similarly, vervet monkeys also fail to infer the presence of a leopard upon encountering a gazelle carcass on a tree (where leopards usually drag their prey so they can feed in solitude; Cheney & Seyfarth, 1985). Critically, this failure appears to persist even after vervet monkeys have, in past events, seen the direct association between the physical evidence and the predator (Cheney & Seyfarth, 1985, 2008). These results might point to the possibility that the form of event reconstruction that we present here might require capacities that go beyond simple physical and social reasoning, as they involve an ability to combine the two capacities to derive richer inferences than would be otherwise possible.

Overall, our results illustrate the sophistication of human social intelligence. Beyond being able to make social inferences about agents that we are personally interacting with, we can also make social inferences about 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 (Herrmann et al., 2007). Our work shows that this ability is not confined to social interactions, but can fundamentally affect how we reason about the physical world, allowing us to see social meaning embedded in physical structures, like a pile of rocks, where other animals may see merely just that: a pile of rocks.

References

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*, *1*, 1–10.

Baker, C. L., Saxe, R., & Tenenbaum, J. B. (2009). Action understanding as inverse planning. *Cognition*, *113*, 329–349.

Battaglia, P. W., Hamrick, J. B., & Tenenbaum, J. B. (2013). Simulation as an engine of physical scene understanding. *Proceedings of the National Academy of Sciences*, *110*, 18327–18332.

Bellman, R. (1957). A markovian decision process. *Journal of Mathematics and Mechanics*, (pp. 679–684).

Bonawitz, E., Denison, S., Gopnik, A., & Griffiths, T. L. (2014). Win-stay, lose-sample: A simple sequential algorithm for approximating bayesian inference. *Cognitive psychology*, *74*, 35–65.

Bridgers, S., Jara-Ettinger, J., & Gweon, H. (2020). Young children consider the expected utility of others' learning to decide what to teach. *Nature Human Behaviour*, *4*, 144–152.

Brunswik, E. (1956). *Perception and the representative design of psychological experiments*. University of California Press.

Cheney, D. L., & Seyfarth, R. M. (1985). Social and non-social knowledge in vervet monkeys. *Philosophical Transactions of the Royal Society of London. B, Biological Sciences*, *308*, 187–201.

Cheney, D. L., & Seyfarth, R. M. (2008). *Baboon metaphysics: The evolution of a social mind*. University of Chicago Press.

Collette, S., Pauli, W. M., Bossaerts, P., & O'Doherty, J. (2017). Neural computations underlying inverse reinforcement learning in the human brain. *Elife*, *6*, e29718.

Fischer, J., Mikhael, J. G., Tenenbaum, J. B., & Kanwisher, N. (2016). Functional neuroanatomy of intuitive physical inference. *Proceedings of the National Academy of Sciences*, *113*, E5072–E5081.

Gergely, G. (2003). The development of teleological versus mentalizing observational learning strategies in infancy. *Bulletin of the Menninger clinic*, *67*, 113–131.

Gergely, G., & Csibra, G. (1997). Teleological reasoning in infancy: The infant's naive theory of rational action: A reply to premack and premack. *Cognition*, *63*, 227–233.

Gergely, G., & Csibra, G. (2003). Teleological reasoning in infancy: The naïve theory of rational action. *Trends in cognitive sciences*, *7*, 287–292.

Gopnik, A., Meltzoff, A. N., & Bryant, P. (1997). Words, thoughts, and theories.

Gosling, S. D., Gaddis, S., & Vazire, S. (2008). First impressions based on the environments we create and inhabit. *First Impressions*, (pp. 334–356).

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*, 379.

Gweon, H., Asaba, M., & Bennett-Pierre, G. (2017). Reverse-engineering the process: Adults' and preschoolers' ability to infer the difficulty of novel tasks. In *CogSci*.

Hamrick, J. B., Smith, K. A., Griffiths, T. L., & Vul, E. (2015). Think again? the amount of mental simulation tracks uncertainty in the outcome. In *CogSci*. Citeseer.

Herrmann, E., Call, J., Hernàndez-Lloreda, M. V., Hare, B., & Tomasello, M. (2007). Humans have evolved specialized skills of social cognition: The cultural intelligence hypothesis. *Science*, *317*, 1360–1366. doi:`10.1126/science.1146282`.

Ho, M. K., Cushman, F. A., Littman, M., & Austerweil, J. L. (2019). Communication in action: Planning and interpreting communicative demonstrations. *Journal of Experimental Psychology: General*, .

