Learning What is Where from Social Observations

Julian Jara-Ettinger (jjara@mit.edu)
Chris L. Baker (clbaker@mit.edu)
Joshua B. Tenenbaum (jbt@mit.edu)
Department of Brain and Cognitive Sciences, MIT
Cambridge, MA 02139

Abstract

Observing the actions of other people allows us to learn not only about their mental states, but also about hidden aspects of a shared environmental situation – things we cannot see, but they can, and that influence their behavior in predictable ways. This paper presents a computational model of how people can learn about the world through these social inferences, supported by the same *Theory of Mind* (ToM) that enables representing and reasoning about an agent's mental states such as beliefs, desires, and intentions. The model is an extension of the Bayesian Theory of Mind (BToM) model of Baker et al. (2011), which treats observed intentional actions as the output of an approximately rational planning process and then reasons backwards to infer the most likely inputs to the agent's planner – in this case, the locations and states of utility sources (potential goal objects) in the environment. We conducted a large-scale experiment comparing the world-state inferences of the BToM model and those of human subjects, given observations of agents moving along various trajectories in simple spatial environments. The model quantitatively predicts subjects' graded beliefs about possible world states with high accuracy – and substantially better than a non-mentalistic feature-based model with many more free parameters. These results show the power of social learning for acquiring surprisingly fine-grained knowledge about the world.

Keywords: Social Cognition; Theory of Mind; Social Learning; Reinforcement Learning

Introduction

The most obvious way to learn about the world is by direct observation. You may believe there is a Starbucks across the street from your office because you have passed it many times, and believe it is open at this moment because you just passed by a few minutes ago and saw a number of people going in and out. But many aspects of the world are unobservable and must be inferred indirectly, often based on observing the actions of other people who know or perceive what you do not. Consider the situation of driving or biking and needing to turn left at an intersection onto a busy street, across oncoming traffic. Of course before turning you will check to see whether there are any cars coming down the busy street from the left, but suppose there is a large truck parked on the street, blocking your view so that you cannot see whether there is any oncoming traffic. You may inch out slowly until you can see, but you may also observe what other drivers or pedestrians are doing. If they are in a position to see the oncoming cars that you cannot, and if they are crossing the busy street at the same point you wish to turn, then it is a good bet that your turn would also be safe.

Making such a judgment is literally betting your life on a mental model of another person's cognitive processes – a Theory of Mind (ToM) (e.g. Dennett, 1987; Wellman, 1990; Gopnik & Meltzoff, 1997). Implicitly you assume basic aspects of rationality in the person you see crossing the street: that they want to cross safely, that they update their beliefs about the presence of oncoming cars based on what they can see, and that they plan their actions appropriately to achieve their goals given their beliefs. If they are clearly paying attention to the side of the street you cannot see, and they are walking across unhurriedly and unworriedly, it is then a good bet that no traffic is headed imminently toward them; if they are jumping or dashing out of the way, that is another story.

Accounts of distinctively human cognition often emphasize the sophisticated representational power of people's ToM, as in the capacity to represent arbitrary belief states, false beliefs as well as true ones, and predict how people will act accordingly. But just as or more important is the sophisticated inferential power of ToM: how we can learn about the contents of other agents' mental states, or even the structure of the world, by reasoning backwards to the best explanations of agents' observed behaviors. This kind of inverse reasoning underlies not only the traffic example above, but many other situations of practical importance for everyday cognition. For example, if you see people filing out of a new restaurant with contented looks, it is a good bet the food inside is satisfying. If you see someone enter the restaurant with an expression of eager anticipation, then exit a moment later and start looking for a different place to eat, you might guess that the restaurant is unexpectedly closed – or perhaps he mistook it for a different place. If your friend the foodie goes far out of his way while visiting a new city to visit a particular restaurant, you can bet that place is one of the city's best.