Hurwitz, E., Brady, T., & Schachner, A. (2019). Detecting social transmission in the design of artifacts via inverse planning.

Jacobs, C., Lopez-Brau, M., & Jara-Ettinger, J. (2021). What happened here? children integrate physical reasoning to infer actions from indirect evidence. *CogSci*, .

Jara-Ettinger, J. (2019). Theory of mind as inverse reinforcement learning. *Current Opinion in Behavioral Sciences*, *29*, 105–110.

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*, 589–604.

Jara-Ettinger, J., Schulz, L. E., & Tenenbaum, J. B. (2020). The naïve utility calculus as a unified, quantitative framework for action understanding. *Cognitive Psychology*, *123*, 101334.

Jern, A., & Kemp, C. (2015). A decision network account of reasoning about other people's choices. *Cognition*, *142*, 12–38.

Jern, A., Lucas, C., & Kemp, C. (2011). Evaluating the inverse decision-making approach to preference learning. *Advances in Neural Information Processing Systems*, *24*, 2276–2284.

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

Keil, F. C., & Newman, G. E. (2015). Order, order everywhere, and only an agent to think: The cognitive compulsion to infer intentional agents. *Mind & Language*, *30*, 117–139.

Liu, S., Ullman, T. D., Tenenbaum, J. B., & Spelke, E. S. (2017). Ten-month-old infants infer the value of goals from the costs of actions. *Science*, *358*, 1038–1041.

Lopez-Brau, M. (2021). Social inferences from physical evidence. URL: `https://osf.io/q3ct5/`.

Lopez-Brau, M., Colombatto, C., Jara-Ettinger, J., & Scholl, B. (2021). Attentional prioritization for historical traces of agency. *Journal of Vision*, *21*, 2748–2748.

Lopez-Brau, M., & Jara-Ettinger, J. (2020). Physical pragmatics: Inferring the social meaning of objects.

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

Marr, D. (1982). *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*. MIT Press.

Newman, G. E., Keil, F. C., Kuhlmeier, V. A., & Wynn, K. (2010). Early understandings of the link between agents and order. *Proceedings of the National Academy of Sciences*, *107*, 17140–17145.

Onishi, K. H., & Baillargeon, R. (2005). Do 15-month-old infants understand false beliefs? *Science*, *308*, 255–258.

Pelz, M., Schulz, L., & Jara-Ettinger, J. (2020). The signature of all things: Children infer knowledge states from static images. *PsyArXiv*, .

Pesowski, M., Quy, A., Lee, M., & Schachner, A. (2020). Children use inverse planning to detect social transmission in design of artifacts. *PsyArXiv*, .

Sarin, A., Ho, M. K., Martin, J. W., & Cushman, F. A. (2021). Punishment is organized around principles of communicative inference. *Cognition*, *208*, 104544.

Saxe, R., & Powell, L. J. (2006). It's the thought that counts: Specific brain regions for one component of theory of mind. *Psychological Science*, *17*, 692–699.

Schachner, A., Brady, T., Oro, K., & Lee, M. (2018). Intuitive archeology: Detecting social transmission in the design of artifacts.

Seyfarth, R. M., Cheney, D. L., & Marler, P. (1980a). Monkey responses to three different alarm calls: evidence of predator classification and semantic communication. *Science*, *210*, 801–803.

Seyfarth, R. M., Cheney, D. L., & Marler, P. (1980b). Vervet monkey alarm calls: semantic communication in a free-ranging primate. *Animal Behaviour*, *28*, 1070–1094.

Tenenbaum, J. B., Kemp, C., Griffiths, T. L., & Goodman, N. D. (2011). How to grow a mind: Statistics, structure, and abstraction. *Science*, *331*, 1279–1285.

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

Ullman, T., Baker, C., Macindoe, O., Evans, O., Goodman, N., & Tenenbaum, J. B. (2009). Help or hinder: Bayesian models of social goal inference. In *Advances in Neural Information Processing Systems* (pp. 1874–1882).

Webb, E. J., Campbell, D. T., Schwartz, R. D., & Sechrest, L. (1966). *Unobtrusive measures: Nonreactive research in the social sciences*. Rand McNally Chicago.

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*, 1–34.

Yildirim, I., Saeed, B., Bennett-Pierre, G., Gerstenberg, T., Tenenbaum, J., & Gweon, H. (2019). Explaining intuitive difficulty judgments by modeling physical effort and risk.