In this paper, we present a computational model of this social-learning capacity – inferring the world's state from observing other agents' behavior, guided by ToM. Similar inferential abilities have been studied in infants (Csibra, Biró, Koós, & Gergely, 2003) and adults (Goodman, Baker, & Tenenbaum, 2009), and the latter paper presented a computational model similar to ours in key respects (but focused on causal learning). Our work is the first to test people's social learning against rational model predictions in a large-scale quantitative experiment, showing that people can form surprisingly accurate fine-grained beliefs about the relative probabilities of different possible worlds from sparse social observations – just a single agent moving along a single goal-directed path of intentional action. We contrast our model with a non-intentional, non-ToM account based on low-level features of the agent's motion. Even when we introduce many free parameters in the form of variable feature weights, and optimize their values to best fit people's world-state inferences, the feature-based alternative performs substantially worse than a ToM-based model with many fewer parameters.

Computational framework

A rapidly growing body of research suggests that human judgments about intentional agents' mental states (goals, preferences, beliefs) can be modeled as probabilistic inverse planning, inverse optimal control, or inverse decision-making: Bayesian inferences over predictive models of agents' rational behavior (Baker, Saxe, & Tenenbaum, 2009; Lucas, Griffiths, Xu, & Fawcett, 2009; Bergen, Evans, & Tenenbaum, 2010; Jern, Lucas, & Kemp, 2012; Baker, Goodman, & Tenenbaum, 2008; Ullman et al., 2010; Tauber & Steyvers, 2011). Here we adopt the Bayesian ToM (BToM) formulation of Baker, Saxe, and Tenenbaum (2011), expressing relations between the world's state, an agent's state, and the agent's observations, beliefs, desires, and actions in terms of a rational-agent model known as a partially observable Markov decision process (POMDP) (Kaelbling, Littman, & Cassandra, 1998). This captures a probabilistic version of the classical rational agent who updates their beliefs to conform with their observations and chooses sequences of actions expected to achieve their desires given their beliefs. The causal schema for BToM is shown in Fig. 1(a).

Baker et al. (2011) used the BToM model to explain human observers' joint inferences about agents' beliefs and desires, based on how these mental states guided agents' actions exploring a small, spatially structured world with different sources of utility (candidate goals) in different locations. Observers had full knowledge of the agent's situation and world state, but the agent only learned about the world piecemeal (based on line-of-sight perceptual access) as it explored. In contrast, in this paper we consider scenarios where *neither* the agent nor the observer have full access to the state of the world. The agent again has line-of-sight perceptual access, but the observer sees none of the utility sources (candidate goals) in the environment; these must be inferred from observing the agent's movements. At first blush, this inference problem might seem hopelessly underconstrained; however, we will show that when the observer knows the agent's preferences, and if those preferences are strong enough, then joint inferences about the agent's beliefs and the unobservable world state are possible.

To illustrate how this works, consider the scenario shown in Fig. 2. At a certain university food hall, every day at lunchtime three different food carts arrive: an Afghani (A) cart, a Burmese (B) cart, and a Colombian (C) cart. The food hall contains three rooms, West (W), North (N) and East (E), and on any given day, any cart can be in any room. Harold, the student shown in the figure, always prefers to eat at cart A over carts B and C, and prefers to eat at cart B over cart C. Furthermore, carts A and B can be *open* or *closed* when Harold arrives; he only goes to a cart if he sees that it is open. Cart C is always open and is the last resort when all others are closed.

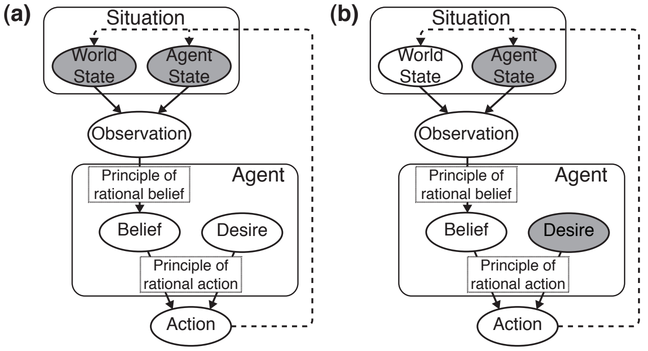

Figure 1: Causal structure of Theory of Mind. Traditional accounts of ToM (e.g., Dennett, 1987; Wellman, 1990; Gopnik & Meltzoff, 1997) have proposed informal versions of these schemata, characterizing the content and causal relations of ToM in commonsense terms, e.g., "seeing is believing" for the principle of rational belief. **(a)** Schematic of the Bayesian theory of mind (BToM) model proposed by Baker et al. (2011). Grey shaded nodes – World State and Agent State – are assumed to be observed (for the observer; not necessarily for the agent, as described in the main text). **(b)** Our extension of BToM to allow inference of hidden aspects of the World State by observing an agent's behavior. Here, the Agent State and Desire are observed, but the World State is only partially observable for both agent and observer.

Fig. 2(a) shows a hypothetical path that Harold could take, ending in the North room. What, if anything, does this tell us about the cart locations? From where Harold enters the food hall, he can observe the cart in the North room. Next, he checks the East room, indicating that either cart A is not in the North room, or that cart A is in the North room, but is closed. When Harold returns to the North room, only one possibility remains: that he saw cart A in the East room, but it was closed, so he returned to the North room to eat at cart B, which was open (this cart configuration is shown in Fig. 2(d), row 1, column 3). Crucially, this inference also depends on Harold's *not* checking the West room, which is consistent with several other configurations in Fig. 2(d). In our experiment, 66% of participants rated the correct configuration to be the most likely in this condition (chance = 17%).

Informal Model Sketch

Fig. 1 sketches the causal schema for BToM. For concreteness, we will describe the content of the model in terms of our food carts scenario, but in principle the BToM framework can be defined over arbitrary state and action spaces. In our food cart examples, there are 24 possible World States: 6 possible cart configurations (shown in Fig. 2(d)) times 4 possible joint combinations of open/closed for carts A and B. There are 12 possible Agent States, one for each grid square in the food hall scenario. Agents' Observations provide information about the World State, conditioned on the Agent State, and are based on line-of-sight visibility – Fig. 2(b,c) give examples of what can be seen from different vantage points. The observer represents an agent's Belief as a probability distribution over possible World States. The observer maintains a

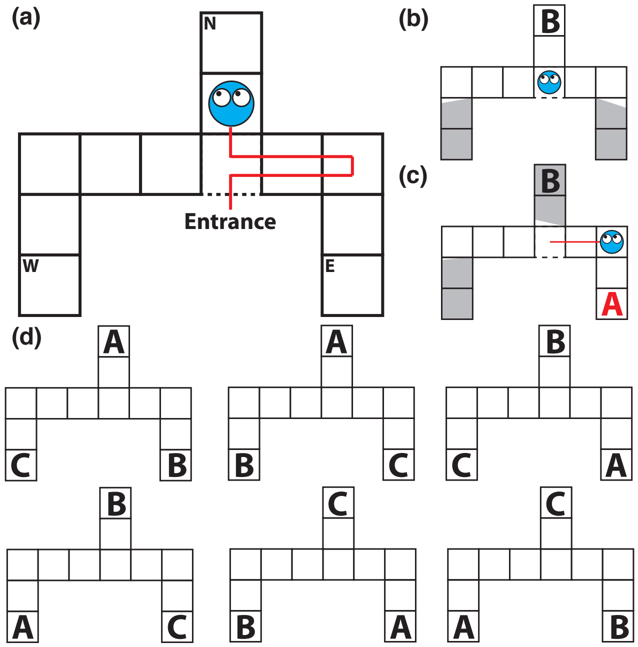

Figure 2: Example experimental stimulus. **(a)** Example of an observed path. The task is to figure out where each of three food carts is, given the trajectory the agent took. **(b)** In the agent's initial position he can observe the North spot with food truck B in it. However, he doesn't know where his favorite cart A is. **(c)** Agent's state when he travels to the entrance of the East hallway. He can now observe cart A being closed and remembers having seen cart B in the North spot. With this information he can deduce that cart C is in the West room and so his best option is to choose the North spot, producing the path shown in (a). **(d)** Possible configurations the carts can take independent of them being closed or open. In the experiment, subjects ranked these six configurations for each path.

finite set of possible Beliefs the agent could hold, drawn from a prior over initial Beliefs, and simulates the agent's Belief update for each possible Observation, given the Agent State and World State. An agent's Desire is captured by utilities for each cart which capture the preference relation $A \succ B \succ C$.

Given the representational content of the nodes, BToM expresses the functional form of the causal relations in Fig. 1 in terms of POMDPs, which capture the dual principles of rational Belief and Action in Fig. 1. To generate a POMDP policy for each initial Belief point, we employ an implementation of the SARSOP algorithm (Kurniawati, Hsu, & Lee, 2008), provided by the APPL POMDP solver. These policies represent a predictive, Belief- and Desire-dependent distribution over the agent's actions.

The schema in Fig. 1 illustrates the conditional dependencies involved in the model of the agent. For clarity, in this informal sketch we suppress the temporal nature of the model; technical details of dynamic inference are provided in Baker et al. (2011). The predictive distribution over the agent's Action, given its Desire, Beliefs, the World State and the Agent State (abbreviating variable names as A, D, B, W, S respectively) is:

$$p(A|D,W,S) \quad = \quad \sum_{B} p(A|B,D) \sum_{O} p(B|O) p(O|W,S). \qquad (1)$$

In Fig. 1(b), the World State is unknown, and the problem of "learning what is where" involves inferring the World State, given an agent's Desire and Action using Bayes' rule:

$$p(W|A,D,S) \propto p(A|D,W,S) p(W). \qquad (2)$$

Intuitively, this involves evaluating the likelihood of every possible World State, given the agent's Action, Desire and Agent State, and integrating these likelihoods with the prior over possible World States. Evaluation of each likelihood also requires simultaneously inferring and updating the agent's Beliefs over time.

An alternative cue-based model

To assess the intrinsic difficulty or logical complexity of our task, we formulated a cue-based alternative to our BToM account of social inference of food cart locations. We name the alternative model F-40; the model considered 7 key features and fit 40 free parameters (one for each feature, plus an additive constant, multiplied by 5 independent response variables) using multinomial logistic regression to minimize the error in prediction of human judgments. The features were chosen to capture key moments in the paths that were strongly indicative of a preferred cart being in a particular location. Specifically, for each room, we assigned a unique vantage point at which the agent could see what was in that room, and could choose to either commit to eating at that room by moving North/South, or commit to moving to another vantage point by moving East/West. The set of vantage points is indicated by the marked cells in Fig. 5. Features Toward and Away were computed for each room by counting the number of times the agent moved to or away from that room, starting from that room's vantage point. In addition to the 6 Toward and Away features, the 7th feature recorded whether or not the condition was part of the introduction (in which carts could not be closed) or the main experiment. Because of its large number of free parameters, we hypothesized that F-40 would capture those regularities in people's judgments that could be explained by low-level movement properties.

Experiment

Design

Fig. 2 illustrates our experimental design. On each trial, subjects were shown either a complete or an incomplete path that the agent took. They were then asked to rate on a scale from 0 to 10 (with 0 meaning "Definitely Not"; 10 "Definitely"; and 5 "Maybe") how much they believed each possible configuration of carts was the real one. Fig. 2(d) shows the six possible configurations of carts that subjects rated on each trial. Food cart names as well as stimulus order were randomized across subjects. For simplicity we will refer to the carts as Afghani (A), Burmese (B), and Colombian (C), always with the preference order: $A \succ B \succ C$.

In this scenario there are 24 possible worlds (6 possible permutations of the cart's locations multiplied by 4 permutations of carts A and B being open or closed). Stimuli were generated as follows. We assume that the agent always starts at the entrance of the North hallway, being able to chose between entering that hall, going to the West hall, or going to the East hall. An exhaustive list of possible paths was constructed by listing all possible combinations of the short-term goals of the agent (go to entrance of W hall, go to entrance of N hall, and go to entrance of W hall), assuming that the first time a hall is selected it is for the purpose of exploration, and any selection of a hall that had been selected before is for exploitation, meaning the agent has chosen where to eat. From the eleven exhaustively enumerated paths, two paths that only produced permutations of beliefs were removed, leaving a total of 9 complete paths. In addition, 7 incomplete paths (subsequences of the 9 complete paths) which produce different judgments were selected. Lastly, three of these paths were duplicated in initial displays in which all carts are assumed to be open, shown to subjects to familiarize them with the task. This produced a total of 19 different paths (see Fig. 3) for which each subject rated the six possible configurations of carts, for a total of 114 judgments per subject.

Participants

200 U.S. residents were recruited using the Amazon Mechanical Turk. 176 subjects were included in the analysis, with 24 excluded due to server error.

Procedure

Subjects first completed a familiarization stage, which began with an explanation of the basic food cart setting, and allowed subjects to provide judgments for three paths where the food carts were assumed to always be open. Next, the possibility that carts could be closed was introduced with a step by step example. The experimental stage immediately followed.

Results

We begin by analyzing the fit between people's judgments and our two models. Fig. 3 shows the average human rating of the likelihood of each cart configuration, the BToM model, and the F-40 model. In Fig. 3 it is clear that both models perform well in capturing the general contours of the mean subject belief, but with a quantitative difference in their explanatory power.

The BToM model has four parameters that were not fit to the data: three parameters indicating how strong the preference for each food cart is, and a discount parameter indicating the tradeoff between immediate and delayed rewards. Intuitively, these four parameters together determine whether an agent is willing to spend time and energy finding food carts he likes better or whether he should settle for a closer cart. These parameters were set only qualitatively, to ensure that the agent would have a strong preference order that would motivate him to explore the environment until he finds the best option.

In contrast, the F-40 model has forty free parameters fit to the average subject ratings, and so, by construction, the fit is very close to human judgment. Looking deeper into the model, there were no outstanding predictive features of the path that would determine the food cart ordering. That is, F-40 shows a great capacity to mimic human reasoning, but it fails to capture the essence of the task. This clear in Fig. 4, where we can see that mean human judgments have a $r = 0.91$ correlation with the BToM model, but a correlation of $r = 0.64$ with the F-40 model. As we can see in the scatterplot, BToM comes much closer to explaining the variance in the human data, while F-40 is much less accurate overall. Fig. 4(c) shows that for a strong majority of individual subjects, the BToM model provides a superior fit. To further assess the statistical significance of the models we performed a Bootstrap Cross-Validated Correlational Analysis (Cohen, 1995). For 10,000 iterations, we trained F-40 on randomly selected subsets of paths and compared its performance on the remaining untrained paths. This produced average correlations of $r = -0.0733$, -0.0832 and 0.0830, for training sets of size 16, 17 and 18 (and testing sets of size 3, 2, and 1), respectively. A similar analysis with BToM (for which no parameters were fit to data) yielded correlations of $r = 0.9015$, 0.8922 and 0.8714 for testing sets of size 3, 2 and 1, respectively. These analyses suggest that the feature-based model is not tapping into the cognitive mechanisms underlying human performance, but rather just fitting the data without strong predictive power.

This is clear in Fig. 5, where two paths that contrast the models' performance are shown. For path 1, BToM is capable of realizing that if the carts were set as C, B, A/closed in positions West, North, and East respectively, then the agent would have no reason to visit the West position, since by the time it has observed B and A/closed it already has all the information it needs to make its final choice. It is this type of fine grained reasoning that allows BToM to make subtle inferences when F-40 fails as a result of mimicking the data rather than predicting it.

Discussion

In this work we have proposed a Bayesian Theory of Mind model to explain how we make sense of the world by observing how others interact with it. Our experiment shows that subjects produce very similar predictions to that of the ideal Bayesian observer. We have compared the BToM model to a feature-based regression model (F-40) that was fit to subjects' mean judgments. Although the F-40 model appears to be a good competitor, we show that at both the individual and average level, the correlation with the BToM model is substantially higher compared to the correlation with F-40. Further analysis showed how BToM is capable of more subtle, fine-grained reasoning, making sensible inferences in several situations where F-40 gives counter-intuitive predictions.

One interesting point is that the BToM model is more sensitive to the precise geometry of the environment than humans

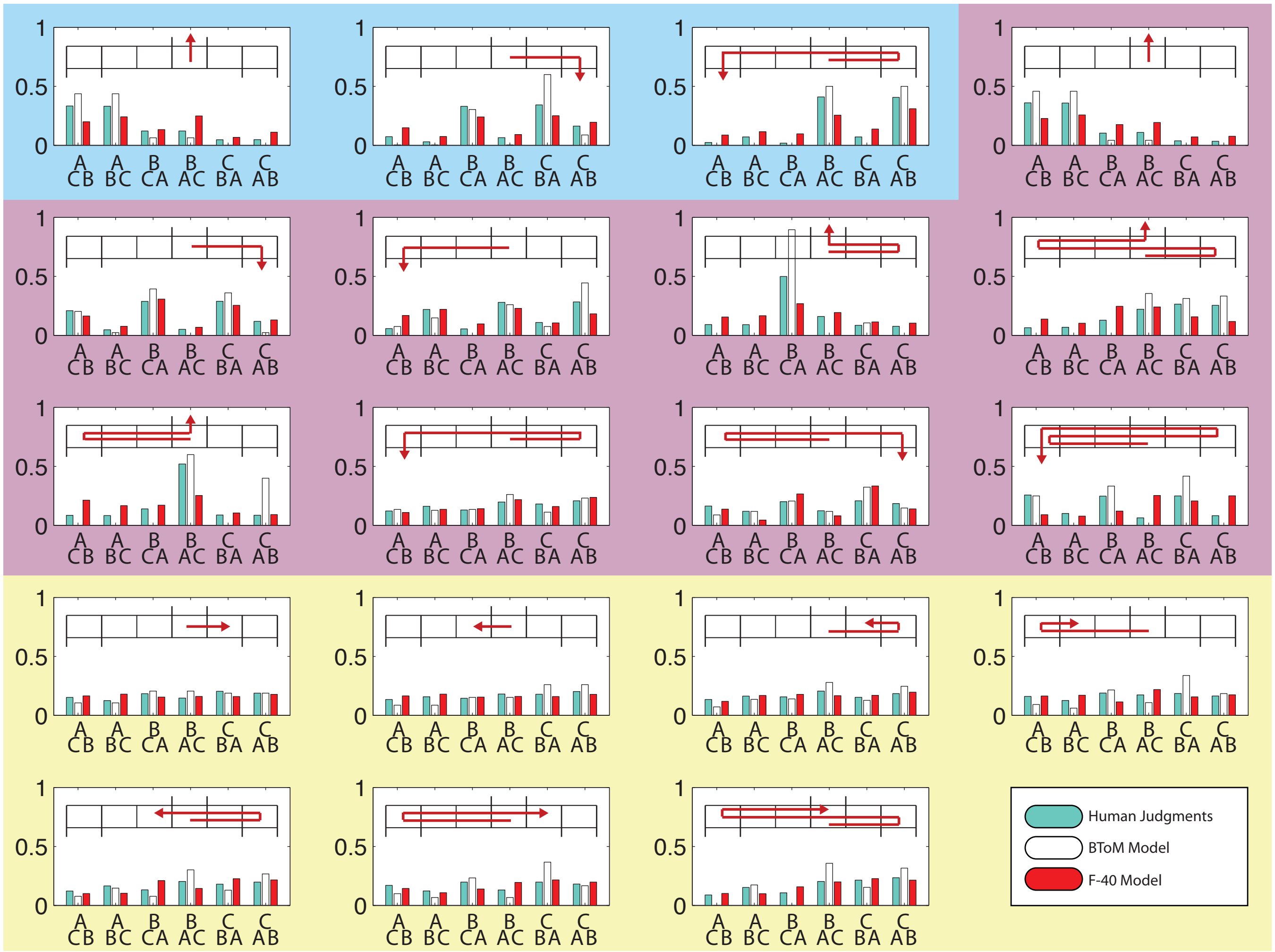

Figure 3: Mean subjects' judgments (normalized degrees of belief in each of six possible configurations of the food carts), along with BToM and F-40 predictions for the 19 displayed paths. The agent's preference order for these carts is always known to be $A \succ B \succ C$, and carts A and B may be open or closed. The first three conditions were those used in the familiarization phase. The second block used "completed" paths, in which the agent committed to a particular cart in the last frame. The last block of conditions used "incomplete" paths, in which the agent's final destination had not yet resolved.

seem to be. Specifically, because of the asymmetry of the hallway in our experiment, the model assigned a significantly higher cost to checking the West hallway versus checking the East hallway. Thus, when the model observed the agent going West, it reasoned that the agent must have had some prior belief in the presence of a high value cart in the West hallway or a low value cart in the East hallway that made him go through the more lengthy path versus the shorter path to check the East hallway. In contrast, subjects did not appear to be sensitive to the distance mismatch and produced relatively symmetric judgments on paths that had the same structure but traveled in opposite directions. This is particularly evident in the plot (3,1) of Fig. 3. In this path, subjects believed that the agent had already found carts (A) and (B) and therefore had no need to visit the East room. The model however, when observing the agent choose the longer path, reasoned that there was some prior belief the agent had that could have been wrong. This leads the model to consider it possible that the (B) food truck was in the East room but that the agent had a prior belief that it was closed and therefore did not bother checking it. Analogous disparities were found by Baker et al. (2011), and in ongoing work we are investigating more qualitative representations of the spatial structure of the environment that might support a closer match between BToM model reasoning and human judgments.

In sum, these results show the power of social inference for acquiring surprisingly fine-grained knowledge about the world. ToM is typically thought of as a system of knowledge for reasoning about the mental states and actions of intentional agents, but it is not only that. In the context of a Bayesian framework, actions of other agents become clues to any aspects of the environment that causally influence their behavior – sometimes the only clues available. ToM thus also provides an essential tool for learning about the world.

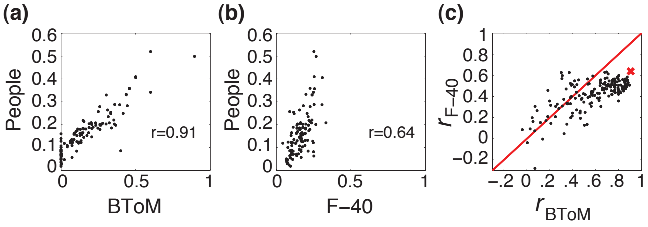

Figure 4: Comparison of models and normalized human judgments. **(a)** BToM vs. mean normalized human judgments. Each point represents a mean human rating plotted against the corresponding model prediction; there are 114 points in all (19 conditions times ratings for 6 possible cart configurations), with an overall correlation of $r = 0.91$. **(b)** F-40 vs. mean normalized human judgments; analogous to analysis (a), with an overall correlation of $r = 0.64$. **(c)** Scatter plot of individual subjects' correlations with BToM vs. individual subjects' correlations with F-40; 176 points in all (one point for each subject). For 80% of subjects, the correlation between BToM and that subject's ratings is higher than the correlation of F-40 with that subject's ratings. The bold "X" plots the correlation of BToM with the mean human judgments vs. that of F-40.

Acknowledgements This work was supported by ARO MURI contract W911NF-08-1-0242 and ONR MURI contract 1015GNA126. We thank the Motion Modeling, Analysis and Planning group at the National University of Singapore for providing the APPL POMDP solver. APPL is available at: `http://bigbird.comp.nus.edu.sg/pmwiki/farm/appl`.

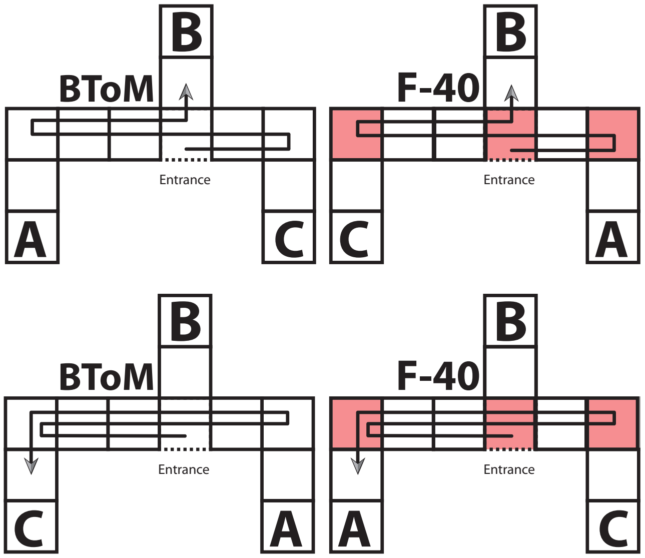

Figure 5: Comparing MAP predictions (the configuration shown for each path) of BToM and F-40 models for two different paths. F-40's errors reveal how a non-mentalistic approach fails to use context specific reasoning to make accurate predictions. For the F-40 model, marked grid squares indicate "vantage points" used to compute the features.

References

Baker, C. L., Goodman, N. D., & Tenenbaum, J. B. (2008). Theory-based social goal inference. In *Proceedings of the Thirtieth Annual Conference of the Cognitive Science Society* (pp. 1447–1455).

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

Baker, C. L., Saxe, R., & Tenenbaum, J. B. (2011). Bayesian theory of mind: Modeling joint belief-desire attribution. In *Proceedings of the Thirtieth Third Annual Conference of the Cognitive Science Society* (p. 2469-2474).

Bergen, L., Evans, O. R., & Tenenbaum, J. B. (2010). Learning structured preferences. In *Proceedings of the Thirty-Second Annual Conference of the Cognitive Science Society* (pp. 853–858).

Cohen, P. R. (1995). *Empirical methods in artificial intelligence*. Cambridge, MA: MIT Press.

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

Dennett, D. C. (1987). *The intentional stance*. Cambridge, MA: MIT Press.

Goodman, N. D., Baker, C. L., & Tenenbaum, J. B. (2009). Cause and intent: Social reasoning in causal learning. In *Proceedings of the Thirty-First Annual Conference of the Cognitive Science Society* (pp. 2759–2764).

Gopnik, A., & Meltzoff, A. N. (1997). *Words, thoughts, and theories*. Cambridge, MA: MIT Press.

Jern, A., Lucas, C. G., & Kemp, C. (2012). Evaluating the inverse decision-making approach to preference learning. In *Advances in Neural Information Processing Systems.*

Kaelbling, L. P., Littman, M. L., & Cassandra, A. R. (1998). Planning and acting in partially observable stochastic domains. *Artificial Intelligence*, *101*, 99–134.

Kurniawati, H., Hsu, D., & Lee, W. (2008). SARSOP: Efficient point-based POMDP planning by approximating optimally reachable belief spaces. In *Proc. Robotics: Science and Systems.*

Lucas, C. G., Griffiths, T. L., Xu, F., & Fawcett, C. (2009). A rational model of preference learning and choice prediction by children. In *Advances in Neural Information Processing Systems 21* (pp. 985–992).

Tauber, S., & Steyvers, M. (2011). Using inverse planning and theory of mind for social goal inference. In *Proceedings of the Thirtieth Third Annual Conference of the Cognitive Science Society.*

Ullman, T. D., Baker, C. L., Macindoe, O., Evans, O., Goodman, N. D., & Tenenbaum, J. B. (2010). Help or hinder: Bayesian models of social goal inference. In *Advances in Neural Information Processing Systems 22* (pp. 1874–1882).

Wellman, H. M. (1990). *The child's theory of mind*. Cambridge, MA: MIT Press.